Exercises

Some of Euclid's axioms correspond to familiar facts about equations, for example, the axiom that a (unique) line can be drawn through any two points.

3.1.1. Find the (unique) linear equation satisfied by two points (a_1, b_1) and (a_2, b_2).

Also prove the parallel axiom.

3.1.2. How would you recognize a line parallel to the line $ax + by = c$ from its equation? Show that its equation has no solution in common with $ax + by = c$ and that there is only one such line through a given point outside the line $ax + by = c$.

3.2 Intersections

The difference between analytic and synthetic geometry can be illustrated with Euclid's very first proposition, the construction of the equilateral triangle on a line segment AB. Recall from Section 2.2 that Euclid did this by finding the intersection of two circles, one with center A and radius AB and the other with center B and radius BA. It follows that each point of intersection is distant from both A and B by the length of AB and hence forms an equilateral triangle. Analytic geometry takes its cue from this construction, but it also finds the intersections by finding the common solutions of the equations to the two circles.

For example, if $A = (-1, 0)$ and $B = (1, 0)$ the two circles have radius 2 and hence their equations are

$$(x + 1)^2 + y^2 = 2^2,$$
$$(x - 1)^2 + y^2 = 2^2.$$

Subtracting the second equation from the first leads to $x = 0$ (as you would expect), and substituting this back in the first equation gives

$$1 + y^2 = 4,$$

whence $y = \pm\sqrt{3}$. Thus the points of intersection are $(0, \pm\sqrt{3})$, either of which can be taken as the third vertex of the equilateral triangle.

Euclid's argument is short and sweet, but it has one defect. It does not follow from his axioms! His axioms guarantee only the existence of circles, not their intersections. This defect can be repaired by introducing axioms about intersections, but only with difficulty, because it is hard to foresee all the situations that may arise. The great advantage of coordinates is that all questions about intersections become questions about solutions of equations, which algebra can answer. In this case, the algebra shows that existence of the intersection depends on existence of the number $\sqrt{3}$.

In fact, we have the following theorem.

Nature of constructible points   *Points constructible by ruler and compass have coordinates obtainable from 1 by the rational operations $+$, $-$, $\times$, $\div$, and square roots.*

Proof   Recall from Section 2.2 that the given constructions are:

- To draw a straight line between any two given points.

- To draw a circle with given center and radius.

In the beginning, we are given only the unit of length, which we may take to be the line segment between $(0, 0)$ and $(1, 0)$. All points are constructed as intersections of lines and circles, so it will suffice to show the following:

1. The line through (a_1, b_1) and (a_2, b_2) has an equation with coefficients obtainable from a_1, a_2, b_1, and b_2 by rational operations.

2. The circle with center (a, b) and radius r has an equation with coefficients obtainable from a, b, and r by rational operations.

3. The intersection of two lines has coordinates obtainable from the coefficients of their equations by rational operations.

4. The intersection of a line and a circle has coordinates obtainable from the coefficients of their equations by rational operations and square roots.

5. The intersection of two circles has coordinates obtainable from the coefficients of their equations by rational operations and square roots.

These facts are confirmed by calculations like those we have already considered.

1. The line through (a_1, b_1) and (a_2, b_2) has equation

$$(b_1 - b_2)x - (a_1 - a_2)y = a_2 b_1 - a_1 b_2,$$

as may be checked by substituting the points $x = a_1$, $y = b_1$ and $x = a_2$, $y = b_2$. All the coefficients come from a_1, a_2, b_1, b_2 by rational operations.

2. The circle with center (a, b) and radius r has equation

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

as we already know.

3. The intersection of two lines

$$a_1 x + b_1 y = c_1 \quad \text{and} \quad a_2 x + b_2 y = c_2$$

is computed from a_1, b_1, c_1 and a_2, b_2, c_2 by rational operations. Just recall the usual process for solving a pair of linear equations.

4. The intersection of the line

$$a_1 x + b_1 y = c$$

with the circle

$$(x - a_2)^2 + (y - b_2)^2 = r^2$$

is found by substituting $x = (c - b_1 y)/a_1$, from the equation of the line, in the equation of the circle (unless $a_1 = 0$, in which case we substitute the similar expression for y). This gives a quadratic equation for x, the coefficients of which are rational in the coefficients of the line and the circle. The quadratic formula gives x from the new coefficients by further rational operations and (possibly) a square root. Finally, by substituting x back in the equation of the line, we obtain y by further rational operations.

