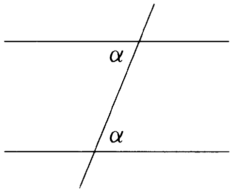

FIGURE 3.7 Alternate angles.

3.5.5. Deduce from Exercise 3.5.4 a proof that the angle sum of any triangle is two right angles.

3.6 Isometries of the Euclidean Plane

Isometries of the Euclidean plane are actually not much more complicated than the example in the previous section, though this is not clear from the bare definition. The situation is crucially simplified by the following.

Basic property of isometries *An isometry is determined by the images of three points not in a line.*

Proof This theorem is based on the third basic property of distance, that each point is determined by its distances from three points not in a line, but first we have to show that the image of any isometry includes three such points.

Three particular points not in a line are the vertices A, B, C of an equilateral triangle. C is not in the line AB because AB is the equidistant set of two points C and D (Figure 3.8), and C is certainly not equidistant from C and D. Notice that this argument depends only on the distances between A, B, C, and D, and these distances

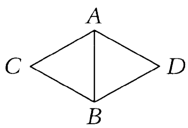

FIGURE 3.8 Points not in a line.

are preserved by any isometry. Hence if f is any isometry, and A, B, C are the vertices of an equilateral triangle, then $f(A)$, $f(B)$, $f(C)$ are three points not in a line.

If Q is any point, we know that Q is determined by its distances from A, B, and C. Its image $f(Q)$ has the same distances from $f(A)$, $f(B)$, and $f(C)$, respectively, and because $f(A)$, $f(B)$ and $f(C)$ are not in a line, there is only one point with the same distances from them as $f(Q)$. Thus any isometry that agrees with f on A, B, and C agrees with f on any point Q. □

This theorem will allow us to express any isometry f of the Euclidean plane as a composite of simple isometries, if only we can find enough isometries to move A, B, C to $f(A)$, $f(B)$, $f(C)$, respectively. The most convenient isometries for this purpose arise from the fact that each line is the equidistant set of two points. The calculations in the previous section show that the line $ax+by=c$ is the equidistant set of any two points (x_1, y_1), (x_2, y_2) that satisfy the relation

$$x_2 = x_1 + \frac{k}{2}a, \qquad y_2 = y_1 + \frac{k}{2}b,$$
$$\text{where} \qquad k = \frac{4(c - ax_1 - by_1)}{a^2 + b^2}.$$

Because these points can be regarded as "mirror images" in the line $ax + by = c$, it is reasonable to make the following definition.

Definition *Reflection in the line* $ax + by = c$ is the map that sends each point (x_1, y_1) to the point (x_2, y_2) defined by

$$x_2 = x_1 + \frac{2a(c - ax_1 - by_1)}{a^2 + b^2},$$
$$y_2 = y_1 + \frac{2b(c - ax_1 - by_1)}{a^2 + b^2}.$$

It follows that any two points can be exchanged by reflection in their equidistant line, but we have to check that reflection in $ax+by=c$ is an isometry. This is easier if we first arrange that $a^2 + b^2 = 1$, which can always be done because an equivalent equation is obtained if a, b, and c are multiplied by any nonzero constant. The reflection that sends (x_1, y_1) to (x_2, y_2) is then expressed by the equations

$$x_2 = x_1 + 2a(c - ax_1 - by_1) = x_1(1 - 2a^2) - 2aby_1 + 2ac,$$
$$y_2 = y_1 + 2b(c - ax_1 - by_1) = y_1(1 - 2b^2) - 2abx_1 + 2bc.$$

Let $P_1 = (x_1, y_1)$ and $P_1' = (x_1', y_1')$ be any two points, and consider the square of the distance between their reflections P_2 and P_2' in $ax + by = c$:

$$\left[(x_1' - x_1)(1 - 2a^2) - (y_1' - y_1)2ab\right]^2$$
$$+ \left[(y_1' - y_1)(1 - 2b^2) - (x_1' - x_1)2ab\right]^2.$$

Expanding the two main terms, one finds that the coefficients of $(x_1' - x_1)^2$, $(x_1' - x_1)(y_1' - y_1)$, and $(y_1' - y_1)^2$ are 1, 0, and 1, respectively, because $a^2 + b^2 = 1$. Hence $d(P_2, P_2') = d(P_1, P_1')$ as required. □

The work involved in formalizing the intuitively simple idea of reflection in a line is worthwhile, because it gives us *all* isometries of the Euclidean plane. We get them as *composites* of reflections, that is, as the result of successive reflections.

