# *Introduction to*

# GEOMETRY

*second edition*

**H. S. M. COXETER, F. R. S.**

Professor of Mathematics

University of Toronto

**JOHN WILEY & SONS, INC.**

New York · London · Sydney · Toronto

# Contents

## *Part I*

# Part II

## Part III

# 6

# Circles and spheres

The present chapter shows how Euclidean geometry, in which lines and planes play a fundamental role, can be extended to *inversive* geometry, in which this role is taken over by circles and spheres. We shall see how the obvious statement, that lines and planes are circles and spheres of infinite radius, can be replaced by the sophisticated statement that lines and planes are those circles and spheres which pass through an "ideal" point, called "the point at infinity." In § 6.9 we shall briefly discuss a still more unusual geometry, called *elliptic*, which is one of the celebrated "non-Euclidean" geometries.

## 6.1 INVERSION IN A CIRCLE

*Can it be that all the great scientists of the past were really playing a game, a game in which the rules are written not by man but by God? . . . When we play, we do not ask why we are playing—we just play. Play serves no moral code except that strange code which, for some unknown reason, imposes itself on the play. . . . You will search in vain through scientific literature for hints of motivation. And as for the strange moral code observed by scientists, what could be stranger than an abstract regard for truth in a world which is full of concealment, deception, and taboos? . . . In submitting to your consideration the idea that the human mind is at its best when playing, I am myself playing, and that makes me feel that what I am saying may have in it an element of truth.*

J. L. Synge (1897 - )*

All the transformations so far discussed have been similarities, which transform straight lines into straight lines and angles into equal angles. The transformation called *inversion*, which was invented by L. J. Magnus in 1831, is new in one respect but familiar in another: it transforms some

* *Hermathena*, **19** (1958), p. 40; quoted with the editor's permission.

straight lines into circles, but it still transforms angles into equal angles. Like the reflection and the half-turn, it is involutory (that is, of period 2). Like the reflection, it has infinitely many invariant points; these do not lie on a straight line but on a circle, and the center of the circle is "singular:" it has no image!

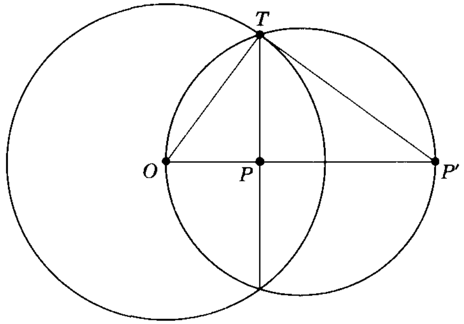



**Figure 6.1a**

Given a fixed circle with center $O$ and radius $k$, we define the *inverse* of any point $P$ (distinct from $O$) to be the point $P'$ on the ray $OP$ whose distance from $O$ satisfies the equation

$$OP \times OP' = k^2.$$

It follows from this definition that the inverse of $P'$ is $P$ itself. Moreover, every point outside the circle of inversion is transformed into a point inside, and every point inside (except the center $O$) into a point outside. The circle is invariant in the strict sense that every point on it is invariant. Every line through $O$ is invariant as a whole, but not point by point.

To construct the inverse of a given point $P$ (other than $O$) inside the circle of inversion, let $T$ be one end of the chord through $P$ perpendicular to $OP$, as in Figure 6.1*a*. Then the tangent at $T$ meets $OP$ (extended) in the desired point $P'$. For, since the right-angled triangles $OPT$, $OTP'$ are similar, and $OT = k$,

$$\frac{OP}{k} = \frac{k}{OP'}.$$

To construct the inverse of a given point $P'$ outside the circle of inversion, let $T$ be one of the points of intersection of this circle with the circle on $OP'$ as diameter (Figure 6.1*a*). Then the desired point $P$ is the foot of the perpendicular from $T$ to $OP'$.

If $OP > \frac{1}{2}k$, the inverse of $P$ can easily be constructed by the use of compasses alone, without a ruler. To do so, let the circle through $O$ with center $P$ cut the circle of inversion in $Q$ and $Q'$. Then $P'$ is the second intersection of the circles through $O$ with centers $Q$ and $Q'$. (This is easily seen by considering the similar isosceles triangles $POQ$, $QOP'$.)

There is an interesting connection between inversion and dilatation:

**6.11** *The product of inversions in two concentric circles with radii $k$ and $k'$ is the dilatation $O(\mu)$ where $\mu = (k'/k)^2$.*

To prove this, we observe that this product transforms $P$ into $P''$ (on $OP$) where

$$OP \times OP' = k^2, \qquad OP' \times OP'' = k'^2$$

and therefore

$$\frac{OP''}{OP} = \left(\frac{k'}{k}\right)^2.$$

**EXERCISES**

1. Using compasses alone, construct the vertices of a regular hexagon.

2. Using compasses alone, locate a point $B$ so that the segment $OB$ is twice as long as a given segment $OA$.

3. Using compasses alone, construct the inverse of a point distant $\frac{1}{2}k$ from the center $O$ of the circle of inversion. Describe a procedure for inverting points arbitrarily near to $O$.

4. Using compasses alone, bisect a given segment.

5. Using compasses alone, trisect a given segment. Describe a procedure for dividing a segment into any given number of equal parts.

*Note.* The above problems belong to the Geometry of Compasses, which was developed independently by G. Mohr in Denmark (1672) and L. Mascheroni in Italy (1797). For a concise version of the whole story, see Pedoe [**1**, pp. 23–25] or Courant and Robbins [**1**, pp. 145–151].

## 6.2 ORTHOGONAL CIRCLES

*A circle is a happy thing to be—*
*Think how the joyful perpendicular*
*Erected at the kiss of tangency*
*Must meet my central point, my avatar.*
*And lovely as I am, yet only 3*
*Points are needed to determine me.*

Christopher Morley (1890- )

Two circles are said to be *orthogonal* if they cut at right angles, that is, if they intersect in two points at either of which the radius of each is a tangent to the other (Figure 6.2*a*).

By Euclid III.36 (see p. 8) any circle, through a pair of inverse points is invariant: the circle of inversion decomposes it into two arcs which invert into each other. Moreover, such a circle is orthogonal to the circle of inversion, and every circle orthogonal to the circle of inversion is invariant in this sense. Through a pair of inverse points we can draw a whole *pencil* of circles (infinitely many), and they are all orthogonal to the circle of inversion. Hence

**6.21** *The inverse of a given point P is the second intersection of any two circles through P orthogonal to the circle of inversion.*

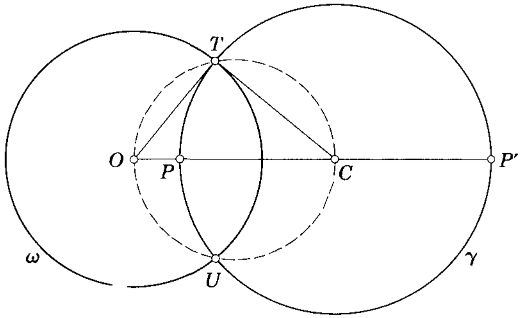



**Figure 6.2a**

The above remarks provide a simple solution for the problem of drawing, through a given point $P$, a circle (or line) orthogonal to two given circles. Let $P_1$, $P_2$ be the inverses of $P$ in the two circles. Then the circle $PP_1P_2$ (or the line through these three points, if they happen to be collinear) is orthogonal to the two given circles.

If $O$ and $C$ are the centers of two orthogonal circles $\omega$ and $\gamma$, as in Figure 6.2*a*, the circle on $OC$ as diameter passes through the points of intersection $T$, $U$. Every other point on this circle is inside one of the two orthogonal circles and outside the other. It follows that, if $a$ and $b$ are two perpendicular lines through $O$ and $C$ respectively, either $a$ touches $\gamma$ and $b$ touches $\omega$, or $a$ cuts $\gamma$ and $b$ lies outside $\omega$, or $a$ lies outside $\gamma$ and $b$ cuts $\omega$.

## 6.3 INVERSION OF LINES AND CIRCLES

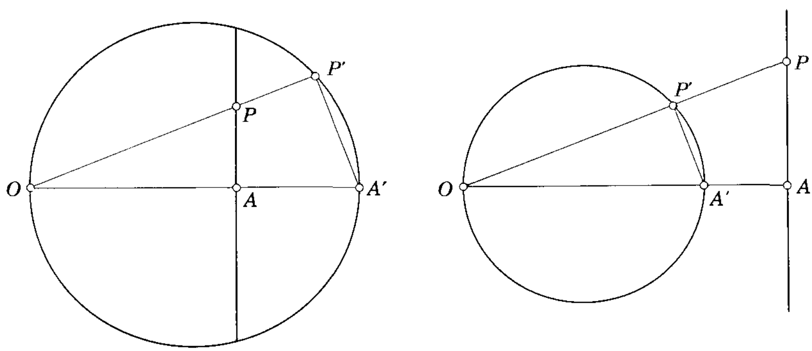



**Figure 6.3a**

We have seen that lines through $O$ invert into themselves. What about other lines? Let $A$ be the foot of the perpendicular from $O$ to a line not through $O$. Let $A'$ be the inverse of $A$, and $P'$ the inverse of any other point $P$ on the line. (See Figure 6.3*a* where, for simplicity, the circle of inversion has not been drawn.) Since

$$OP \times OP' = k^2 = OA \times OA',$$

the triangles $OAP$, $OP'A'$ are similar, and the line $AP$ inverts into the circle on $OA'$ as diameter, which is the locus of points $P'$ from which $OA'$ subtends a right angle. Thus any line not through $O$ inverts into a circle through $O$, and vice versa.

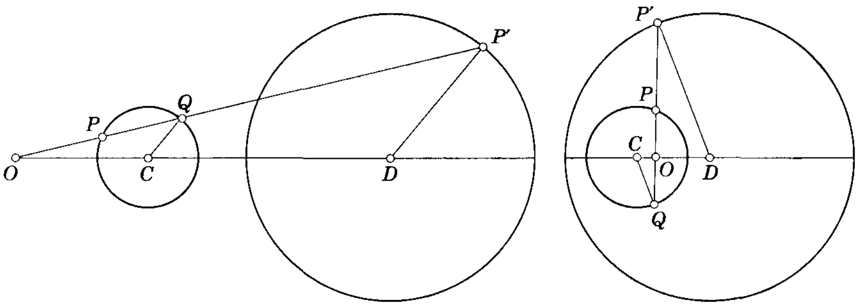



**Figure 6.3b**

Finally, what about a circle not through $O$? Let $P$ be any point on such a circle, with center $C$, and let $OP$ meet the circle again in $Q$. By Euclid III.35 again, the product

$$p = OP \times OQ$$

is independent of the position of $P$ on the circle. Following Jacob Steiner (1796–1863), we call this product the *power* of $O$ with respect to the circle. It is positive when $O$ is outside the circle, zero when $O$ lies on the circle, and we naturally regard it as being negative when $O$ is inside (so that $OP$ and $OQ$ are measured in opposite directions). Let the dilatation $O(k^2/p)$ transform the given circle and its radius $CQ$ into another circle (or possibly the same) and its parallel radius $DP'$ (Figure 6.3*b*, cf. Figure 5.2*a*), so that

$$\frac{OP'}{OQ} = \frac{OD}{OC} = \frac{k^2}{p}.$$

Since $OP \times OQ = p$, we have, by multiplication,

$$OP \times OP' = k^2.$$

Thus $P'$ is the inverse of $P$, and the circle with center $D$ is the desired inverse of the given circle with center $C$. (The point $D$ is usually *not* the inverse of $C$.)

We have thus proved that the inverse of a circle not through $O$ is another circle of the same kind, or possibly the same circle again. The latter possibility occurs in just two cases: (1) when the given circle is orthogonal to the circle of inversion, so that $p = k^2$ and the dilatation is the identity; (2) when the given circle is the circle of inversion itself, so that $p = -k^2$ and the dilatation is a half-turn.