5. The intersection of the two circles

$$(x - a_1)^2 + (y - b_1)^2 = r_1^2 \quad \text{and} \quad (x - a_2)^2 + (y - b_2)^2 = r_2^2$$

is found by expanding these equations to

$$x^2 + 2a_1x + a_1^2 + y^2 + 2b_1y + b_1^2 = r_1^2,$$
$$x^2 + 2a_2x + a_2^2 + y^2 + 2b_2y + b_2^2 = r_2^2,$$

and subtracting the second from the first to obtain the linear equation

$$2(a_1 - a_2)x + 2(b_1 - b_2)y = r_1^2 - r_2^2 + a_2^2 - a_1^2 + b_2^2 - b_1^2.$$

All the coefficients are rational combinations of the original coefficients, so we are now reduced to the situation just considered, the intersection of a circle and a line. It follows that the coordinates of the intersections may be found by rational operations and square roots. □

It follows from this theorem that if a number is *not* expressible by rational operations and square roots, the corresponding point is not constructible by ruler and compass. This opens the way for algebraic attack on problems of constructibility, and in this way some of the Greek problems were shown to be unsolvable, after 2000 years of unsuccessful attempts to solve them. The simplest example of a nonconstructible number is $\sqrt[3]{2}$; see the exercises.

Exercises

The theorem on the nature of constructible points also has a converse: *there is a ruler and compass construction of any point with coordinates obtainable by rational operations and square roots (of positive numbers)*.

3.2.1. Explain how to do addition and subtraction.

The keys to multiplication and division are the similar triangles in Figure 3.3.

3.2.2. Explain why the lengths are as shown in Figure 3.3.

Likewise, similar right-angled triangles are the key to constructing a square root (Figure 3.4).

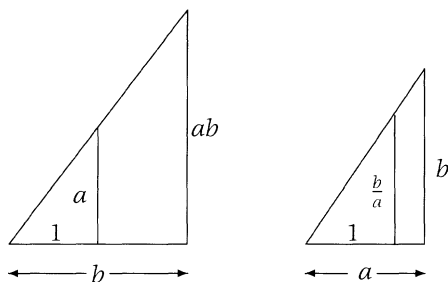3.2.3. Explain why the lengths are as shown in Figure 3.4.

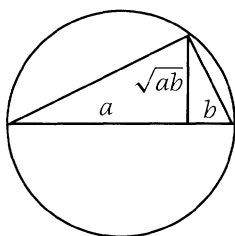**FIGURE 3.3**   Constructing the product and quotient of lengths.

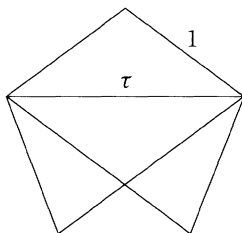**FIGURE 3.4**   Constructing the square root of a length.

FIGURE 3.5   Regular pentagon.

It follows from this converse theorem that any length expressible by rational operations and square roots is constructible by ruler and compass. This gives an easy way to see that certain figures are constructible. For example, we can say immediately that the length $\tau = \frac{1+\sqrt{5}}{2}$ is constructible. This shows that the regular pentagon is constructible, because the regular pentagon with unit sides has diagonal τ (Figure 3.5).

3.2.4.* Prove that the diagonal is τ. (It may help to use the lines shown in Figure 3.5 and find some similar triangles.)

Now let us see why the number $\sqrt[3]{2}$ is *not* constructible. An elementary proof was discovered by the number theorist Edmund Landau (1877–1938) when he was still a student. He starts with the set F_0 of rational numbers and considers successively larger sets $F_1, F_2, \dots$, each obtained from the one before by adding the square root of one of its members, and then applying rational operations. The aim is to show that $\sqrt[3]{2}$ is never reached in this way, because its presence would yield a contradiction.

3.2.5.* Let F_0 be the set of rational numbers and let $F_{k+1} = \{a + b\sqrt{c_k} : a, b \in F_k\}$ for some $c_k \in F_k$. Show that the sum, difference, product, and quotient of any members of F_{k+1} are also members. (Hence each F_k is a *field*, as defined in the exercises in Section 1.4.)

