Exercises

An interesting application of equidistant lines is the following.

3.3.1 Show that any three points not in a line lie on a unique circle. (Hint: the center of the circle is equidistant from the three points.)

The equations of lines and circles enable us to prove many geometric theorems by algebra, as Descartes realized. In fact, they greatly expand the scope of geometry by allowing many curves to be described by equations. But algebra is also useful in proving that certain quantities are *not* equal. One example is the *triangle inequality*.

3.3.2 Consider any triangle, which for convenience we take to have one vertex at $O=(0,0)$, one at $P=(x_1,0)$ with $x_1>0$, and one at $Q=(x_2,y_2)$. Show that

$$|OP|=x_1,\quad |PQ|=\sqrt{(x_2-x_1)^2+y_2^2},\quad |OQ|=\sqrt{x_2^2+y_2^2}.$$

The triangle inequality states that $|OP|+|PQ|>|OQ|$ (any two sides of a triangle are together greater than the third). To prove this statement, it suffices to show that

$$(|OP|+|PQ|)^2>|OQ|^2.$$

3.3.3 Show that $(|OP|+|PQ|)^2-|OQ|^2=2x_1\left[\sqrt{(x_2-x_1)^2+y_2^2}-(x_2-x_1)\right]$.

3.3.4 Show that the term in square brackets in Exercise 3.3.3 is positive if $y_2\neq 0$, and hence that the triangle inequality holds in this case.

3.3.5 If $y_2=0$, why is this not a problem?

Later we will give a more sophisticated approach to the triangle inequality, which does not depend on choosing a special position for the triangle.

3.4 Intersections of lines and circles

Now that lines and circles are defined by equations, we can give exact algebraic equivalents of straightedge and compass operations:

- Drawing a line through given points corresponds to finding the equation of the line through given points (x_1,y_1) and (x_2,y_2). The slope between these two points is $\frac{y_2-y_1}{x_2-x_1}$, which must equal the slope $\frac{y-y_1}{x-x_1}$ between the general point (x,y) and the special point (x_1,y_1), so the equation is

$$\frac{y-y_1}{x-x_1}=\frac{y_2-y_1}{x_2-x_1}.$$

Multiplying both sides by $(x-x_1)(x_2-x_1)$, we get the equivalent equation

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

or

$$(y_2-y_1)x-(x_2-x_1)y-x_1y_2+y_1x_2 = 0.$$

- Drawing a circle with given center and radius corresponds to finding the equation of the circle with given center (a,b) and given radius r, which is
$$(x-a)^2+(y-b)^2 = r^2.$$
- Finding new points as intersections of previously drawn lines and circles corresponds to finding the solution points of
  - a pair of equations of lines,
  - a pair of equations of circles,
  - the equation of a line and the equation of a circle.

For example, to find the intersection of the two circles

$$(x-a_1)^2+(y-b_1)^2 = r_1^2$$

and

$$(x-a_2)^2+(y-b_2)^2 = r_2^2,$$

we expand the equations of the circles as

$$x^2-2a_1x+a_1^2+y^2-2b_1y+b_1^2-r_1^2 = 0, \tag{1}$$
$$x^2-2a_2x+a_2^2+y^2-2b_2y+b_2^2-r_2^2 = 0, \tag{2}$$

and subtract Equation (2) from Equation (1). The x^2 and y^2 terms cancel, and we are left with the linear equation in x and y:

$$2(a_2-a_1)x+2(b_2-b_2)y+r_2^2-r_1^2 = 0. \tag{3}$$

We can solve Equation (3) for either x or y. Then substituting the solution of (3) in (1) gives a quadratic equation for either y or x. If the equation is of the form $Ax^2+Bx+C=0$, then we know that the solutions are

$$x = \frac{-B \pm \sqrt{B^2-4AC}}{2A}.$$

Solving linear equations requires only the operations $+,-,\times$, and $\div$, and the quadratic formula shows that $\sqrt{\ }$ is the only additional operation needed to solve quadratic equations.

Thus, all intersection points involved in a straightedge and compass construction can be found with the operations $+,-,\times,\div$, and $\sqrt{\ }$.

Now recall from Chapters 1 and 2 that the operations $+,-,\times,\div$, and $\sqrt{\ }$ can be carried out by straightedge and compass. Hence, we get the following result:

Algebraic criterion for constructibility. *A point is constructible (starting from the points 0 and 1) if and only if its coordinates are obtainable from the number 1 by the operations* $+,-,\times,\div$, *and* $\sqrt{\ }$.