Three reflections theorem *Each isometry of the Euclidean plane is the composite of one, two, or three reflections.*

Proof Suppose that P_1, P_2, and P_3 are three points not in a line, and f is any isometry. By the basic property of isometries, it suffices to find a composite of one, two, or three reflections that send P_1, P_2, and P_3 to $f(P_1)$, $f(P_2)$, and $f(P_3)$ respectively, because the latter isometry necessarily coincides with f. This can be done with the following reflections f_1, f_2, and f_3.

1. Let f_1 be reflection in the equidistant line of P_1 and $f(P_1)$. It sends P_1 to $f(P_1)$ (by definition of reflection), P_2 to $f_1(P_2)$, and P_3 to $f_1(P_3)$.

2. If $f_1(P_2) \neq f(P_2)$, let f_2 be reflection in the equidistant line of $f_1(P_2)$ and $f(P_2)$. Then f_2 sends $f_1(P_2)$ to $f(P_2)$, as required. Also $f_1(P_1) = f(P_1)$ is equidistant from $f_1(P_2)$ and $f(P_2)$ (namely, at the distance between P_1 and P_2), hence it is fixed by the reflection f_2.

3. We now have $f_2f_1(P_1) = f(P_1)$ and $f_2f_1(P_2) = f(P_2)$. If $f_2f_1(P_3) \neq f(P_3)$ we send $f_2f_1(P_3)$ to $f(P_3)$ by reflection f_3 in their equidistant line. Again, $f(P_1) = f_2f_1(P_1)$ is equidistant from $f(P_3)$ and $f_2f_1(P_3)$, hence it is fixed by f_3. So is $f(P_2) = f_2f_1(P_2)$.

It follows that P_1, P_2, and P_3 are sent to $f(P_1)$, $f(P_2)$, and $f(P_3)$, respectively, by either f_1, f_2f_1, or $f_3f_2f_1$. Thus one of these composites of reflections is the required isometry. □

While this proof is in front of you, two aspects of terminology and notation should be pointed out.

- We speak of the "composite" of reflections because reflections are functions, and taking a "function of a function" is called *composition*. The notation for composition of functions (for example, $f_2f_1(P_1)$) is the usual product notation, but we prefer not to call this a "product" because it does not have all the properties of other products.
- In particular, the composite f_2f_1 is not necessarily the same as f_1f_2. For example, let f_1 be reflection in the x-axis and let f_2 be reflection in the line $x = y$; then f_2f_1 is a quarter turn anticlockwise and f_1f_2 is a quarter turn clockwise. So don't forget that f_2f_1 means "f_1 first, then f_2."

Exercises

Several general properties of isometries follow from the three reflections theorem, because they are easily proved for reflections.

3.6.1. Show that the following are true of reflections, and hence of all isometries.

1. They are invertible functions, and their inverses are also isometries.
2. They map lines to lines.

The three reflections theorem should also account for the isometries we already know from Section 3.5, the half turn about O and the translations.

3.6.2. Show that a half turn about O is the composite of reflections in OX and OY.

3.6.3. Show that the translation $\mathrm{tran}_{a,b}$ is the composite of reflection in the equidistant line of O and (a, b) and reflection in the parallel to this line through (a, b).

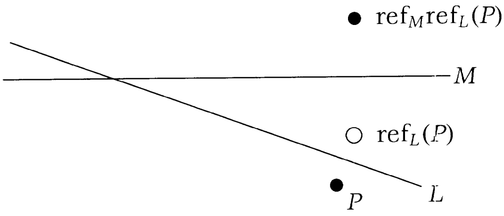

FIGURE 3.9 Rotation via reflections in intersecting lines.

The half turn is an example of a *rotation*, which is defined in general to be the composite of reflections in intersecting lines. Because reflection in a line L leaves all points of L fixed (this is clear from the defining formulas), the composite of reflections in lines L and M leaves the intersection of L and M fixed. A picture also suggests that this rotation moves any other point through *twice* the angle between L and M (Figure 3.9).

3.6.4. Show that the composite ref_Mref_L of reflections ref_L in L and ref_M in M moves the line L through twice the angle between L and M.

3.6.5. Show that the rotation about O obtained by successive reflections in OX and the line $y = mx$ sends (x, y) to $(\frac{1-m^2}{1+m^2}x - \frac{2m}{1+m^2}y, \frac{2m}{1+m^2}x + \frac{1-m^2}{1+m^2}y)$.