3.2.6.* Show that if $a, b, c \in F_k$ but $\sqrt{c} \notin F_k$ then $a + b\sqrt{c} = 0 \Leftrightarrow a = b = 0$.

3.2.7.* Suppose $\sqrt[3]{2} = a + b\sqrt{c}$ where $a, b, c \in F_k$, but that $\sqrt[3]{2} \notin F_k$. (We know $\sqrt[3]{2} \notin F_0$ because $\sqrt[3]{2}$ is irrational by Exercise 1.6.8.) Cube both sides and deduce that

$$2 = a^3 + 3ab^2c \quad \text{and} \quad 0 = 3a^2b + b^3c.$$

3.2.8.* Deduce from Exercise 3.2.7* that $a - b\sqrt{c} = \sqrt[3]{2}$ also, which is a contradiction.

3.3   The Real Numbers

The results of the last section throw new light on the relationship between numbers and geometry. We knew from the beginning that the rational numbers cannot represent all lengths, because irrational square roots occur in even the simplest figures. Now we know that all lengths actually needed in the geometry of straight lines and circles arise from rational operations and square roots. If we want to treat these lengths as numbers, therefore, it suffices to understand square roots.

Nevertheless, this is a hard problem. Not only do we have to understand $\sqrt{2}$, $\sqrt{3}$, and so on, but also more complicated numbers such as $\sqrt{1 + \sqrt{2}}$ and $\sqrt{\sqrt{2} + \sqrt{3}}$, because the corresponding lengths

are all constructible by Exercises 3.2.1 to 3.2.3. The simplest solution was discovered by Dedekind in 1858, when he decided that a better understanding of *all* irrational numbers was desirable. It comes from reflecting on the "position" an irrational number occupies among the rationals.

Consider $\sqrt{2}$. It is less than each of the numbers

$$2$$
$$1.5$$
$$1.42$$
$$1.415$$
$$1.4143$$
$$1.41422$$
$$1.414214$$
$$1.4142136$$
$$\cdots$$

because each of these numbers has a square greater than 2. Likewise, it is greater than each of

$$\cdots$$
$$1.4142135$$
$$1.414213$$
$$1.41421$$
$$1.4142$$
$$1.414$$
$$1.41$$
$$1.4$$
$$1$$

because each of the latter numbers has square less than 2.

These two lists of rational numbers give a rough idea of the position of $\sqrt{2}$ among all the rationals. Its exact position can be specified by the set L of *all* the positive rationals with squares < 2 (L is for "lower"), and the set U of positive rationals with squares > 2 (U is for "upper"), because there is no other number that fits between these two sets. Thus $\sqrt{2}$ can be recognized in its absence,

as it were, by the two sets into which the positive rationals separate. It dawned on Dedekind that this is all we need to know about $\sqrt{2}$; it may as well be *defined* as the pair of sets of rationals (L, U).

It takes a while to get your breath back after first seeing this idea, because everyone thinks that $\sqrt{2}$ is already "there," and we only have to compute it. But no, the real problem is to *define* it, and hence to know what it is we are computing. The beauty of Dedekind's definition is that it requires nothing new, only sets of objects already assumed to exist: the rational numbers. Since we know that $\sqrt{2}$ is not itself a rational number, this is as simple as the definition can be.

The general idea is to imagine that the rationals are separated by an irrational number, then take the separation—or *cut* as Dedekind called it—to *be* the number. This enables us to define all positive irrational numbers in one fell swoop as follows.

Definition A *positive irrational number* is a pair (L, U) of sets of positive rationals such that

- L and U together include all positive rationals.

- Each member of L is less than every member of U.

- L has no greatest member and U has no least member.

It is not necessary to stick to positive irrationals either, because the sets L and U can be taken from the set of all rationals. However, it is convenient to keep the restriction to positive rationals a little longer, as it makes it easier to define addition and (especially) multiplication.

Each rational number a also makes a cut in the set of rationals, of course, into the set $L_a = \{$rationals $< a\}$ and the set $U_a = \{$rationals $\geq a\}$. The only difference is that U_a has a least member, namely a. This prompts us to define positive *real* numbers so as to include both rationals and irrationals, by weakening the third condition in the definition of irrational number to say only that L has no greatest member.

Definition A *positive real number* is a pair (L, U) of sets of positive rationals such that

- L and U together include all positive rationals.