When $p$ is positive (see the left half of Figure 6.3*b*), so that $O$ is outside the circle with center $C$, this circle is orthogonal to the circle with center $O$ and radius $\sqrt{p}$; that is, the former circle is invariant under inversion with respect to the latter. In effect, we have expressed the given inversion as the product of this new inversion, which takes $P$ to $Q$, and the dilatation $O(k^2/p)$, which takes $Q$ to $P'$. When $p$ is negative (as in the right half of Figure 6.3*b*), $P$ and $Q$ are interchanged by an "anti-inversion:" the product of an inversion with radius $\sqrt{-p}$ and a half-turn [Forder **3**, p. 20].

When discussing isometries and other similarities, we distinguished between *direct* and *opposite* transformations by observing their effect on a triangle. Since we are concerned only with *sense*, the triangle could have been replaced by its circumcircle. Such a distinction can still be made for inversions (and products of inversions), which transform circles into circles. Instead of a triangle we use a circle: not an arbitrary circle but a "small" circle whose inverse is also "small," that is, a circle not surrounding $O$. Referring again to the left half of Figure 6.3*b*, we observe that $P$ and $Q$ describe the circle with center $C$ in opposite senses, whereas $Q$ and $P'$ describe the two circles in the same sense. Thus the inverse points $P$ and $P'$ proceed oppositely, and

*Inversion is an opposite transformation.*

It follows that the product of an even number of inversions is direct. One instance is familiar: the product of inversions with respect to two concentric circles is a dilatation.

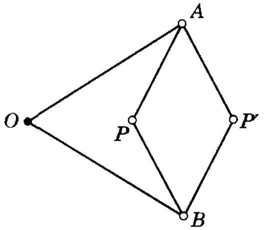



**Figure 6.3c**

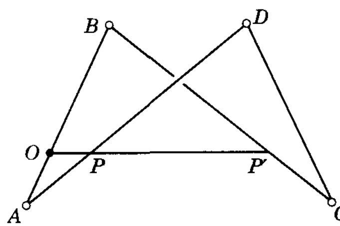



**Figure 6.3d**

**EXERCISES**

**1.** For any two unequal circles that do not intersect, one of the two centers of similitude (§ 5.2) is the center of a circle which inverts either of the given circles into the other. For two unequal intersecting circles, both centers of similitude have this property. What happens in the case of equal intersecting circles?

**2.** Explain the action of *Peaucellier's cell* (Figure 6.3*c*), an instrument invented by A. Peaucellier in 1864 for the purpose of drawing the inverse of any given locus. It is formed by four equal rods, hinged at the corners of a rhombus $APBP'$, and two equal (longer) rods connecting two opposite corners, $A$ and $B$, to a fixed pivot $O$. When a pencil point is inserted at $P'$ and a tracing point at $P$ (or vice versa), and the latter is traced over the curves of a given figure, the pencil point draws the inverse figure. In particular, if a seventh rod and another pivot are introduced so as to keep $P$ on a circle passing through $O$, the locus of $P'$ will be a straight line. This linkage gives an exact solution of the important mechanical problem of converting circular into rectilinear motion. [Lamb **2**, p. 314.]

**3.** Explain the action of *Hart's linkage* (Figure 6.3*d*), an instrument invented by H. Hart in 1874 for the same purpose as Peaucellier's cell. It requires only four rods, hinged at the corners of a "crossed parallelogram" $ABCD$ (with $AB = CD$, $BC = DA$). The three collinear points $O$, $P$, $P'$, on the respective rods $AB$, $AD$, $BC$, remain collinear (on a line parallel to $AC$ and $BD$) when the shape of the crossed parallelogram is changed. As before, the instrument is pivoted at $O$. [Lamb **2**, p. 315.]

**4.** With respect to a circle $\gamma$ of radius $r$, let $p$ be the power of an outside point $O$. Then the circle with center $O$ and radius $k$ inverts $\gamma$ into a circle of radius $k^2r/p$.

## 6.4 THE INVERSIVE PLANE

*Whereupon the Plumber said in tones of disgust:*
*"I suggest that we proceed at once to infinity."*

J. L. Synge [**2**, p. 131]

We have seen that the image of a given point $P$ by reflection in a line (Figure 1.3*b*) is the second intersection of any two circles through $P$ orthogonal to the mirror, and that the inverse of $P$ in a circle is the second intersection of any two circles through $P$ orthogonal to the circle of inversion. Because of this analogy, inversion is sometimes called "reflection in a circle" [Blaschke **1**, p. 47], and we extend the definition of a circle so as to include a straight line as a special (or "limiting") case: a circle of infinite radius. We can then say that *any* three distinct points lie on a unique circle, and that any circle inverts into a circle.

In the same spirit, we extend the Euclidean plane by inventing an "ideal" *point at infinity* $O'$, which is both a common point and the common center of all straight lines, regarded as circles of infinite radius. Two circles with a common point either touch each other or intersect again. This remains obvious when one of the circles reduces to a straight line. When both of them are straight, the lines are either parallel, in which case they touch at $O'$, or intersecting, in which case $O'$ is their second point of intersection [Hilbert and Cohn-Vossen **1**, p. 251].

We can now assert that *every* point has an inverse. All the lines through $O$, being "circles" orthogonal to the circle of inversion, meet again in $O'$, the inverse of $O$. When the center $O$ is $O'$ itself, the "circle" of inversion is straight, and the inversion reduces to a reflection.

The Euclidean plane with $O'$ added is called the *inversive* (or "conformal") *plane.** It gives inversion its full status as a "transformation" (§ 2.3): a one-to-one correspondence without exception.

Where two curves cross each other, their angle of intersection is naturally defined to be the angle between their tangents. In this spirit, two intersecting circles, being symmetrical by reflection in their line of centers, make equal angles at the two points of intersection. This will enable us to prove

**6.41** *Any angle inverts into an equal angle (or, more strictly, an opposite angle).*

We consider first an angle at a point $P$ which is not on the circle of inversion. Since any direction at such a point $P$ may be described as the direction of a suitable circle through $P$ and its inverse $P'$, two such directions are determined by two such circles. Since these circles are self-inverse, they serve to determine the corresponding directions at $P'$. To show that an angle at $P$ is still preserved when $P$ is self-inverse, we use 6.11 to express the given inversion as the product of a dilatation and the inversion in a concentric circle that does not pass through $P$. Since both these transformations preserve angles, their product does likewise.

In particular, right angles invert into right angles, and

**6.42** *Orthogonal circles invert into orthogonal circles (including lines as special cases).*

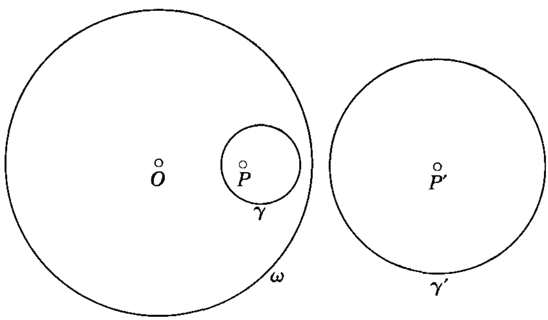



**Figure 6.4a**

By 6.21, inversion can be defined in terms of orthogonality. Therefore a circle and a pair of inverse points invert (in another circle) into a circle and a pair of inverse points. More precisely, if a circle $\gamma$ inverts $P$ into $Q$ and

* M. Bôcher, *Bulletin of the American Mathematical Society,* **20** (1914), p. 194.

a circle $\omega$ inverts $\gamma$, $P$, $Q$ into $\gamma'$, $P'$, $Q'$, then the circle $\gamma'$ inverts $P'$ into $Q'$. An important special case (Figure 6.4*a*) arises when $Q$ coincides with $O$, the center of $\omega$, so that $Q'$ is $O'$, the point at infinity. Then $P$ is the inverse of $O$ in $\gamma$, and $P'$ is the center of $\gamma'$. In other words, if $\gamma$ inverts $O$ into $P$, whereas $\omega$ inverts $\gamma$ and $P$ into $\gamma'$ and $P'$, then $P'$ is the center of $\gamma'$.

Two circles either touch, or cut each other twice, or have no common point. In the last case (when each circle lies entirely outside the other, or else one encloses the other), we may conveniently say that the circles *miss* each other.

If two circles, $\alpha_1$ and $\alpha_2$, are both orthogonal to two circles $\beta_1$ and $\beta_2$, we can invert the four circles in a circle whose center is one of the points of intersection of $\alpha_1$ and $\beta_1$, obtaining two orthogonal circles and two perpendicular diameters, as in the remark at the end of § 6.2. Hence, either $\alpha_1$ touches $\alpha_2$ and $\beta_1$ touches $\beta_2$, or $\alpha_1$ cuts $\alpha_2$ and $\beta_1$ misses $\beta_2$, or $\alpha_1$ misses $\alpha_2$ and $\beta_1$ cuts $\beta_2$.

## 6.5 COAXAL CIRCLES

In this section we leave the inversive plane and return to the Euclidean plane, in order to be able to speak of distances.

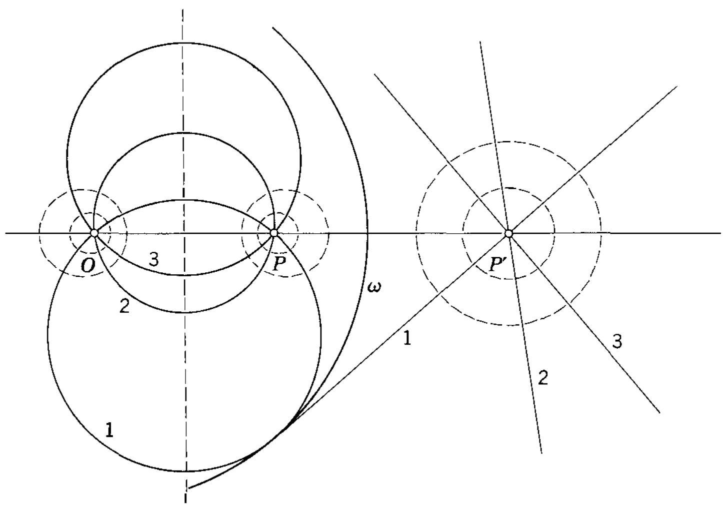



**Figure 6.5a**

If $P$ and $P'$ are inverse points in the circle $\omega$ (with center $O$), as in Figure 6.5*a*, all the lines through $P'$ invert into all the circles through $O$ and $P$: an *intersecting* (or "elliptic") *pencil of coaxal circles,* including the straight line $OPP'$ as a degenerate case. The system of concentric circles with center $P'$, consisting of circles orthogonal to these lines, inverts into a *nonintersecting* (or "hyperbolic") *pencil of coaxal circles* (drawn in broken lines). These circles all miss one another and are all orthogonal to the intersecting pencil. One of them degenerates to a (vertical) line, whose inverse is the circle (with center $P'$) passing through $O$.

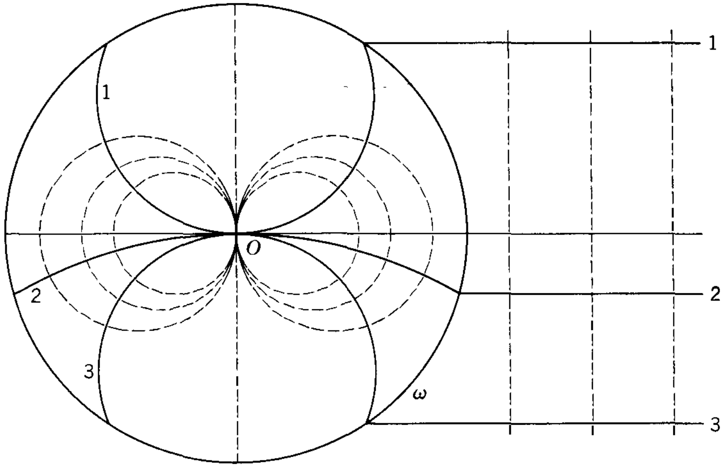



**Figure 6.5b**

As a kind of limiting case when $O$ and $P$ coincide (Figure 6.5*b*), the circles that touch a fixed line at a fixed point $O$ constitute a *tangent* (or "parabolic") *pencil of coaxal circles*. They invert (in a circle with center $O$) into all the lines parallel to the fixed line. Orthogonal to these lines we have another system of the same kind, inverting into an orthogonal tangent pencil of coaxal circles. Again each member of either pencil is orthogonal to every member of the other.

