- Each member of L is less than every member of U.
- L has no greatest member.

Then we are ready to define addition and multiplication for all positive real numbers (or *reals*, for short).

Definition If (L_1, U_1) and (L_2, U_2) are *positive reals* then

- Their sum is the real number (L, U) such that

$$L = \{x_1 + x_2 : x_1 \in L_1 \text{ and } x_2 \in L_2\}$$

  and U consists of the remaining positive rationals.
- Their product is the real number (L, U) such that

$$L = \{x_1 x_2 : x_1 \in L_1 \text{ and } x_2 \in L_2\}$$

  and U consists of the remaining positive rationals.

After this, $+$ and $\times$ can be extended to negative reals the same way they were for negative integers in Section 1.4. The set of all real numbers is denoted by $\mathbb{R}$.

To see that there is method in this madness, let us check that $\sqrt{2}\sqrt{2} = 2$.

By definition of $\sqrt{2}$ and the definition of multiplication, the L for $\sqrt{2}\sqrt{2}$ is $\{x_1x_2 : x_1^2 < 2 \text{ and } x_2^2 < 2\}$, where the x_1 and x_2 are rational.

It follows that each $x_1^2x_2^2$ is less than $2 \times 2 = 4$, and therefore x_1x_2 is a rational x less than 2.

Conversely, any rational x less than 2 can be written as $x = x_1x_2$, where x_1 and x_2 are rationals with $x_1^2 < 2$ and $x_2^2 < 2$. This is because the rationals crowd together arbitrarily closely, and hence so do their squares. It follows that there are rational squares as close as we please to x, and if x_1 is chosen with x_1^2 sufficiently close to x and $x_2 = x/x_1$, then $x_1x_2 = x$ and both $x_1^2 < 2$ and $x_2^2 < 2$.

Thus the L for $\sqrt{2}\sqrt{2}$ is $\{x < 2\}$, which is the L for 2, as required.

Exercises

It is now possible to appreciate Dedekind's claim that

> in this way we arrive at real proofs of theorems (as, e.g., $\sqrt{2}\sqrt{3} = \sqrt{6}$), which to the best of my knowledge have never been established before. [Dedekind (1872), p.22]

As we can see from the example $\sqrt{2}\sqrt{2} = 2$, proving such equations for numbers is very different from proving them for lengths, mainly because the product of irrational numbers is defined so differently from the product of lengths. Recall from Section 2.5 that the Greeks defined the product of lengths $\sqrt{2} \times \sqrt{3}$ to be the rectangle with sides $\sqrt{2}$ and $\sqrt{3}$, and it could be shown equal to $\sqrt{6}$ only by cutting and pasting to form a rectangle with sides $\sqrt{6}$ and 1. Dedekind's theory of irrational numbers gives us a rigorous alternative.

3.3.1. Prove that the numbers $\sqrt{2}$, $\sqrt{3}$, and $\sqrt{6}$ satisfy $\sqrt{2}\sqrt{3} = \sqrt{6}$.

I admit this proof is tedious, but once one such proof has been done, the same routine can be followed in other cases, like the following.

3.3.2. Prove that the numbers $\sqrt[3]{2}$, $\sqrt[3]{3}$, and $\sqrt[3]{6}$ satisfy $\sqrt[3]{2}\sqrt[3]{3} = \sqrt[3]{6}$.

The corresponding theorem about lengths cannot be proved geometrically, because the lengths are not constructible! Thus Dedekind's definition of product of numbers gives us everything we could previously do with the product of lengths, and more. It is not only *possible* to treat lengths as numbers, but it is an advantage.

While on the subject of defining irrational numbers, it should be explained where infinite decimals like $\sqrt{2} = 1.41421356\ldots$ fit in. As the arrangement of numbers $> \sqrt{2}$ and numbers $< \sqrt{2}$, on page 78 suggests, the symbol $1.41421356\ldots$ is a concise way to describe the infinite sequence of rationals $1, 1.4, 1.414, 1.4142, 1.41421, \ldots$, which in turn is part of the lower set L for $\sqrt{2}$. The sequence is said to be *cofinal* with L, because they "end at the same place"; L consists of the rationals less than members of the sequence. For this reason, $1.41421356\ldots$ contains the same information as L, and hence can also serve to represent $\sqrt{2}$.

The main advantage of $1.41421356\ldots$ is that we understand its finite decimal approximations $1, 1.4, 1.414, 1.4142, 1.41421, \ldots$ and we are used to computing with them. However, it is not as easy to define sum and product for infinite decimals as it is for Dedekind cuts.