The algebraic criterion for constructibility was discovered by Descartes, and its greatest virtue is that it enables us to prove that certain figures or points are *not* constructible. For example, one can prove that the number $\sqrt[3]{2}$ is not constructible by showing that it cannot be expressed by a finite number of square roots, and one can prove that the angle $\pi/3$ cannot be trisected by showing that $\cos\frac{\pi}{9}$ also cannot be expressed by a finite number of square roots. These results were not proved until the 19th century, by Pierre Wantzel. Rather sophisticated algebra is required, because one has to go beyond Descartes' concept of constructibility to survey the *totality* of constructible numbers.

Exercises

3.4.1 Find the intersections of the circles $x^2+y^2=1$ and $(x-1)^2+(y-2)^2=4$.

3.4.2 Check the plausibility of your answer to Exercise 3.4.1 by a sketch of the two circles.

3.4.3 The line $x+2y-1=0$ found by eliminating the x^2 and y^2 from the equations of the circles should have some geometric meaning. What is it?

3.5 Angle and slope

The concept of distance is easy to handle in coordinate geometry because the distance between points (x_1,y_1) and (x_2,y_2) is an algebraic function of their coordinates, namely

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}.$$

This is *not* the case for the concept of angle. The angle θ between a line $y = tx$ and the x-axis is $\tan^{-1} t$, and the function $\tan^{-1} t$ is not an algebraic function. Nor is its inverse function $t = \tan\theta$ or the related functions $\sin\theta$ (sine) and $\cos\theta$ (cosine).

To stay within the world of algebra, we have to work with the slope t rather than the angle θ. Lines make the same angle with the x-axis if they have the same slope, but to test equality of angles in general we need the concept of *relative slope*: If line $\mathscr{L}_1$ has slope t_1 and line $\mathscr{L}_2$ has slope t_2, then the *slope of* $\mathscr{L}_1$ *relative to* $\mathscr{L}_2$ is defined to be

$$\pm\left|\frac{t_1 - t_2}{1 + t_1 t_2}\right|.$$

This awkward definition comes from the formula you have probably seen in trigonometry,

$$\tan(\theta_1 - \theta_2) = \frac{\tan\theta_1 - \tan\theta_2}{1 + \tan\theta_1 \tan\theta_2},$$

by taking $t_1 = \tan\theta_1$ and $t_2 = \tan\theta_2$. The reason for the $\pm$ sign and the absolute value is that the slopes t_1, t_2 alone do not specify an angle—they specify only a pair of lines and hence a pair of angles that add to a straight angle. (For more on using relative slope to discuss equality of angles, see Hartshorne's *Geometry: Euclid and Beyond*, particularly pp. 141–155.)

At any rate, with some care it is possible to use the concept of relative slope to test algebraically whether angles are equal. The concept also makes it possible to state the SAS and ASA axioms in coordinate geometry, and to verify that all of Euclid's and Hilbert's axioms hold. We omit the details because they are laborious, and because we can approach SAS and ASA differently now that we have coordinates. Specifically, *it becomes possible to define the concept of "motion" that Euclid appealed to in his proof of SAS!* This will be done in the next section.

Exercises

The most useful instance of relative slope is where the lines are perpendicular.

3.5.1 Show that lines of slopes t_1 and t_2 are perpendicular just in case $t_1 t_2 = -1$.

3.5.2 Use the condition for perpendicularity found in Exercise 3.5.1 to show that the line from $(1,0)$ to $(3,4)$ is perpendicular to the line from $(0,2)$ to $(4,0)$.

In the next section, we will define a *rotation about* O to be a transformation $r_{c,s}$ of $\mathbb{R}^2$ depending on two real numbers c and s such that $c^2+s^2=1$. The transformation $r_{c,s}$ sends the point (x,y) to the point $(cx-sy, sx+cy)$. It will be explained in the next section why it is reasonable to call this a "rotation about O," and why $c=\cos\theta$ and $s=\sin\theta$, where θ is the angle of rotation.