Any two given circles belong to a pencil of coaxal circles of one of these three types, consisting of *all the circles orthogonal to both of any two circles orthogonal to both the given circles.* (More concisely, the coaxal circles consist of all the circles orthogonal to all the circles orthogonal to the given circles.) Two circles that cut each other belong to an intersecting pencil (and can be inverted into intersecting lines); two circles that touch each other belong to a tangent pencil (and can be inverted into parallel lines); two circles that miss each other belong to a nonintersecting pencil (by the remark at the end of § 6.4).

Each pencil contains one straight line (a circle of infinite radius) called the *radical axis* (of the pencil, or of any two of its members).* For an intersecting pencil, this is the line joining the two points common to all the circles ($OP$ for the "unbroken" circles in Figure 6.5*a*). For a tangent pencil,

* Louis Gaultier, *Journal de l'École Polytechnique,* **16** (1813), p. 147.

it is the common tangent. For a nonintersecting pencil, it is the line midway between the two *limiting points* (or circles of zero radius) which are the common points of the orthogonal intersecting pencil. For each pencil there is a *line of centers,* which is the radical axis of the orthogonal pencil. Hence

**6.51** *If tangents can be drawn to the circles of a coaxal pencil from a point on the radical axis, all these tangents have the same length.*

The radical axis of two given circles may be defined as the locus of points of equal power (§ 6.3) with respect to the two circles. This power can be measured as the square of a tangent except in the case when the given circles intersect in two points $O$, $P$, and we are considering a point $A$ on the segment $OP$; then the power is the negative number $AO \times AP$.

It follows that, for three circles whose centers form a triangle, the three radical axes (of the circles taken in pairs) concur in a point called the *radical center,* which has the same power with respect to all three circles. If this power is positive, its square root is the length of the tangents to any of the circles, and the radical center is the center of a circle (of this radius) which is orthogonal to all the given circles. But if the power is negative, no such orthogonal circle exists.

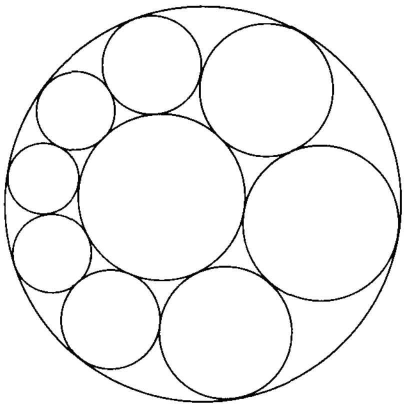


**Figure 6.5c**

The possibility of inverting any two nonintersecting circles into concentric circles (by taking $O$ at either of the limiting points) provides a remarkably simple proof for Steiner's porism:* If we have two (nonconcentric) circles, one inside the other, and circles are drawn successively touching them and one another, as in Figure 6.5*c*, it may happen that the ring of touching circles closes, that is, that the last touches the first. Steiner's statement is that, if this happens once, it will always happen, whatever be the position of the first circle of the ring. To prove this we need only invert the original two circles into concentric circles, for which the statement is obvious.

* Forder [**3**, p. 23]. See also Coxeter, Interlocked rings of spheres, *Scripta Mathematica,* **18** (1952), pp. 113–121, or Yaglom [**2**, p. 199].

**EXERCISES**

1. In a pencil of coaxal circles, each member, used as a circle of inversion, interchanges the remaining members in pairs and inverts each member of the orthogonal pencil into itself.

2. The two limiting points of a nonintersecting pencil are inverses of each other in any member of the pencil.

3. If two circles have two or four common tangents, their radical axis joins the midpoints of these common tangents. If two circles have no common tangent (i.e., if one entirely surrounds the other), how can we construct their radical axis?

4. When a nonintersecting pencil of coaxal circles is inverted into a pencil of concentric circles, what happens to the limiting points?

5. In Steiner's porism, the points of contact of successive circles in the ring all lie on a circle, and this will serve to invert the two original circles into each other. Do the *centers* of the circles in the ring lie on a circle?

6. For the triangle considered in Exercise 10 of § 1.5 (page 16), the incircle is coaxal with the "two other circles" (Soddy's circles).

## 6.6 THE CIRCLE OF APOLLONIUS

The analogy between reflection and inversion is reinforced by the following

**PROBLEM.** *To find the locus of a point $P$ whose distances from two fixed points $A$, $A'$ are in a constant ratio $1 : \mu$, so that*

$$A'P = \mu AP.$$

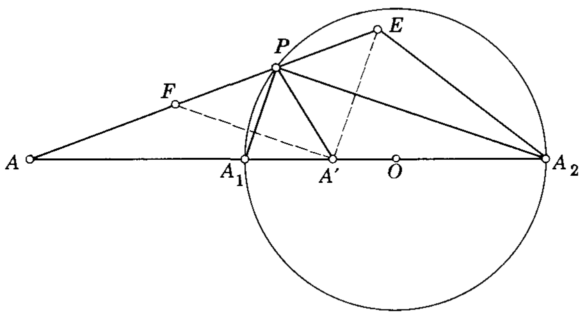



**Figure 6.6a**

When $\mu = 1$, the locus is evidently the perpendicular bisector of $AA'$, that is, the line that reflects $A$ into $A'$. We shall see that for other values of $\mu$ it is a circle that inverts $A$ into $A'$. (Apollonius of Perga, *c.* 260–190 B.C.)

Assuming $\mu \neq 1$, let $P$ be any point for which $A'P = \mu AP$. Let the internal and external bisectors of $\angle APA'$ meet $AA'$ in $A_1$ and $A_2$ (as in Figure 6.6$a$, where $\mu = \frac{1}{2}$). Take $E$ and $F$ on $AP$ so that $A'E$ is parallel to $A_1P$ and $A'F$ is parallel to $A_2P$, that is, perpendicular to $A_1P$. Since $FP = PA' = PE$, we have

$$\frac{AA_1}{A_1A'} = \frac{AP}{PE} = \frac{AP}{PA'}, \qquad \frac{AA_2}{A'A_2} = \frac{AP}{FP} = \frac{AP}{PA'}.$$

(The former result is Euclid VI.3.) Thus $A_1$ and $A_2$ divide the segment $AA'$ internally and externally in the ratio $1 : \mu$, and their location is independent of the position of $P$. Since $\angle A_1PA_2$ is a right angle, $P$ lies on the circle with diameter $A_1A_2$.

Conversely, if $A_1$ and $A_2$ are defined by their property of dividing $AA'$ in the ratio $1 : \mu$, and $P$ is any point on the circle with diameter $A_1A_2$, we have

$$\frac{AP}{PE} = \frac{AA_1}{A_1A'} = \frac{1}{\mu} = \frac{AA_2}{A'A_2} = \frac{AP}{FP}.$$

Thus $FP = PE$, and $P$, being the midpoint of $FE$, is the circumcenter of the right-angled triangle $EFA'$. Therefore $PA' = PE$ and

$$\frac{AP}{PA'} = \frac{AP}{PE} = \frac{1}{\mu}$$

[Court **2**, p. 15].

Finally, the *circle of Apollonius* $A_1A_2P$ inverts $A$ into $A'$. For, if $O$ is its center and $k$ its radius, the distances $a = AO$ and $a' = A'O$ satisfy

$$\frac{a-k}{k-a'} = \frac{AA_1}{A_1A'} = \frac{AA_2}{A'A_2} = \frac{a+k}{a'+k},$$

whence

$$aa' = k^2.$$

## EXERCISES

**1.** When $\mu$ varies while $A$ and $A'$ remain fixed, the circles of Apollonius form a nonintersecting pencil with $A$ and $A'$ for limiting points.

**2.** Given a line $l$ and two points $A$, $A'$ (not on $l$), locate points $P$ on $l$ for which the ratio $A'P/AP$ is maximum or minimum. (*Hint*: Consider the circle through $A$, $A'$ with its center on $l$. The problem is due to N. S. Mendelsohn, and the hint to Richard Blum.)

**3.** Express $k/AA'$ in terms of $\mu$.

**4.** In the notation of Figure 5.6$a$ (which is embodied in Figure 6.6$b$), the circles on $A_1A_2$ and $B_1B_2$ as diameters meet in two points $O$ and $\bar{O}$, such that the triangles $OAB$ and $OA'B'$ are similar, and likewise the triangles $\bar{O}AB$ and $\bar{O}A'B'$. Of the two similarities

$$OAB \to OA'B' \quad \text{and} \quad \bar{O}AB \to \bar{O}A'B',$$

one is opposite and the other direct. In fact, $O$ is where $A_1B_1$ meets $A_2B_2$, and $\bar{O}$ lies on the four further circles $AA'P$, $BB'P$, $ABT$, $A'B'T$ (cf. Ex. 2 of § 5.5). [Casey **1**, p. 185.] If $A'$ coincides with $B$, $O$ lies on $AB'$.

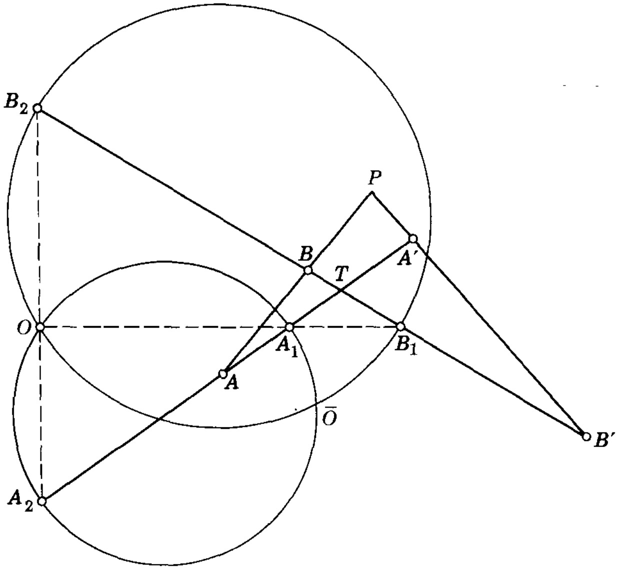



**Figure 6.6b**

5. Let the *inversive distance* between two nonintersecting circles be defined as the natural logarithm of the ratio of the radii (the larger to the smaller) of two concentric circles into which the given circles can be inverted. Then, if a nonintersecting pencil of coaxal circles includes $\alpha_1$, $\alpha_2$, $\alpha_3$ (in this order), the three inversive distances satisfy

$$(\alpha_1, \alpha_2) + (\alpha_2, \alpha_3) = (\alpha_1, \alpha_3).$$

6. Two given unequal circles are related by infinitely many dilative rotations and by infinitely many dilative reflections. The locus of invariant points (in either case) is the circle having for diameter the segment joining the two centers of similitude of the given circles. (This locus is known as the *circle of similitude* of the given circles.) What is the corresponding result for two given *equal* circles?

7. The inverses, in two given circles, of a point on their circle of similitude, are images of each other by reflection in the radical axis of the two circles [Court **2,** p. 199].

## 6.7 CIRCLE-PRESERVING TRANSFORMATIONS

Having observed that inversion is a transformation of the whole inversive plane (including the point at infinity) into itself, taking circles into circles, we naturally ask what is the most general transformation of this kind. We distinguish two cases, according as the point at infinity is, or is not, invariant.

In the former case, not only are circles transformed into circles but also lines into lines. With the help of Euclid III.21 (see p. 7) we deduce that equality of angles is preserved, and consequently the measurement of angles is preserved, so that every triangle is transformed into a similar triangle, and the transformation is a similarity (§ 5.4).