3.3.3. Try to define sum and product for infinite decimals.

Apart from this, the main disadvantage of $1.41421356\ldots$ is the lack of any apparent pattern in the sequence of its digits. In fact, the simplest way to describe its finite decimal approximations is to say that they are respectively the largest 1-digit, 2-digit, 3-digit, ... decimals whose squares are less than 2. Thus we end up essentially repeating Dedekind's definition.

There is in fact another way to describe $\sqrt{2}$ by a process with an infinite, but repeating, pattern. This is the *continued fraction algorithm*, which will be described in Chapter 8. It is definitely not the case that the infinite decimal for $\sqrt{2}$ eventually becomes repeating, because this does not happen for any irrational number.

3.3.4. Let $x = 0.\overline{a_1 a_2 \dots a_k}$ be a number whose infinite decimal consists of the sequence $a_1 a_2 \dots a_k$ repeated indefinitely. Using the infinite geometric series, or otherwise, show that x is rational.

3.3.5. Let $y = 0.b_1 b_2 \dots b_j \overline{a_1 a_2 \dots a_k}$ be a number whose infinite decimal, after the first j places, consists of the sequence $a_1 a_2 \dots a_k$ repeated indefinitely. Show that y is also rational. (Such a decimal is called *ultimately periodic*.)

3.3.6. Show that any rational number has an ultimately periodic decimal.

3.4 The Line

Having seen how individual lengths, like $\sqrt{2}$, can be reborn as numbers, the next step is to see whether these numbers make up anything we would recognize as a line.

One crucial property they have is *order*: if α and β are any distinct real numbers, then either $\alpha \le \beta$ or $\beta \le \alpha$. In fact, if $\alpha = (L_\alpha, U_\alpha)$ and $\beta = (L_\beta, U_\beta)$, it is natural to say that $\alpha \le \beta$ if and only if L_α is *contained in* L_β, because this captures the idea that α separates the rationals at a position $\le$ the position where they are separated by β. If L_α is not contained in L_β then there is a rational r in L_α but not in L_β, in which case all members of L_β are less than r. Then L_β is contained in L_α, and hence $\beta \le \alpha$ by our definition. Thus the real numbers have an order, like points on a line.

The second crucial property of the line is what Dedekind called its *continuity*, or absence of gaps. Do the real numbers have this property? Well, the real numbers were created precisely by filling all the gaps in the rationals. A gap occurs where the rationals split into a lower set L with no greatest member and an upper set U with no least member, and we filled each such gap by the irrational number (L, U). We could even say that the number (L, U) *is* the gap in the rationals!

Thus the irrationals fill all gaps in the set $\mathbb{Q}$ of rationals, by definition. Can the resulting set $\mathbb{R}$ of reals have gaps? The answer is *no*, because a gap in $\mathbb{R}$ implies an "unfilled gap" in $\mathbb{Q}$. In fact, if $\mathbb{R}$ is separated into a lower set $\mathcal{L}$ and an upper set $\mathcal{U}$, consider the following sets of rationals r:

$$L = \{r : r \leq \text{some member of } \mathcal{L}\},$$
$$U = \{r : r \geq \text{some member of } \mathcal{U}\}.$$

Because $\mathcal{L}$ and $\mathcal{U}$ together include all reals, L and U together include all rationals. And because $\mathcal{L}$ and $\mathcal{U}$ have no members in common, neither do L and U. L and U therefore define a number (L, U). But then (L, U) is either the least member of $\mathcal{U}$ or the greatest member of $\mathcal{L}$, so there is no gap where $\mathbb{R}$ is separated.

The "no gaps" property of $\mathbb{R}$ is now called *completeness*, because Dedekind's word "continuity" is used for a related property of functions or curves. We also say that $\mathbb{R}$ is the *completion of* the set $\mathbb{Q}$ of rationals. At any rate, ordering and completeness are exactly what we were looking for to model the concept of line in geometry, so $\mathbb{R}$ fits the bill. We often call $\mathbb{R}$ the *real line*. It now remains to check that pairs (x, y) of real numbers can be made to behave like points of the plane, and the conversion of geometry to arithmetic will be complete. We shall do this in the next section.

Identifying the line with the real numbers has other advantages, apart from allowing the free use of arithmetic in geometry. It gives answers to questions that cannot really be settled by geometric intuition, because they involve the "infinitely small." For example, most people have the feeling, at first, that $0.999999999\ldots$ cannot be equal to 1, because it seems to be less than 1 by an "infinitesimal amount"; maybe 1 is the "next number" after $0.999999999\ldots$. Such feelings are dispelled by Dedekind's picture of real numbers. In fact, we can say definitely that:

1. There is no such thing as the "next point," because there is no such thing as the "next real number." If α and β are distinct real

numbers then $(\alpha + \beta)/2$ is a number that lies strictly between them.