For the moment, suppose that this is the case, and consider the effect of two rotations r_{c_1,s_1} and r_{c_2,s_2}, where

$$c_1=\cos\theta_1, \quad s_1=\sin\theta_1; \qquad c_2=\cos\theta_2, \quad s_2=\sin\theta_2.$$

This thought experiment leads us to proofs of the formulas for cos, sin, and tan of $\theta_1+\theta_2$:

3.5.3 Show that the outcome of r_{c_1,s_1} and r_{c_2,s_2} is to send (x,y) to

$$((c_1c_2-s_1s_2)x-(s_1c_2+c_1s_2)y, (s_1c_2+c_1s_2)x+(c_1c_2-s_1s_2)y).$$

3.5.4 Assuming that r_{c_1,s_1} really is a rotation about O through angle θ_1, and r_{c_2,s_2} really is a rotation about O through angle θ_2, deduce from Exercise 3.5.3 that

$$\begin{aligned}\cos(\theta_1+\theta_2)&=\cos\theta_1\cos\theta_2-\sin\theta_1\sin\theta_2,\\ \sin(\theta_1+\theta_2)&=\sin\theta_1\cos\theta_2+\cos\theta_1\sin\theta_2.\end{aligned}$$

3.5.5 Deduce from Exercise 3.5.4 that

$$\tan(\theta_1+\theta_2)=\frac{\tan\theta_1+\tan\theta_2}{1-\tan\theta_1\tan\theta_2},$$

hence

$$\tan(\theta_1-\theta_2)=\frac{\tan\theta_1-\tan\theta_2}{1+\tan\theta_1\tan\theta_2}.$$

3.6 Isometries

A possible weakness of our model of the plane is that it seems to single out a particular point (the origin O) and particular lines (the x- and y-axes). In Euclid's plane, each point is like any other point and each line is like any other line. We can overcome the apparent bias of $\mathbb{R}^2$ by considering *transformations* that allow any point to become the origin and any line to become the x-axis. As a bonus, this idea gives meaning to the idea of "motion" that Euclid tried to use in his attempt to prove SAS.

A transformation of the plane is simply a function $f:\mathbb{R}^2\to\mathbb{R}^2$, in other words, a function that sends points to points.

A transformation f is called an *isometry* (from the Greek for "same length") if it sends any two points, P_1 and P_2, to points $f(P_1)$ and $f(P_2)$ the same distance apart. Thus, an isometry is a function f with the property

$$|f(P_1)f(P_2)| = |P_1P_2|$$

for any two points P_1, P_2. Intuitively speaking, an isometry "moves the plane rigidly" because it preserves the distance between points. There are many isometries of the plane, but they can be divided into a few simple and obvious types. We show examples of each type below, and, in the next section, we explain why only these types exist.

You will notice that certain isometries (translations and rotations) make it possible to move the origin to any point in the plane and the x-axis to any line. Thus, $\mathbb{R}^2$ is really like Euclid's plane, in the sense that each point is like any other point and each line is like any other line. This property entitles us to choose axes wherever it is convenient. For example, we are entitled to prove the triangle inequality, as suggested in the Exercises to Section 3.3, by choosing one vertex of the triangle at O and another on the positive x-axis.

Translations

A translation moves each point of the plane the same distance in the same direction. Each translation depends on two constants a and b, so we denote it by $t_{a,b}$. It sends each point (x,y) to the point $(x+a, y+b)$. It is obvious that a translation preserves the distance between any two points, but it is worth checking this formally—so as to know what to do in less obvious cases.

So let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$. It follows that

$$t_{a,b}(P_1) = (x_1 + a, y_1 + b), \quad t_{a,b}(P_2) = (x_2 + a, y_2 + b)$$

and therefore,

$$\begin{aligned} |t_{a,b}(P_1)t_{a,b}(P_2)| &= \sqrt{(x_2+a-x_1-a)^2 + (y_2+b-y_1-b)^2} \\ &= \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2} \\ &= |P_1P_2|, \quad \text{as required.} \end{aligned}$$

Rotations

We think of a rotation as something involving an angle θ, but, as mentioned in the previous section, it is more convenient to work algebraically with $\cos\theta$ and $\sin\theta$. These are simply two numbers c and s such that $c^2+s^2=1$, so we will denote a rotation of the plane about the origin by $r_{c,s}$.

The rotation $r_{c,s}$ sends the point (x,y) to the point $(cx-sy, sx+cy)$. It is not obvious why this transformation should be called a rotation, but it becomes clearer after we check that $r_{c,s}$ preserves lengths.