If, on the other hand, the given transformation T takes an ordinary point

$O$ into the point at infinity $O'$, we consider the product $J_1T$, where $J_1$ is the inversion in the unit circle with center $O$. This product $J_1T$, leaving $O'$ invariant, is a similarity. Let $k^2$ be its ratio of magnification, and $J_k$ the inversion in the circle with center $O$ and radius $k$. Since, by 6.11, $J_1J_k$ is the dilatation $O(k^2)$, the similarity $J_1T$ can be expressed as $J_1J_kS$, where S is an isometry. Thus

$$T = J_kS,$$

the product of an inversion and an isometry.

To sum up,

**6.71** *Every circle-preserving transformation of the inversive plane is either a similarity or the product of an inversion and an isometry.*

It follows that every circle-preserving transformation is the product of at most four inversions (provided we regard a reflection as a special kind of inversion) [Ford **1**, p. 26]. Such a transformation is called a *homography* or an *antihomography* according as the number of inversions is even or odd. The product of two inversions (either of which could be just a reflection) is called a *rotary* or *parabolic* or *dilative* homography according as the two inverting circles are intersecting, tangent, or nonintersecting (i.e., according as the orthogonal pencil of invariant circles is nonintersecting, tangent, or intersecting). As special cases we have, respectively, a rotation, a translation, and a dilatation. The most important kind of rotary homography is the *Möbius involution,* which, being the inversive counterpart of a half-turn, is the product of inversions in two orthogonal circles (e.g., the product of the inversion in a circle and the reflection in a diameter). Any product of four inversions that cannot be reduced to a product of two is called a *loxodromic* homography [Ford **1**, p. 20].

### EXERCISE

When a given circle-preserving transformation is expressed as JS (where J is an inversion and S an isometry), J and S are unique. There is an equally valid expression SJ′, in which the isometry precedes the inversion. Why does this revised product involve the same S? Under what circumstances will we have $J' = J$?

## 6.8 INVERSION IN A SPHERE

By revolving Figures 6.1*a*, 6.2*a*, 6.3*a*, 6.3*b*, and 6.4*a* about the line of centers ($OP$ or $OA$ or $OC$), we see that the whole theory of inversion extends readily from circles in the plane to spheres in space. Given a sphere with center $O$ and radius $k$, we define the inverse of any point $P$ (distinct from $O$) to be the point $P'$ on the ray $OP$ whose distance from $O$ satisfies

$$OP \times OP' = k^2.$$

Alternatively, $P'$ is the second intersection of three spheres through $P$ orthogonal to the sphere of inversion. Every sphere inverts into a sphere, rovided we include, as a sphere of infinite radius, a plane, which is the inverse of a sphere through $O$. Thus, inversion is a transformation of *inversive* (or "conformal") *space,* which is derived from Euclidean space by postulating a *point at infinity,* which lies on all planes and lines.

Revolving the circle of Apollonius (Figure 6.6*a*) about the line $AA'$, we obtain the *sphere of Apollonius,* which may be described as follows:

**6.81** *Given two points $A$, $A'$ and a positive number $\mu$, let $A_1$ and $A_2$ divide $AA'$ internally and externally in the ratio $1 : \mu$. Then the sphere on $A_1A_2$ as diameter is the locus of a point $P$ whose distances from $A$ and $A'$ are in this ratio.*

**EXERCISES**

**1.** If a sphere with center $O$ inverts $A$ into $A'$ and $B$ into $B'$, the triangles $OAB$ and $OB'A'$ are similar.

**2.** In terms of $a = OA$ and $b = OB$, we have (in the notation of Ex. 1)

$$A'B' = \frac{k^2}{ab} AB.$$

**3.** The "cross ratio" of any four points is preserved by any inversion:

$$\frac{AB/BD}{AC/CD} = \frac{A'B'/B'D'}{A'C'/C'D'}.$$

[Casey **1**, p. 100.]

**4.** Two spheres which touch each other at $O$ invert into parallel planes.

**5.** Let $\alpha$, $\beta$, $\gamma$ be three spheres all touching one another. Let $\sigma_1, \sigma_2, \ldots$ be a sequence of spheres touching one another successively and all touching $\alpha$, $\beta$, $\gamma$. Then $\sigma_6$ touches $\sigma_1$, so that we have a ring of six spheres interlocked with the original ring of three.* (*Hint:* Invert in a sphere whose center is the point of contact of $\alpha$ and $\beta$.)

## 6.9 THE ELLIPTIC PLANE

*In some unaccountable way, while he [Davidson] moved hither and thither in London, his sight moved hither and thither in a manner that corresponded, about this distant island. . . . When I said that nothing would alter the fact that the place [Antipodes Island] is eight thousand miles away, he answered that two points might be a yard away on a sheet of paper, and yet be brought together by bending the paper round.*

H. G. Wells (1866-1946)

(*The Remarkable Case of Davidson's Eyes*)

Let $S$ be the foot of the perpendicular from a point $N$ to a plane $\sigma$, as in Figure 6.9*a*. A sphere (not drawn) with center $N$ and radius $NS$ inverts the plane $\sigma$ into the sphere $\sigma'$ on $NS$ as diameter [Johnson **1**, p. 108]. We have

* Frederick Soddy, The Hexlet, *Nature,* **138** (1936), p. 958; **139** (1937), p. 77.

seen that spheres invert into spheres (or planes); therefore circles, being intersections of spheres, invert into circles (or lines). In particular, all the circles in $\sigma$ invert into circles (great or small) on the sphere $\sigma'$, and all the lines in $\sigma$ invert into circles through $N$. Each point $P$ in $\sigma$ yields a corresponding point $P'$ on $\sigma'$, namely, the second intersection of the line $NP$ with $\sigma'$. Conversely, each point $P'$ on $\sigma'$, except $N$, corresponds to the point $P$ in which $NP'$ meets $\sigma$. The exception can be removed by making $\sigma$ an inversive plane whose point at infinity is the inverse of $N$.

This inversion, which puts the points of the inversive plane into one-to-one correspondence with the points of a sphere, is known as *stereographic projection*. It serves as one of the simplest ways to map the geographical globe on a plane. Since angles are preserved, small islands are mapped with the correct shape, though on various scales according to their latitude.

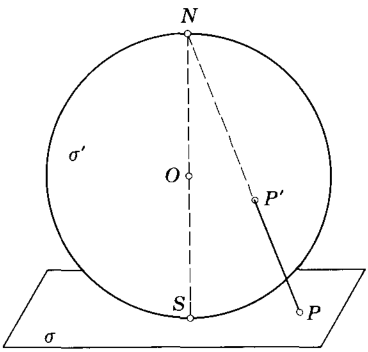



**Figure 6.9a**

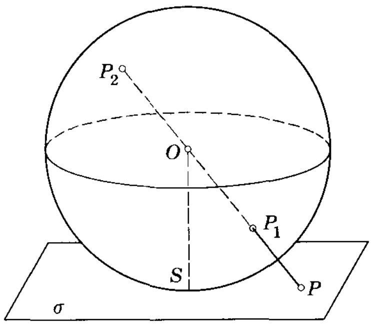



**Figure 6.9b**

Another way is by *gnomonic* (or central) *projection*, in which the point from which we project is not $N$ but $O$, the center of the sphere, as in Figure 6.9*b*. Each point $P$ in $\sigma$ yields a line $OP$, joining it to $O$. This diameter meets the sphere in two antipodal points $P_1$, $P_2$, which are both mapped on the same point $P$. Each line $m$ in $\sigma$ yields a plane $Om$, joining it to $O$. This diametral plane meets the sphere in a *great circle*. Conversely, each great circle of the sphere, except the "equator" (whose plane is parallel to $\sigma$), corresponds to a line in $\sigma$. This time the exception can be removed by adding to the Euclidean plane $\sigma$ a *line at infinity* (representing the equator) with all its points, called *points at infinity*, which represent pairs of antipodal points on the equator. Thus, all the lines parallel to a given line contain the same point at infinity, but lines in different directions have different points at infinity, all lying on the same line at infinity. (This idea is due to Kepler and Desargues.)

When the line at infinity is treated just like any other line, the plane so extended is called the *projective plane* or, more precisely, the *real* projective plane [Coxeter **2**]. Two parallel lines meet in a point at infinity, and an ordinary line meets the line at infinity in a point at infinity. Hence

**6.91** *Any two lines of the projective plane meet in a point.*

Instead of taking a section of all the lines and planes through $O$, we could more symmetrically (though more abstractly) declare that, by definition, the points and lines of the projective plane *are* the lines and planes through $O$. The statement 6.91 is no longer surprising; it merely says that any two planes through $O$ meet in a line through $O$.

Equivalently we could declare that, by definition, the lines of the projective plane are the great circles on a sphere, any two of which meet in a pair of antipodal points. Then the points of the projective plane are the pairs of antipodal points, abstractly identified. This abstract identification was vividly described by H. G. Wells in his short story, *The Remarkable Case of Davidson's Eyes.* (A sudden catastrophe distorted Davidson's field of vision so that he saw everything as it would have appeared from an exactly antipodal position on the earth.)

When the inversive plane is derived from the sphere by stereographic projection, distances are inevitably distorted, but the angle at which two circles intersect is preserved. In this sense, the inversive plane has a partial metric: angles are measured in the usual way, but distances are never mentioned [Graustein **1**, pp. 377, 388, 395].

On the other hand, gnomonic projection enables us, if we wish, to give the *projective* plane a *complete* metric. The distance between two points $P$ and $Q$ in $\sigma$ (Figure 6.9*a*) is defined to be the angle $POQ$ (in radian measure), and the angle between two lines $m$ and $n$ in $\sigma$ is defined to be the angle between the planes $Om$ and $On$. (This agrees with the customary measurement of distances and angles on a sphere, as used in spherical trigonometry.) We have thus obtained the *elliptic* plane* or, more precisely, the real projective plane with an elliptic metric [Coxeter **3**, Chapter VI; E. T. Bell **2**, pp. 302–311; Bachmann **1**, p. 21].

Since the points of the elliptic plane are in one-to-two correspondence with the points of the unit sphere, whose total area is $4\pi$, it follows that the total area of the elliptic plane (according to the most natural definition of "area") is $2\pi$. Likewise, the total length of a line (represented by a "great semicircle") is $\pi$. The simplification that results from using the elliptic plane instead of the sphere is well illustrated by the problem of computing the area of a spherical triangle $ABC$, whose sides are arcs of three great circles. Figure 6.9*c* shows these great circles, first in stereographic projection and then in gnomonic projection. The elliptic plane is decomposed, by the three lines $BC$, $CA$, $AB$, into four triangular regions. One of them is the given triangle $\Delta$ with angles $A$, $B$, $C$; the other three are marked $\alpha$, $\beta$, $\gamma$ in Figure 6.9*c*. (On the sphere, we have, of course, not only four regions but eight.) The two

* The name "elliptic" is possibly misleading. It does not imply any direct connection with the curve called an ellipse, but only a rather far-fetched analogy. A central conic is called an ellipse or a hyperbola according as it has no asymptote or two asymptotes. Analogously, a non-Euclidean plane is said to be elliptic or hyperbolic (Chapter 16) according as each of its lines contains no point at infinity or two points at infinity.

lines $CA$, $AB$ decompose the plane into two *lunes* whose areas, being proportional to the supplementary angles $A$ and $\pi - A$, are exactly $2A$ and $2(\pi - A)$. The lune with angle $A$ is made up of the two regions $\Delta$ and $\alpha$. Hence

$$\Delta + \alpha = 2A.$$

Similarly $\Delta + \beta = 2B$ and $\Delta + \gamma = 2C$. Adding these three equations and subtracting

$$\Delta + \alpha + \beta + \gamma = 2\pi,$$

we deduce Girard's "spherical excess" formula

**6.92** $$\Delta = A + B + C - \pi,$$

which is equally valid for the sphere and the elliptic plane. (A. Girard, *Invention nouvelle en algèbre,* Amsterdam, 1629.)