In fact, there is a rational number strictly between them. For example, if $\alpha < \beta$, take any number in the lower set for β that is not in the lower set for α. (Here it is convenient that we defined reals so that their lower sets never have greatest members.)

2. There are no "infinitesimal distances" between points, that is, distances that are nonzero yet less than any positive rational. This is because there is no positive number less than all positive rationals. In fact, if α is a positive real number, then the lower set for α must include a positive rational, and all numbers in the lower set are less than α.

Exercises

Another important property of $\mathbb{R}$ is the existence of *least upper bounds*: if the numbers in some set S are all $\leq$ some number α, then there is a least number $\lambda \geq$ all members of S. This number $\lambda \leq \alpha$ is called the *least upper bound of* S.

We can obtain λ by taking its lower set to be the union of all the lower sets L_β of members β of S. (That is, L_λ is the set of all the members of all the sets L_β.)

3.4.1. Deduce from this definition that $\beta \leq \lambda$ for each β in S.

3.4.2. Show also that if $\mu < \lambda$ then $\mu <$ some β in S.

The existence of least upper bounds is in fact another way to state the completeness of $\mathbb{R}$.

3.4.3. Suppose $\mathbb{R}$ is separated into a lower set $\mathcal{L}$ and an upper set $\mathcal{U}$. Use the existence of least upper bounds to show this separation is not a gap. (That is, either $\mathcal{L}$ has a greatest member or $\mathcal{U}$ has a least member.)

3.4.4. Conversely, use the nonexistence of gaps to find a least upper bound for any bounded set S.

The "no infinitesimals" property of $\mathbb{R}$ can be stated in another way that goes back to Archimedes. It is called the *Archimedean axiom*; it says that if α and β are positive numbers, with $\alpha < \beta$, then there is a natural number n with $n\alpha > \beta$.

3.4.5. Prove the Archimedean axiom.

3.5 The Euclidean Plane

Now that we have the line, as the set of real numbers x, the plane is obtained by a simple trick. It is the set of *ordered pairs* of real numbers, (x, y). In honor of Descartes, this set is called the *cartesian product*, $\mathbb{R} \times \mathbb{R}$, of the set $\mathbb{R}$ of reals with itself. The main difference between Descartes and us is that he supposed the plane to exist, then gave each point in it a coordinate pair (x, y); we suppose only that numbers exist, we say the coordinate pair (x, y) *is* a point, and that the set of these points is the plane.

We also have to define the distance between points, which is not hard, because we know what it should be from previous experience.

Definition The *Euclidean distance* between $P_1 = (x_1, y_2)$ and $P_2(x_2, y_2)$ is

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

This is prompted by the Pythagorean theorem, because we expect the line segment from (x_1, y_1) to (x_2, y_2) to be the hypotenuse of a right-angled triangle with sides $x_2 - x_1$ and $y_2 - y_1$ (Figure 3.6).

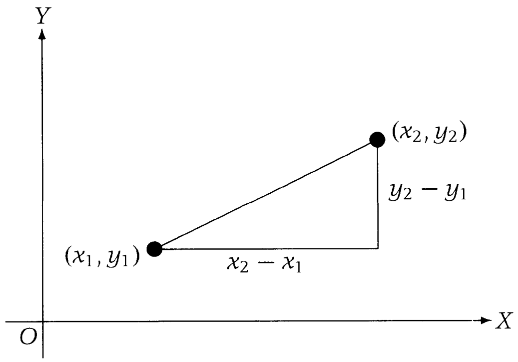

FIGURE 3.6 The distance-defining triangle.

The set $\mathbb{R} \times \mathbb{R}$ with this distance function is called the *Euclidean plane*. As we know, the Pythagorean theorem is a characteristic statement of Euclid's geometry, and by defining distance as we did we have made the theorem true *by definition* in $\mathbb{R}\times\mathbb{R}$. With a different choice of distance function we can get a *non*-Euclidean plane, as we shall see in Section 3.8*.

A *line* is defined to be the set of points (x, y) satisfying an equation of the form $ax+by = c$. A *circle* is defined to be the set of points (x, y) at constant distance r from a point (a, b). It follows from the definition of distance that the equation of the circle is $(x-a)^2+(y-b)^2 = r^2$, as expected. Thus we can re-create the basic concepts of Euclid's geometry in terms of numbers, with the added advantage that Euclid's unstated assumptions about the existence of intersections are guaranteed. There are enough real numbers to solve all the equations that arise when we seek intersections of lines and circles.