If we let $P_1=(x_1,y_1)$ and $P_2=(x_2,y_2)$ again, it follows that

$$r_{c,s}(P_1)=(cx_1-sy_1, sx_1+cy_1), \quad r_{c,s}(P_2)=(cx_2-sy_2, sx_2+cy_2)$$

and therefore,

$$\begin{aligned}
|r_{c,s}(P_1)r_{c,s}(P_2)| &= \sqrt{[c(x_2-x_1)-s(y_2-y_1)]^2+[s(x_2-x_1)+c(y_2-y_1)]^2} \\
&= \sqrt{\begin{array}{c} c^2(x_2-x_1)^2-2cs(x_2-x_1)(y_2-y_1)+s^2(y_2-y_1)^2 \\ +s^2(x_2-x_1)^2+2cs(x_2-x_1)(y_2-y_1)+c^2(y_2-y_1)^2 \end{array}} \\
&= \sqrt{(c^2+s^2)(x_2-x_1)^2+(c^2+s^2)(y_2-y_1)^2} \\
&= \sqrt{(x_2-x_1)^2+(y_2-y_1)^2} \quad \text{because} \quad c^2+s^2=1 \\
&= |P_1P_2|.
\end{aligned}$$

Thus, $r_{c,s}$ preserves lengths. Also, $r_{c,s}$ sends $(0,0)$ to itself, and it moves $(1,0)$ to (c,s) and $(0,1)$ to $(-s,c)$, *which is exactly what rotation about O through angle θ does* (see Figure 3.5). We will see in the next section that only one isometry of the plane moves these three points in this manner.

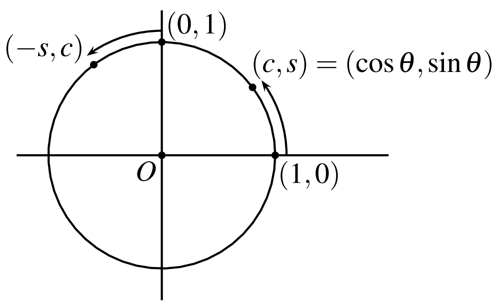

Figure 3.5: Movement of points by a rotation

Reflections

The easiest reflection to describe is *reflection in the x-axis*, which sends $P=(x,y)$ to $\overline{P}=(x,-y)$. Again it is obvious that this is an isometry, but we can check by calculating the distance between reflected points $\overline{P_1}$ and $\overline{P_2}$ (Exercise 3.6.1).

We can reflect the plane in any line, and we can do this by combining reflection in the x-axis with translations and rotations. For example, reflection in the line $y=1$ (which is parallel to the x-axis) is the result of the following three isometries:

- $t_{0,-1}$, a translation that moves the line $y=1$ to the x-axis,
- reflection in the x-axis,
- $t_{0,1}$, which moves the x-axis back to the line $y=1$.

In general, we can do a reflection in any line $\mathscr{L}$ by moving $\mathscr{L}$ to the x-axis by some combination of translation and rotation, reflecting in the x-axis, and then moving the x-axis back to $\mathscr{L}$.

Reflections are the most fundamental isometries, because any isometry is a combination of them, as we will see in the next section. In particular, any translation is a combination of two reflections, and any rotation is a combination of two reflections (see Exercises 3.6.2–3.6.4).

Glide reflections

A glide reflection is the result of a reflection followed by a translation in the direction of the line of reflection. For example, if we reflect in the x-axis, sending (x,y) to $(x,-y)$, and follow this with the translation $t_{1,0}$ of length 1 in the x-direction, then (x,y) ends up at $(x+1,-y)$.

A glide reflection with nonzero translation length is different from the three types of isometry previously considered.

- It is not a translation, because a translation maps any line in the direction of translation into itself, whereas a glide reflection maps only one line into itself (namely, the line of reflection).
- It is not a rotation, because a rotation has a fixed point and a glide reflection does not.
- It is not a reflection, because a reflection also has fixed points (all points on the line of reflection).

Exercises

3.6.1 Check that reflection in the x-axis preserves the distance between any two points.

When we combine reflections in two lines, the nature of the outcome depends on whether the lines are parallel.

3.6.2 Reflect the plane in the x-axis, and then in the line $y = 1/2$. Show that the resulting isometry sends (x,y) to $(x,y+1)$, so it is the translation $t_{0,1}$.