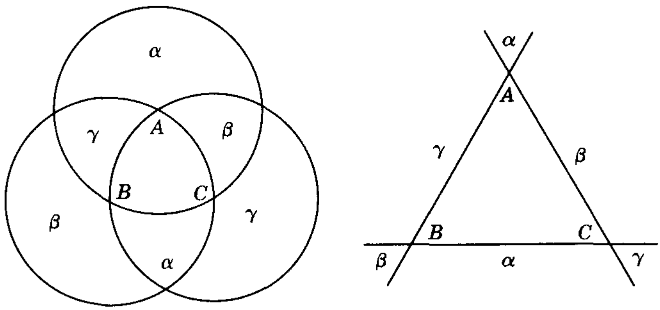



**Figure 6.9c**

## EXERCISES

**1.** Two circles in the elliptic plane may have as many as four points of intersection.

**2.** The area of a $p$-gon in the elliptic plane is equal to the excess of its angle sum over the angle sum of a $p$-gon in the Euclidean plane.

# Part III

# 12
# Ordered geometry

During the last 2000 years, the two most widely read books have undoubtedly been the Bible and the Elements. Scholars find it an interesting task to disentangle the various accounts of the Creation that are woven together in the Book of Genesis. Similarly, as Euclid collected his material from various sources, it is not surprising that we can extract from the Elements two self-contained geometries that differ in their logical foundation, their primitive concepts and axioms. They are known as *absolute* geometry and *affine* geometry. After describing them briefly in § 12.1, we shall devote the rest of this chapter to those propositions which belong to both: propositions so fundamental and "obvious" that Euclid never troubled to mention them.

## 12.1 THE EXTRACTION OF TWO DISTINCT GEOMETRIES FROM EUCLID

*The pursuit of an idea is as exciting as the pursuit of a whale.*

Henry Norris Russell (1877-1957)

Absolute geometry, first recognized by Bolyai (1802–1860), is the part of Euclidean geometry that depends on the first four Postulates without the fifth. Thus it includes the propositions I.1–28, III.1–19, 25, 28–30; IV.4–9 (with a suitably modified definition of "square"). The study of absolute geometry is motivated by the fact that these propositions hold not only in Euclidean geometry but also in hyperbolic geometry, which we shall study in Chapter 16. In brief, absolute geometry is geometry without the assumption of a unique parallel (through a given point) to a given line.

On the other hand, in affine geometry, first recognized by Euler (1707–1783), the unique parallel plays a leading role. Euclid's third and fourth postulates become meaningless, as circles are never mentioned and angles are never measured. In fact, the only admissible isometries are half-turns and translations. The affine propositions in Euclid are those which are preserved by parallel projection from one plane to another [Yaglom **2**, p. 17]:

for example, I. 30, 33–45, and VI. 1, 2, 4, 9, 10, 24–26. The importance of affine geometry has lately been enhanced by the observation that these propositions hold not only in Euclidean geometry but also in Minkowski's geometry of time and space, which Einstein used in his special theory of relativity.

Since each of Euclid's propositions is affine or absolute or neither, we might at first imagine that the two geometries (which we shall discuss in Chapters 13 and 15, respectively) had nothing in common except Postulates I and II. However, we shall see in the present chapter that there is a quite impressive nucleus of propositions belonging properly to both. The essential idea in this nucleus is *intermediacy* (or "betweenness"), which Euclid used in his famous definition:

*A line (segment) is that which lies evenly between its ends.*

This suggests the possibility of regarding intermediacy as a primitive concept and using it to define a line segment as the set of all points between two given points. In the same spirit we can extend the segment to a whole infinite line. Then, if $B$ lies between $A$ and $C$, we can say that the three points $A$, $B$, $C$ lie in *order* on their line. This relation of order can be extended from three points to four or more.

Euclid himself made no explicit use of order, except in connection with measurement: saying that one magnitude is greater or less than another. It was Pasch, in 1882, who first pointed out that a geometry of order could be developed without reference to measurement. His system of axioms was gradually improved by Peano (1889), Hilbert (1899), and Veblen (1904).

Etymologically, "geometry without measurement" looks like a contradiction in terms. But we shall find that the passage from axioms and simple theorems to "interesting" theorems resembles Euclid's work in spirit, though not in detail.

This basic geometry, the common foundation for the affine and absolute geometries, is sufficiently important to have a name. The name *descriptive* geometry, used by Bertrand Russell [**1**, p. 382], was not well chosen, because it already had a different meaning. Accordingly, we shall follow Artin [**1**, p. 73] and say *ordered* geometry.

We shall pursue this rigorous development far enough to give the reader its flavor without boring him. The whole story is a long one, adequately told by Veblen [**1**] and Forder [**1**, Chapter II, and the *Canadian Journal of Mathematics,* **19** (1967), pp. 997–1000].

It is important to remember that, in this kind of work, we must define all the concepts used (except the primitive concepts) and prove all the statements (except the axioms), however "obvious" they may seem.

## EXERCISES

**1.** Is the ratio of two lengths along one line a concept belonging to absolute geometry or to affine geometry or to both? (*Hint:* In "one dimension," i.e., when we consider only the points on a single line, the distinction between *absolute* and *affine* disappears.)

**2.** Name a Euclidean theorem that belongs neither to absolute geometry nor to affine geometry.

**3.** The concurrence of the medians of a triangle (1.41) is a theorem belonging to both absolute geometry and affine geometry. To which geometry does the rest of § 1.4 belong?

**4.** Which geometry deals (a) with parallelograms? (b) with regular polygons? (c) with Fagnano's problem (§ 1.8)?

## 12.2 INTERMEDIACY

*A discussion of order . . . has become essential to any understanding of the foundation of mathematics.*

Bertrand Russell (1872- )
[Russell **1**, p. 199]

In Pasch's development of ordered geometry, as simplified by Veblen, the only primitive concepts are *points* $A, B, \ldots$ and the relation of *intermediacy* $[ABC]$, which says that $B$ is between $A$ and $C$. If $B$ is not between $A$ and $C$, we say simply "not $[ABC]$." There are ten axioms (12.21–12.27, 12.42, 12.43, and 12.51), which we shall introduce where they are needed among the various definitions and theorems.

AXIOM **12.21** *There are at least two points.*

AXIOM **12.22** *If $A$ and $B$ are two distinct points, there is at least one point $C$ for which $[ABC]$.*

AXIOM **12.23** *If $[ABC]$, then $A$ and $C$ are distinct: $A \neq C$.*

AXIOM **12.24** *If $[ABC]$, then $[CBA]$ but not $[BCA]$.*

THEOREM **12.241** *If $[ABC]$ then not $[CAB]$.*

*Proof.* By Axiom 12.24, $[CAB]$ would imply not $[ABC]$.

THEOREM **12.242** *If $[ABC]$, then $A \neq B \neq C$* (that is, in view of Axiom 12.23, the three points are all distinct).

*Proof.* If $B = C$, the two conclusions of Axiom 12.24 are contradictory. Similarly, we cannot have $A = B$.

DEFINITIONS. If $A$ and $B$ are two distinct points, the *segment* $AB$ is the set of points $P$ for which $[APB]$. We say that such a point $P$ is *on* the segment. Later we shall apply the same preposition to other sets, such as "lines."

THEOREM **12.243** *Neither $A$ nor $B$ is on the segment $AB$.*

*Proof.* If $A$ or $B$ were on the segment, we would have $[AAB]$ or $[ABB]$, contradicting 12.242.

**THEOREM 12.244** *Segment $AB$ = segment $BA$.*

*Proof.* By Axiom 12.24, $[APB]$ implies $[BPA]$.

**DEFINITIONS.** The *interval* $\overline{AB}$ is the segment $AB$ plus its *end points* $A$ and $B$:

$$\overline{AB} = A + AB + B.$$

The *ray* $A/B$ ("from $A$, away from $B$") is the set of points $P$ for which $[PAB]$. The *line* $AB$ is the interval $\overline{AB}$ plus the two rays $A/B$ and $B/A$:

$$\text{line } AB = A/B + \overline{AB} + B/A.$$

**COROLLARY 12.2441** Interval $\overline{AB}$ = interval $\overline{BA}$; line $AB$ = line $BA$.

**AXIOM 12.25** *If $C$ and $D$ are distinct points on the line $AB$, then $A$ is on the line $CD$.*

**THEOREM 12.251** *If $C$ and $D$ are distinct points on the line $AB$ then*

$$\text{line } AB = \text{line } CD.$$

*Proof.* If $A$, $B$, $C$, $D$ are not all distinct, suppose $D = B$. To prove that line $AB$ = line $BC$, let $X$ be any point on $BC$ except $A$ or $B$. By 12.25, $A$, like $X$, is on $BC$. Therefore $B$ is on $AX$, and $X$ is on $AB$. Thus every point on $BC$ is also on $AB$. Interchanging the roles of $A$ and $C$, we see that similarly every point on $AB$ is also on $BC$. Thus $AB = BC$. Finally, if $A$, $B$, $C$, $D$ are all distinct, we have $AB = BC = CD$.

**COROLLARY 12.2511** Two distinct points lie on just one line. Two distinct lines (if such exist) have at most one common point. (Such a common point $F$ is called a point of intersection, and the lines are said to *meet* in $F$.)

**COROLLARY 12.2512** Any three distinct points $A$, $B$, $C$ on a line satisfy just one of the relations $[ABC]$, $[BCA]$, $[CAB]$.

**AXIOM 12.26** *If $AB$ is a line, there is a point $C$ not on this line.*

**THEOREM 12.261** *If $C$ is not on the line $AB$, then $A$ is not on $BC$, nor $B$ on $CA$: the three lines $BC$, $CA$, $AB$ are distinct.*

*Proof.* By 12.25, if $A$ were on $BC$, $C$ would be on $AB$.

**DEFINITIONS.** Points lying on the same line are said to be *collinear.* Three non-collinear points $A$, $B$, $C$ determine a *triangle* $ABC$, which consists of these three points, called *vertices,* together with the three segments $BC$, $CA$, $AB$, called *sides.*

**AXIOM 12.27** *If $ABC$ is a triangle and $[BCD]$ and $[CEA]$, then there is, on the line $DE$, a point $F$ for which $[AFB]$.* (See Figure 12.2*a*.)

**THEOREM 12.271** *Between two distinct points there is another point.*

*Proof.* Let $A$ and $B$ be the two points. By 12.26, there is a point $E$ not on the line $AB$. By 12.22, there is a point $C$ for which $[AEC]$. By 12.251, the line $AC$ is the same as $AE$. By 12.261 (applied to $ABE$), $B$ is not on this line: therefore $ABC$ is a triangle. By 12.22 again, there is a point $D$ for which $[BCD]$. By 12.27 there is a point $F$ between $A$ and $B$.

THEOREM **12.272** *In the notation of Axiom* 12.27, $[DEF]$.

*Proof.* Since $F$ lies on the line $DE$, there are (by 12.2512) just five possibilities: $F = D$, $F = E$, $[EFD]$, $[FDE]$, $[DEF]$. Either of the first two would make $A$, $B$, $C$ collinear.

If $[EFD]$, we could apply 12.27 to the triangle $DCE$ with $[CEA]$ and $[EFD]$ (Figure 12.2*b*), obtaining $X$ on $AF$ with $[DXC]$. Since $AF$ and $CD$ cannot meet more than once, we have $X = B$, so that $[DBC]$. Since $[BCD]$, this contradicts 12.24.

Similarly (Figure 12.2*c*) we cannot have $[FDE]$. The only remaining possibility is $[DEF]$.

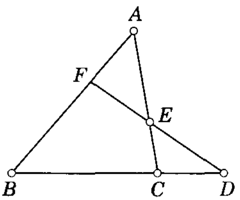



**Figure 12.2a**

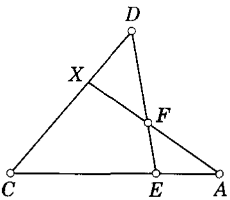



**Figure 12.2b**

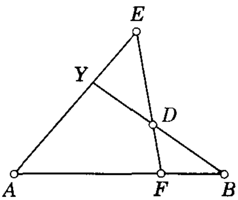



**Figure 12.2c**

This proof is typical; so let us be content to give the remaining theorems without proofs [Veblen **1**, pp. 9–15; Forder **1**, pp. 49–55].