Basic properties of distance *It follows from the definition of distance that*

1. *The set of points equidistant from two distinct points is a line.*
2. *Any line is the equidistant set of two points.*
3. *Each point of the plane is determined by its distances from three points not in a line.*

Proof

1. If $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ are any two points, then the points (x, y) equidistant from them both satisfy

$$(x - x_1)^2 + (y - y_1)^2 = (x - x_2)^2 + (y - y_2)^2,$$

which is equivalent to the equation of a line, namely,

$$2(x_2 - x_1)x + 2(y_2 - y_1)y = x_2^2 - x_1^2 + y_2^2 - y_1^2.$$

2. The latter equation represents an arbitrary line $ax + by = c$, provided we can find some constant k such that

$$\begin{aligned} 2(x_2 - x_1) &= ka \\ 2(y_2 - y_1) &= kb \\ x_2^2 - x_1^2 + y_2^2 - y_1^2 &= kc. \end{aligned}$$

Substituting $x_2 = x_1 + \frac{k}{2}a$ from the first equation, and $y_2 = y_1 + \frac{k}{2}b$ from the second in the third gives an equation from which we find

$$k = \frac{4(c - ax_1 - by_1)}{a^2 + b^2}.$$

3. If Q and Q' are two distinct points with the same respective distances from three points P_1, P_2, and P_3, then P_1, P_2, and P_3 lie on the equidistant line of Q and Q'. Hence if P_1, P_2, and P_3 are not in a line there can be only one point Q with given distances from them. □

As an example of the first property, if $P = (-a, 0)$ and $Q = (a, 0)$ then a point (x, y) is equidistant from P and Q if and only if

$$(x + a)^2 + y^2 = (x - a)^2 + y^2,$$

whence

$$x = 0,$$

which is the equation of the axis OY.

Another advantage of this definition of the Euclidean plane is that it admits a concept of "moving" one figure until it coincides with another, as in Pappus' proof that the base angles of an isosceles triangle are equal (Section 2.4). To formalize this idea, we consider functions that "preserve distance."

Definition A function f on $\mathbb{R} \times \mathbb{R}$ is an *isometry* (from the Greek for "same distance") if $d(f(P_1), f(P_2)) = d(P_1, P_2)$ for any two points P_1 and P_2.

An example of an isometry is the function ref_{OY} that sends each point (x, y) to $(-x, y)$. We call this function *reflection in OY* (hence the symbol ref_{OY}) because it captures the intuitive idea of mirror reflection in the line OY. It preserves distances, because if $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ then $\text{ref}_{OY}(x_1, y_1) = (-x_1, y_1)$ and $\text{ref}_{OY}(x_2, y_2) = (-x_2, y_2)$, hence

$$\begin{aligned} d(\text{ref}_{OY}(P_1), \text{ref}_{OY}(P_2)) &= \sqrt{(-x_2 + x_1)^2 + (y_2 - y_1)^2} \\ &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \\ &= d(P_1, P_2). \end{aligned}$$

Now suppose we have a triangle ABC with $CA = CB$. We can re-create Pappus' proof by placing the triangle with A and C on OX, with O at their midpoint, say $A = (-a, 0)$, and $C = (a, 0)$. Because C is equidistant from A and B, it must be on OY by the preceding calculation. If we then reflect triangle ABC in OY, it is mapped onto itself. In particular, the angle at A is mapped onto the angle at C; hence these two angles are equal.

Proving that two angles are equal can usually be done, as here, by moving one to coincide with the other. Actually measuring angles is harder, but it can also be done with the help of the real numbers, as we shall see in Chapter 5.

Exercises

Another useful isometry is the *half turn*, or rotation through π. The half turn about O is the function $\mathrm{rot}_{O,\pi}$ that sends (x, y) to $(-x, -y)$.

3.5.1. Check that the half turn about O is an isometry, and use it to show that vertically opposite angles between lines through O are equal.

We can prove that vertically opposite angles are equal at any point (a, b) with the help of an isometry that moves O to (a, b). The simplest such isometry is the *translation* $\mathrm{tran}_{a,b}$, which moves each point (x, y) to the point $(x + a, y + b)$. It is reversed by the translation $\mathrm{tran}_{-a,-b}$, which sends (x, y) to $(x - a, y - b)$.

3.5.2. Check that $\mathrm{tran}_{a,b}$ is an isometry.

3.5.3. Show that vertically opposite angles at any point (a, b) are equal by translating (a, b) to O, applying a half turn, then translating the angles back to (a, b).

Other classical results about equal angles can also be proved by using isometries to formalize intuitive movements of one angle onto another. An example is the pair of *alternate* angles that occur where a line crosses two parallels (Figure 3.7).

3.5.4. Prove that alternate angles are equal by a suitable combination of translations and half turns.