You may recognize this as the standard formula for "rotation through angle θ," where $m = \tan\frac{\theta}{2}$, $(1 - m^2)/(1 + m^2) = \cos\theta$, and $2m/(1 + m^2) = \sin\theta$. The same formulas will recur when we study rational points on the circle in the next chapter.

There is one more type of Euclidean isometry that is not a reflection, translation, or rotation. It is called a *glide reflection*, and it is the composite of a translation with a reflection in a line parallel to the direction of translation.

3.6.6.* Show that any composite of three reflections is a glide reflection.

3.7 The Triangle Inequality

Another crucial property of distance is the so-called triangle inequality: *if A, B, and C are three points not in a line then*

$$d(A, C) < d(A, B) + d(B, C).$$

This property is so obvious that even a dog knows it; a dog will go straight from A to a bone at C rather than go via B. However, in mathematics, even obvious statements need not be accepted if they can be proven. It so happens that the triangle inequality *can* be proved from the standard assumptions of geometry, though perhaps not as easily as one would like.

Euclid arrives at the triangle inequality only in his Proposition 20 (*Elements*, Book I), and it depends on most of his earlier propositions. A proof in our setup is also not obvious, but it takes only a few lines of algebra. We can simplify the calculation by applying isometries to move triangle ABC to a convenient position. First we apply a translation to move B to the origin. Then, if A is not already on the x-axis, we exchange it with the point A' on the x-axis such that $d(A', B) = d(A, B)$, by reflecting in their equidistant line.

The result is a triangle with coordinates of the form

$$A = (a_1, 0), \quad B = (0, 0), \quad C = (c_1, c_2),$$

with $c_2 > 0$ if the three points are not in a line. The required inequality

$$d(A, C) < d(A, B) + d(B, C)$$

then takes the form

$$\sqrt{(c_1 - a_1)^2 + c_2^2} < a_1 + \sqrt{c_1^2 + c_2^2},$$

and it is true because

$$\begin{aligned}
(\text{RHS})^2 - (\text{LHS})^2 &= a_1^2 + 2a_1\sqrt{c_1^2 + c_2^2} + c_1^2 + c_2^2 - (c_1 - a_1)^2 - c_2^2 \\
&= 2a_1\sqrt{c_1^2 + c_2^2} - 2a_1c_1 \\
&> 0,
\end{aligned}$$

because $c_2 > 0$ and therefore $\sqrt{c_1^2 + c_2^2} > c_1$. □

This calculation also shows that $d(A, C) = d(A, B) + d(B, C)$ only when $c_2 = 0$, that is, when the three points are a line.

One reason Euclid's proof of the triangle inequality is longer than ours is that he assumes less. He proves it without assuming the parallel axiom, so his argument also applies to the geometry of the non-Euclidean plane (Section 3.9*).

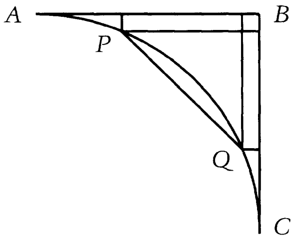

FIGURE 3.10 Bounding the length of polygons in the circle.

Exercises

A less formal way to express the triangle inequality is by the old saying "a straight line is the shortest distance between two points."

3.7.1. Use the triangle inequality and induction to prove that a line segment AB is the shortest *polygonal path* from A to B. (A polygonal path is a sequence of line segments $A_1A_2, A_2A_3, \ldots, A_{n-1}A_n$.)

It follows from this that the line segment is also the shortest *curve* from A to B, because we define the length of a curve K to be the least upper bound of lengths of polygonal paths from A to B with their vertices on K, provided the least upper bound exists.

For certain curves, such as the circle, we can also prove the existence of an upper bound by the triangle inequality. It then follows that the least upper bound exists, by the completeness of the real numbers (Exercises 3.4.1 and 3.4.2).

3.7.2. Deduce from Figure 3.10 and the triangle inequality that any polygonal path with vertices on the quarter circle from A to C has total length $< d(A,B) + d(B,C)$.

3.8* Klein's Definition of Geometry

The history of geometry is a story with a shifting point of view. In Euclid's time the raw materials of geometry were points, lines, and planes, and theorems were proved from visual axioms with the help of constructions and the vague idea of "movement." Numbers had a very limited role, because irrational lengths were not believed to correspond to numbers. In the 17th century, Fermat and Descartes