**12.273** *A line cannot meet all three sides of a triangle.* (Remember that the "sides" are not intervals, nor whole lines, but only segments.)

**12.274** *If* $[ABC]$ *and* $[BCD]$, *then* $[ABD]$.

**12.275** *If* $[ABC]$ *and* $[ABD]$ *and* $C \neq D$, *then* $[BCD]$ *or* $[BDC]$, *and* $[ACD]$ *or* $[ADC]$.

**12.276** *If* $[ABD]$ *and* $[ACD]$ *and* $B \neq C$, *then* $[ABC]$ *or* $[ACB]$.

**12.277** *If* $[ABC]$ *and* $[ACD]$, *then* $[BCD]$ *and* $[ABD]$.

DEFINITION. If $[ABC]$ and $[ACD]$, we write $[ABCD]$.

This four-point order is easily seen to have all the properties that we should expect, for example, if $[ABCD]$, then $[DCBA]$, but all the other orders are false.

Any point $O$ on a segment $AB$ decomposes the segment into two segments: $AO$ and $OB$. (We are using the word *decomposes* in a technical sense [Veblen **1**, p. 21], meaning that every point on the segment $AB$ except $O$ itself is on just one of the two "smaller" segments.) Any point $O$ on a ray from $A$ decomposes the ray into a segment and a ray: $AO$ and $O/A$. Any point $O$ on a line decomposes the line into two "opposite" rays; if $[AOB]$, the rays are $O/A$ and $O/B$. The ray $O/A$, containing $B$, is sometimes more conveniently called *the ray OB*.

For any integer $n > 1$, $n$ distinct collinear points decompose their line into two rays and $n - 1$ segments. The points can be named $P_1, P_2, \ldots\ldots P_n$ in such a way that the two rays are $P_1/P_n$, $P_n/P_1$, and the $n - 1$ segments are

$$P_1P_2,\ P_2P_3,\ \ldots,\ P_{n-1}P_n,$$

each containing none of the points. We say that the points are in the *order* $P_1P_2 \ldots P_n$, and write $[P_1P_2 \ldots P_n]$. Necessary and sufficient conditions for this are

$$[P_1P_2P_3],\quad [P_2P_3P_4],\ \ldots,\quad [P_{n-2}P_{n-1}P_n].$$

Naturally, the best logical development of any subject uses the simplest or "weakest" possible set of axioms. (The worst occurs when we go to the opposite extreme and assume everything, so that there is no development at all!) In his original formulation of Axiom 12.27 [Pasch and Dehn **1**, p. 2 : "IV. Kernsatz"] Pasch made the following far stronger statement: If a line in the plane of a given triangle meets one side, it also meets another side (or else passes through a vertex). Peano's formulation, which we have adopted, excels this in two respects. The word "plane" (which we shall define in § 12.4) is not used at all, and the line $DE$ penetrates the triangle $ABC$ in a special manner, namely, before entering through the side $CA$, it comes from a point $D$ on $C/B$. It might just as easily have come from a point on $A/B$ (which is the same with $C$ and $A$ interchanged) or from a point on $B/A$ or $B/C$ (which is quite a different story). The latter possibility (with a slight change of notation) is covered by the following theorem (12.278). Axiom 12.27 is "only just strong enough"; for, although it enables us to deduce the statement 12.278 of apparently equal strength, we could not reverse the roles: if we tried instead to use 12.278 as an axiom, we would not be able to deduce 12.27 as a theorem!

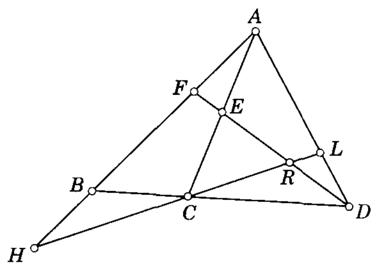



**Figure 12.2d**

THEOREM **12.278** *If $ABC$ is a triangle and $[AFB]$ and $[BCD]$, then there is, on the line $DF$, a point $E$ for which $[CEA]$.*

*Proof.* Take $H$ on $B/F$ (as in Figure 12.2*d*) and consider the triangle $DFB$ with $[FBH]$ and $[BCD]$. By 12.27 and 12.272, there is a point $R$ for which $[DRF]$ and $[HCR]$. By 12.274, $[AFB]$ and $[FBH]$ imply $[AFH]$. Thus we have a triangle $DAF$ with $[AFH]$ and $[FRD]$. By 12.27 and 12.272 again, there is a point $L$ for which $[DLA]$ and $[HRL]$. By 12.277, $[HCR]$ and $[HRL]$

imply $[CRL]$. Thus we have a triangle $CAL$ with $[ALD]$ and $[LRC]$. By 12.27 a third time, there is, on the line $DR$ $(=DF)$, a point $E$ for which $[CEA]$.

**EXERCISES**

1. A line contains infinitely many points.
2. We have defined a segment as a set of points. At what stage in the above development can we assert that this set is never the *null* set? [Forder **1**, p. 50.]
3. In the proof of 12.272, we had to show that the relation $[FDE]$ leads to a contradiction. Do this by applying 12.27 to the triangle $BFD$ (instead of $EAF$).
4. Given a finite set of lines, there are infinitely many points not lying on any of the lines.
5. If $ABC$ is a triangle and $[BLC]$, $[CMA]$, $[ANB]$, then there is a point $E$ for which $[AEL]$ and $[MEN]$. [Forder **1**, p. 56.]
6. If $ABC$ is a triangle, the three rays $B/C$, $A/C$, $A/B$ have a *transversal* (that is, a line meeting them all). (K. B. Leisenring.)
7. If $ABC$ is a triangle, the three rays $B/C$, $C/A$, $A/B$ have no transversal.

## 12.3 SYLVESTER'S PROBLEM OF COLLINEAR POINTS

*Almost any field of mathematics offered an enchanting world for discovery to Sylvester.*

E. T. Bell [**1**, p. 433]

It may seem to some readers that we have been using self-evident axioms to prove trivial results. Any such feeling of irritation is likely to evaporate when it is pointed out that the machinery so far developed is sufficiently powerful to deal effectively with Sylvester's conjecture (§ 4.7), which baffled the world's mathematicians for forty years. This matter of collinearity clearly belongs to ordered geometry. Kelly's Euclidean proof involves the extraneous concept of distance: it is like using a sledge hammer to crack an almond. The really appropriate nutcracker is provided by the following argument.

THEOREM. *If n points are not all collinear, there is at least one line containing exactly two of them.*

*Proof.* Let $P_1, P_2, \ldots, P_n$ be the $n$ points, so named that the first three are not collinear (Figure 12.3$a$). Lines joining $P_1$ to all the other points of the set meet the line $P_2P_3$ in at most $n-1$ points (including $P_2$ and $P_3$). Let $Q$ be any *other* point on this line. Then the line $P_1Q$ contains $P_1$ but no other $P_i$.

Lines joining pairs of $P$'s meet the line $P_1Q$ in at most $\binom{n-1}{2}+1$ points (including $P_1$ and $Q$). Let $P_1A$ be one of the segments that arise in the decomposition of this line by all these points. (Possibly $A = Q$.) Then no joining line $P_iP_j$ can meet the "empty" segment $P_1A$. By its definition, $A$

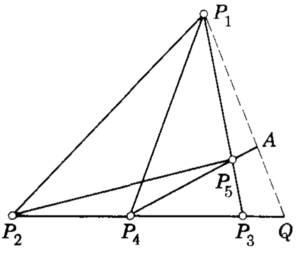



**Figure 12.3a**

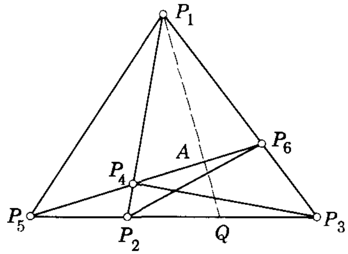



**Figure 12.3b**

lies on at least one joining line, say $P_4P_5$. If $P_4$ and $P_5$ are the only $P$'s on this line (as in Figure 12.3*a*) our task is finished. If not, we have a joining line through $A$ containing at least three of the $P$'s, which we can name $P_4$, $P_5$, $P_6$ in such an order that the segment $AP_5$ contains $P_4$ but not $P_6$. (Since $A$ decomposes the line into two opposite rays, one of which contains at least two of the three $P$'s, this special naming is always possible. See Figure 12.3*b*.) We can now prove that the line $P_1P_5$ contains only these two $P$'s.

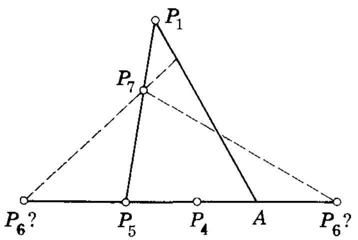



**Figure 12.3c**

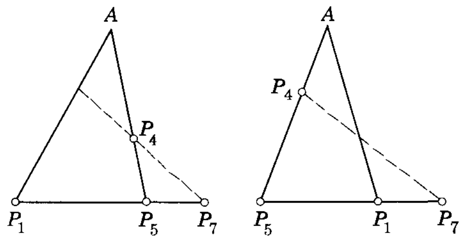



**Figure 12.3d**

We argue by *reductio ad absurdum*. If the line $P_1P_5$ contains (say) $P_7$, we can use 12.27 and 12.278 to deduce that the segment $P_1A$ meets one of the joining lines, namely, $P_6P_7$ or $P_4P_7$. In fact, it meets $P_6P_7$ if $[P_1P_7P_5]$ (as in Figure 12.3*c*), and it meets $P_4P_7$ if $[P_1P_5P_7]$ or $[P_5P_1P_7]$ (as in Figure 12.3*d*). In either case our statement about the "empty" segment is contradicted.

Thus we have found, under all possible circumstances, a line ($P_4P_5$ or $P_1P_5$) containing exactly two of the $P$'s.

### EXERCISE

Justify the statement that the joining lines meet the line $P_1Q$ in at most $\binom{n-1}{2} + 1$ points. In the example shown in Figure 12.3*b*, this number (at most 11) is only 5; why? (The symbol $\binom{i}{j}$ stands for the number of combinations of $i$ things taken $j$ at a time; for instance, $\binom{i}{2}$ is the number of pairs, namely $\frac{1}{2}i(i-1)$.)

## 12.4 PLANES AND HYPERPLANES

*If i hyperplanes in n dimensions are so placed that every n but no n + 1 have a common point, the number of regions into which they decompose the space is*

$$\binom{i}{0}+\binom{i}{1}+\binom{i}{2}+\binom{i}{3}+\cdots+\binom{i}{n}=f(n,i).$$

Ludwig Schläfli (1814-1895)
[Schläfli **1**, p. 209]

It is remarkable that we can do so much plane geometry before defining a plane. But now, as the Walrus said, "The time has come . . . . "

DEFINITIONS. If $A$, $B$, $C$ are three non-collinear points, the *plane* $ABC$ is the set of all points collinear with pairs of points on one or two sides of the triangle $ABC$. A segment, interval, ray, or line is said to be *in* a plane if all its points are.

Axioms 12.21 to 12.27 enable us to prove all the familiar properties of incidence in a plane, including the following two which Hilbert [**1**, p. 4] took as axioms:

Any three non-collinear points in a plane $\alpha$ completely determine that plane.

If two distinct points of a line $a$ lie in a plane $\alpha$, then every point of $a$ lies in $\alpha$.