3.6.3 Generalize the idea of Exercise 3.6.2 to show that the combination of reflections in parallel lines, distance $d/2$ apart, is a translation through distance d, in the direction perpendicular to the lines of reflection.

3.6.4 Show, by a suitable picture, that the combination of reflections in lines that meet at angle $\theta/2$ is a rotation through angle θ, about the point of intersection of the lines.

Another way to put the result of Exercise 3.6.4 is as follows: Reflections in *any* two lines meeting at the same angle $\theta/2$ at the same point P give the same outcome. This observation is important for the next three exercises (where pictures will also be helpful).

3.6.5 Show that reflections in lines $\mathscr{L}$, $\mathscr{M}$, and $\mathscr{N}$ (in that order) have the same outcome as reflections in lines $\mathscr{L}'$, $\mathscr{M}'$, and $\mathscr{N}$, where $\mathscr{M}'$ is perpendicular to $\mathscr{N}$.

3.6.6 Next show that reflections in lines $\mathscr{L}'$, $\mathscr{M}'$, and $\mathscr{N}$ have the same outcome as reflections in lines $\mathscr{L}'$, $\mathscr{M}''$, and $\mathscr{N}'$, where $\mathscr{M}''$ is parallel to $\mathscr{L}'$ and $\mathscr{N}'$ is perpendicular to $\mathscr{M}''$.

3.6.7 Deduce from Exercise 3.6.6 that the combination of any three reflections is a glide reflection.

3.7 The three reflections theorem

We saw in Section 3.3 that the points equidistant from two points A and B form a line, which implies that isometries of the plane are very simple: *An isometry f of $\mathbb{R}^2$ is determined by the images $f(A), f(B), f(C)$ of three points A,B,C not in a line.*

The proof follows from three simple observations:

- Any point P in $\mathbb{R}^2$ is determined by its distances from A,B,C. Because if Q is another point with the same distances from A,B,C as P, then A,B,C lie in the equidistant line of P and Q, contrary to the assumption that A,B,C are *not* in a line.

- The isometry f preserves distances (by definition of isometry), so $f(P)$ lies at the same respective distances from $f(A), f(B), f(C)$ as P does from A, B, C.

- There is only one point at given distances from $f(A), f(B), f(C)$ because these three points are not in a line—in fact they form a triangle congruent to triangle ABC, because f preserves distances.

Thus, the image $f(P)$ of any point P—and hence the whole isometry f—is determined by the images of three points A, B, C not in a line. □

This "three point determination theorem" gives us the:

Three reflections theorem. *Any isometry of* $\mathbb{R}^2$ *is a combination of one, two, or three reflections.*

Given an isometry f, we choose three points A, B, C not in a line, and we look for a combination of reflections that sends A to $f(A)$, B to $f(B)$, and C to $f(C)$. Such a combination is necessarily equal to f. We can certainly send A to $f(A)$ by reflection in the equidistant line of A and $f(A)$. Call this reflection r_A.

Now r_A sends B to $r_A(B)$, so if $r_A(B) = f(B)$ we need to do nothing more for B.

If $r_A(B) \neq f(B)$, we can send $r_A(B)$ to $f(B)$ by reflection r_B in the equidistant line of $r_A(B)$ and $f(B)$. Fortunately, $f(A) = r_A(A)$ lies on this line, because the distance from $f(A)$ to $f(B)$ equals the distance from $r_A(A)$ to $r_A(B)$ (because f and r_A are isometries). Thus, r_B does not move $f(A)$, and the *combination* of r_A followed by r_B sends A to $f(A)$ and B to $f(B)$.

The argument is similar for C. If C has already been sent to $f(C)$, we are done. If not, we reflect in the line equidistant from $f(C)$ and the point where C has been sent so far. It turns out (by a check of equal distances like that made for $f(A)$ above) that $f(A)$ and $f(B)$ already lie on this line, so they are not moved. Thus, we finally have a combination of no more than three reflections that moves A to $f(A)$, B to $f(B)$, and C to $f(C)$, as required. □

Now of course, one reflection is a reflection, and we found in the previous exercise set that combinations of two reflections are translations and rotations, and that combinations of three reflections are glide reflections (which include reflections). Thus, *an isometry of* $\mathbb{R}^2$ *is either a translation, a rotation, or a glide reflection.*