DEFINITIONS. An *angle* consists of a point $O$ and two non-collinear rays going out from $O$. The point $O$ is the *vertex* and the rays are the *sides* of the angle [Veblen **1**, p. 21; Forder **1**, p. 69]. If the sides are the rays $OA$ and $OB$, or $a_1$ and $b_1$, the angle is denoted by $\angle AOB$ or $a_1b_1$ (or $\angle BOA$, or $b_1a_1$). The same angle $a_1b_1$ is determined by any points $A$ and $B$ on its respective sides. If $C$ is any point between $A$ and $B$, the ray $OC$ is said to be *within* the angle.

From here till the statement of Axiom 12.41, we shall assume that all the points and lines considered are *in one plane.*

A *convex region* is a set of points, any two of which can be joined by a segment consisting entirely of points in the set, with the extra condition that each of the points is on at least two non-collinear segments consisting entirely of points in the set. In particular, an *angular* region is the set of all points on rays within an angle, and a *triangular* region is the set of all points between pairs of points on distinct sides of a triangle. An angular (or triangular) region is said to be *bounded* by the angle (or triangle).

It can be proved [Veblen **1**, p. 21] that any line containing a point of a convex region "decomposes" it into two convex regions. In particular, a line $a$ decomposes a plane (in which it lies) into two *half planes.* Two points are said to be on the *same side* of $a$ if they are in the same half plane, on *opposite sides* if they are in opposite half planes, that is, if the segment joining them meets $a$. In the latter case we also say that $a$ *separates* the two points. (It is unfortunate that the word "side" is used with two different meanings, both well established in the literature. However, the context will always show whether we are considering the two sides of an angle, which are rays, or the two sides of a line, which are half planes.)

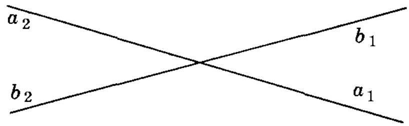



**Figure 12.4a**

As we remarked in § 12.2, any point $O$ on a line $a$ decomposes $a$ into two rays, say $a_1$ and $a_2$. Any other line $b$ through $O$ is likewise decomposed by $O$ into two rays $b_1$ and $b_2$, one in each of the half planes determined by $a$. Each of these rays decomposes the half plane containing it into two angular regions. Thus any two intersecting lines $a$ and $b$ together decompose their plane into four angular regions, bounded by the angles

$$a_1b_1, \quad b_1a_2, \quad a_2b_2, \quad b_2a_1,$$

as in Figure 12.4$a$. The opposite rays $a_1$ and $a_2$ are said to *separate* the rays $b_1$ and $b_2$; they likewise separate all the rays within either of the angles $a_1b_1$, $b_1a_2$ from all the rays within either of the angles $a_2b_2$, $b_2a_1$. We also say that the rays $a_1$ and $b_1$ separate all the rays between them from $a_2$, $b_2$, and from all the rays within $b_1a_2$, $a_2b_2$, or $b_2a_1$.

It follows from the definition of a line that two distinct points, $A$ and $B$, decompose their line into three parts: the segment $AB$ and the two rays $A/B$, $B/A$. Somewhat similarly, two nonintersecting (but coplanar) lines, $a$ and $b$, decompose their plane into three regions. One of these regions lies *between* the other two, in the sense that it contains the segment $AB$ for any $A$ on $a$ and $B$ on $b$. Another line $c$ is said to lie between $a$ and $b$ if it meets such a segment $AB$ but does not meet $a$ or $b$, and we naturally write $[acb]$.

**12.401** *If $ABC$ and $A'B'C'$ are two triads of collinear points, such that the three lines $AA'$, $BB'$, $CC'$ have no intersection, and if $[ACB]$, then $[A'C'B']$.*

Analogous consideration of an angular region yields

**12.402** *If $ABC$ and $A'B'C'$ are two triads of collinear points on distinct lines, such that the three lines $AA'$, $BB'$, $CC'$ have a common point $O$ which is not between $A$ and $A'$, nor between $B$ and $B'$, nor between $C$ and $C'$, and if $[ACB]$, then $[A'C'B']$.*

We need one or more further axioms to determine the number of *dimensions*. If we are content to work in two dimensions we say

AXIOM **12.41** *All points are in one plane.*

If not [Forder **1**, p. 60], we say instead:

AXIOM **12.42** *If ABC is a plane, there is a point D not in this plane.*

We then define the *tetrahedron ABCD*, consisting of the four non-coplanar points $A$, $B$, $C$, $D$, called *vertices*, the six joining segments $AD$, $BD$, $CD$, $BC$, $CA$, $AB$, called *edges*, and the four triangular regions $BCD$, $CDA$, $DAB$, $ABC$, called *faces*. The *space* (or "3-space") $ABCD$ is the set of all points collinear with pairs of points in one or two faces of the tetrahedron $ABCD$.

We can now deduce the familiar properties of incidence of lines and planes [Forder **1**, pp. 61–65]. In particular, any four non-coplanar points of a space determine it, and the line joining any two points of a space lies entirely in the space. If $Q$ is in the space $ABCD$ and $P$ is in a face of the tetrahedron $ABCD$, then $PQ$ meets the tetrahedron again in a point distinct from $P$.

If we are content to work in three dimensions, we say

AXIOM **12.43** *All points are in the same space.*

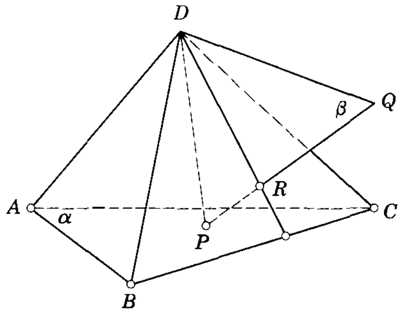



**Figure 12.4b**

Consequently:

THEOREM **12.431** *Two planes which meet in a point meet in another point, and so in a line.*

*Proof.* Let $P$ be the common point and $\alpha$ one of the planes. Take $A$, $B$, $C$ in $\alpha$ so that $P$ is inside the triangle $ABC$. Let $DPQ$ be a triangle in the other plane $\beta$ (Figure 12.4*b*). If $D$ or $Q$ lies in $\alpha$, then $\alpha$ and $\beta$ have two common points. If not, $PQ$ meets the tetrahedron $ABCD$ in a point $R$ distinct from $P$; and $DR$, in $\beta$, meets the triangle $ABC$ in a point common to $\alpha$ and $\beta$.

If, on the other hand, we wish to increase the number of dimensions, we replace 12.43 by

AXIOM **12.44** *If* $A_0A_1A_2A_3$ *is a* 3-*space, there is a point* $A_4$ *not in this* 3-*space.*

We then define the *simplex* $A_0A_1A_2A_3A_4$ which has 5 vertices $A_i$, 10 edges $A_iA_j$ $(i < j)$, 10 faces $A_iA_jA_k$ $(i < j < k)$, and 5 *cells* $A_iA_jA_kA_l$ (which are tetrahedral regions.) The 4-*space* $A_0A_1A_2A_3A_4$ is the set of points collinear with pairs of points on one or two cells of the simplex.

The possible extension to $n$ dimensions (using mathematical induction) is now clear. The $n$-space $A_0A_1 \ldots A_n$ is decomposed into two convex regions (half-spaces) by an $(n-1)$-dimensional subspace such as $A_0A_1 \ldots A_{n-1}$, which is called a *hyperplane* (or "prime," or "$(n-1)$-flat").

### EXERCISES

**1.** Any 5 coplanar points, no 3 collinear, include 4 that form a convex quadrangle.

**2.** A ray $OC$ within $\angle AOB$ decomposes the angular region into two angular regions, bounded by the angles $AOC$ and $COB$. [Veblen **1**, p. 24.]

**3.** If $m$ distinct coplanar lines meet in a point $O$, they decompose their plane into $2m$ angular regions [Veblen **1**, p. 26].

**4.** If $ABC$ is a triangle, the three lines $BC$, $CA$, $AB$ decompose their plane into seven convex regions, just one of which is triangular.

**5.** If $m$ coplanar lines are so placed that every 2 but no 3 have a common point, they decompose their plane into a certain number of convex regions. Call this number $f(2, m)$. Then

$$f(2, m) = f(2, m-1) + m.$$

But $f(2, 0) = 1$. Therefore $f(2, 1) = 2$, $f(2, 2) = 4$, $f(2, 3) = 7$, and $f(2, m) = 1 + m + \binom{m}{2}$.

**6.** If $m$ planes in a 3-space are so placed that every 3 but no 4 have a common point, they decompose their space into (say) $f(3, m)$ convex regions. Then

$$f(3, m) = f(3, m-1) + f(2, m-1).$$

But $f(3, 0) = 1$. Therefore $f(3, 1) = 2$, $f(3, 2) = 4$, $f(3, 3) = 8$, $f(3, 4) = 15$, and $f(3, m) = 1 + \binom{m}{1} + \binom{m}{2} + \binom{m}{3}$. [Steiner **1**, p. 87.]

**7.** Obtain the analogous result for $m$ hyperplanes in an $n$-space.

## 12.5 CONTINUITY

> *Nothing but Geometry can furnish a thread for the labyrinth of the composition of the continuum . . . and no one will arrive at a truly solid metaphysic who has not passed through that labyrinth.*
>
> G. W. Leibniz (1646-1716)
> [Russell **2**, pp. 108-109]

Between any two rational numbers (§ 9.1) there is another rational number, and therefore an infinity of rational numbers; but this does not mean that every real number (§ 9.2) is rational. Similarly, between any two points

(12.271) there is another point, and therefore an infinity of points; but this does not mean that the axioms in § 12.2 make the line "continuous." In fact, continuity requires at least one further axiom. There are two well-recognized approaches to this subtle subject. One, due to Cantor and Weierstrass, defines a monotonic sequence of points, with an axiom stating that *every bounded monotonic sequence has a limit* [Coxeter **2**, Axiom 10.11]. The other, due to Dedekind, obtains a general point on a line as the common origin of two opposite rays [Coxeter **3**, p. 162]. Its arithmetical counterpart is illustrated by describing $\sqrt{2}$ as the "section" between rational numbers whose squares are less than 2 and rational numbers whose squares are greater than 2. Dedekind's Axiom, though formidable in appearance, is the more readily applicable; so we shall use it here:

AXIOM **12.51** *For every partition of all the points on a line into two nonempty sets, such that no point of either lies between two points of the other, there is a point of one set which lies between every other point of that set and every point of the other set.*

This axiom is easily seen to imply several modified versions of the same statement. Instead of "the points on a *line*" we could say "the points on a *ray*" or "the points on a *segment*" or "the points on an *interval*." (In the last case, for instance, the rest of the line consists of two rays which can be added to the two sets in an obvious manner.) Another version [Forder **1**, p. 299] is:

THEOREM **12.52** *For every partition of all the rays within an angle into two nonempty sets, such that no ray of either lies between two rays of the other, there is a ray of one set which lies between every other ray of that set and every ray of the other set.*

To prove this for an angle $\angle AOB$, we consider the section of all the rays by the line $AB$, and apply the "segment" version of 12.51 to the segment $AB$.

## 12.6 PARALLELISM

*In the last few weeks I have begun to put down a few of my own Meditations, which are already to some extent nearly 40 years old. These I had never put in writing, so I have been compelled three or four times to go over the whole matter afresh in my head.*

C. F. Gauss (1777-1855)
(Letter to H. K. Schumacher, May 17, 1831, as translated by Bonola [**1**, p. 67])

The idea of defining, through a given point, two rays parallel to a given line (in opposite senses), was developed independently by Gauss, Bolyai, and Lobachevsky. The following treatment is closest to that of Gauss.

THEOREM **12.61** *For any point $A$ and any line $r$, not through $A$, there are just two rays from $A$, in the plane $Ar$, which do not meet $r$ and which separate all the rays from $A$ that meet $r$ from all the other rays that do not.*

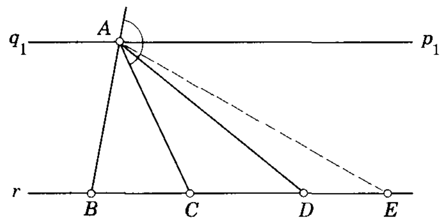



**Figure 12.6a**

*Proof.* Taking any two distinct points $B$ and $C$ on $r$, we apply 12.52 to the angle between the rays $AC$ and $A/B$ (marked in Figure 12.6*a*). We consider the partition of all the rays within this angle into two sets according as they do or do not meet the ray $C/B$. Clearly, these sets are not empty, and no ray in either set lies between two in the other. We conclude that one of the sets contains a special ray $p_1$ which lies between every other ray of that set and every ray of the other set.

In fact, $p_1$ belongs to the second set. For, if it met $C/B$, say in $D$, we would have $[BCD]$. By Axiom 12.22, we could take a point $E$ such that $[CDE]$, with the absurd conclusion that $AE$ belongs to both sets: to the first, because $E$ is on $C/B$, and to the second, because $AD$ lies between $AC$ and $AE$.

We have thus found a ray $p_1$, within the chosen angle, which is the "first" ray that fails to meet the ray $C/B$; this means that every ray within the angle between $AC$ and $p_1$ does meet $C/B$. Interchanging the roles of $B$ and $C$, we obtain another special ray $q_1$, on the other side of $AB$, which may be described (for a counterclockwise rotation) as the "last" ray that fails to meet $B/C$. Since the line $r$ consists of the two rays $B/C$, $C/B$, along with the interval $\overline{BC}$, we have now found two rays $p_1$, $q_1$, which separate all the rays from $A$ that meet $r$ from all the other rays (from $A$) that do not. [Forder **1**, p. 300.]

These special rays from $A$ are said to be *parallel* to the line $r$ in the two senses: $p_1$ parallel to $C/B$, and $q_1$ parallel to $B/C$. (Two rays are said to have the same sense if they lie on the same side of the line joining their initial points.)

For the sake of completeness, we define the rays parallel to $r$ from a point $A$ on $r$ itself to be the two rays into which $A$ decomposes $r$. The distinction between affine geometry and hyperbolic geometry depends on the question whether, for other positions of $A$, the two rays $p_1$, $q_1$ are still the two halves of one line. If they are, this line decomposes the plane into two half planes, one of which contains the whole of the line $r$. If not, the lines $p$ and $q$ (which contain the rays) decompose the plane into four angular regions

$$p_1q_1, \quad q_1p_2, \quad p_2q_2, \quad q_2p_1.$$

In this case, by 12.61, $r$ lies entirely in the region $p_1q_1$.

**COROLLARY 12.62** *For any point A and any line r, not through A, there is at least one line through A, in the plane Ar, which does not meet r.*

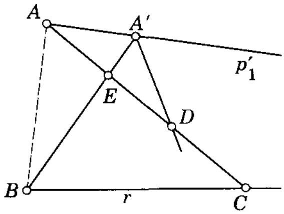



**Figure 12.6b**

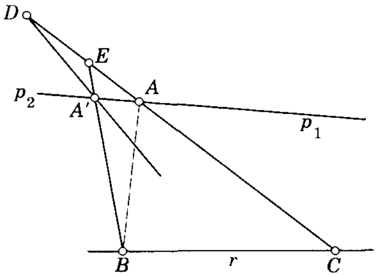



**Figure 12.6c**

Another familiar property of parallelism is its "transmissibility":

**THEOREM 12.63** *The parallelism of a ray and a line is maintained when the beginning of the ray is changed by the subtraction or addition of a segment.*

*Proof* [Gauss **1**, vol. 8, p. 203]. Let $p_1$ be a ray from $A$ which is parallel to a line $r$ through $B$, and let $A'$ be any point on this ray (Figure 12.6*b*) or on the opposite ray $p_2$ (Figure 12.6*c*). The modified ray $p'_1$, beginning at $A'$, is $A'/A$ or $A'A$, respectively; it obviously does not meet $r$. What remains to be proved is that every ray from $A'$, within the angle between $A'B$ and $p'_1$, does meet $r$. Let $D$ be any point on such a ray (Figure 12.6*b*) or on its opposite (Figure 12.6*c*). Since $p_1$ (from $A$) is parallel to $r$, the line $AD$ (containing a ray within the angle between $AB$ and $p_1$) meets $r$, say in $C$. The line $A'B$, separating $A$ from $D$, meets the segment $AD$, say in $E$. By Axiom 2.27, applied to the triangle $CBE$ with $[BEA']$ and $[EDC]$ (Figure 12.6*b*) or to the triangle $BCE$ with $[CED]$ and $[EA'B]$ (Figure 12.6*c*), the line $A'D$ meets $BC$. Thus $p'_1$ is parallel to $r$.

This property of transmissibility enables us to say that the *line* $p = \overline{AA'}$ is parallel to the line $r = \overline{BC}$, provided we remember that this property is associated with a definite "sense" along each line.

Busemann [**1**, p. 139 (23.5)] has proved that it is not possible, within the framework of two-dimensional ordered geometry, to establish the "symmetry" of parallelism: that if $p$ is parallel to $r$ then $r$ is parallel to $p$. To supply this important step we need either Axiom 12.42 [as in Coxeter **3**, pp. 165–177] or the affine axiom of parallelism (13.11) or the absolute axioms of congruence (§ 15.1).

THEOREM **12.64** *If two lines are both parallel to a third in the same sense, there is a line meeting all three.*

*Proof.* We have to show that, if lines $p$ and $s$ are both parallel to $r$ in the same sense, then the three lines $p$, $r$, $s$ have a transversal. In affine geometry this is obvious, so let us assume the geometry to be hyperbolic. Of the two lines parallel to $r$ through a point $A$ on $p$, one is $p$ itself. Let $q$ be the other, and let $r$ be in the angular region $\mathrm{p}_1\mathrm{q}_1$, so that the rays $\mathrm{p}_1$ and $\mathrm{q}_1$ (from $A$) are parallel to $r$ in opposite senses and $s$ is parallel to $r$ in the same sense as $\mathrm{p}_1$. Let $B$ and $D$ be arbitrary points on $r$ and $s$, respectively.

If $D$ is in the region $\mathrm{p}_1\mathrm{q}_1$, the line $AD$ is a transversal. If $D$ is in $\mathrm{p}_1\mathrm{q}_2$, $BD$ is a transversal. If $D$ is in $\mathrm{p}_2\mathrm{q}_2$, both $AD$ and $BD$ are transversals. Finally, if $D$ is in $\mathrm{p}_2\mathrm{q}_1$, $AB$ is a transversal.

Hyperbolic geometry will be considered further in Chapters 15, 16, and 20.

## EXERCISES

**1.** If $p$ is parallel to $s$ and $[prs]$, then $p$ is parallel to $r$. (See Figure 15.2*c* with $s$ for $q$.)

**2.** Consider all the points strictly inside a given circle in the Euclidean plane. Regard all other points as nonexistent. Let chords of the circle be called lines. Then all the axioms 12.21–12.27, 12.41, and 12.51 are satisfied. Locate the two rays through a given point parallel to a given line. Note that they form an angle (as in Figure 16.2*b*).

# 13
# *Affine geometry*

The first three sections of this chapter contain a systematic development of the foundations of affine geometry. In particular, we shall see how length may be measured along a line, though independent units are required for lines in different directions. In §§ 13.4–7 we shall investigate such topics as area, affine transformations, lattices, vectors, barycentric coordinates, and the theorems of Ceva and Menelaus. Finally, in § 13.8 and § 13.9, we shall extend these ideas from two dimensions to three.

According to Blaschke [**1**, p. 31; **2**, p. 12], the word "affine" (German *affin*) was coined by Euler. But it was only after the launching of Klein's Erlangen program (see Chapter 5) that this geometry became recognized as a self-contained discipline. Many of the propositions may seem familiar; in fact, most readers will discover that they have often been working in the affine plane without realizing that it could be so designated.

Our treatment is somewhat more geometric and less algebraic than that of Artin's *Geometric Algebra* [Artin **1**; see especially pp. 58, 63, 71]. Incidentally, we shall find that our Axiom 13.12 (which he calls DP) implies Theorem 13.122 (his $D_a$): this presumably means that his Axiom 4b implies 4a.

## 13.1 THE AXIOM OF PARALLELISM AND THE "DESARGUES" AXIOM

> *Mathematical language is difficult but imperishable. I do not believe that any Greek scholar of to-day can understand the idiomatic undertones of Plato's dialogues, or the jokes of Aristophanes, as thoroughly as mathematicians can understand every shade of meaning in Archimedes' works.*
>
> M. H. A. Newman
>
> (*Mathematical Gazette* **43,** 1959, p. 167)

In this axiomatic treatment, we regard the real affine plane as a special case of the ordered plane. Accordingly, the primitive concepts are *point* and *intermediacy,* satisfying Axioms 12.21–12.27, 12.41 and 12.51. Affine geometry is derived from ordered geometry by adding the following two extra axioms:

**AXIOM 13.11** *For any point $A$ and any line $r$, not through $A$, there is at most one line through $A$, in the plane $Ar$, which does not meet $r$.*

**AXIOM 13.12** *If $A, A', B, B', C, C', O$ are seven distinct points, such that $AA', BB', CC'$ are three distinct lines through $O$, and if the line $AB$ is parallel to $A'B'$, and $BC$ to $B'C'$, then also $CA$ is parallel to $C'A'$.*

The affine axiom of parallelism (13.11) combines with 12.62 to tell us that, for any point $A$ and any line $r$, there is exactly one line through $A$, in the plane $Ar$, which does not meet $r$. Hence the two rays from $A$ parallel to $r$ are always collinear, *any two lines in a plane that do not meet are parallel,* and parallelism is an *equivalence relation.* The last remark comprises three properties:

Parallelism is *reflexive.* (Each line is parallel to itself.)
Parallelism is *symmetric.* (If $p$ is parallel to $r$, then $r$ is parallel to $p$.)
Parallelism is *transitive.* (If $p$ and $q$ are parallel to $r$, then $p$ is parallel to $q$. Euclid I. 30.)

In the manner that is characteristic of equivalence relations, every line belongs to a *pencil* of parallels whose members are all parallel to one another.

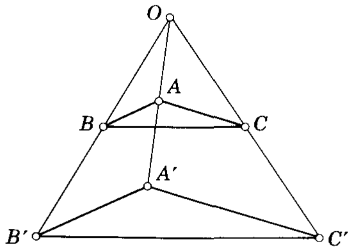


**Figure 13.1a**

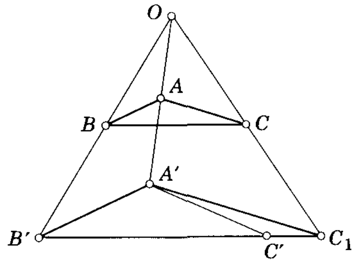


**Figure 13.1b**

Axiom 13.12 (see Figure 13.1*a*) is probably familiar to most readers either as a corollary of Euclid VI.2 or as an affine form of Desargues's theorem. We shall see that it implies

**THEOREM 13.121** *If $ABC$ and $A'B'C'$ are two triangles with distinct vertices, so placed that the line $BC$ is parallel to $B'C'$, $CA$ to $C'A'$, and $AB$ to $A'B'$, then the three lines $AA'$, $BB'$, $CC'$ are either concurrent or parallel.*

*Proof.* If the three lines $AA'$, $BB'$, $CC'$ are not all parallel, some two of them must meet. The notation being symmetrical, we may suppose that these two are $AA'$ and $BB'$, meeting in $O$, as in Figure 13.1*b*. Let $OC$ meet $B'C'$ in $C_1$. By Axiom 13.12, applied to $AA'$, $BB'$, $CC_1$, the line $AC$ is parallel to $A'C_1$ as well as to $A'C'$. By Axiom 13.11, $C_1$ lies on $A'C'$ as well as on $B'C'$. Since $A'B'C'$ is a triangle, $C_1$ coincides with $C'$. Thus, if $AA'$, $BB'$, $CC'$ are not parallel, they are concurrent [Forder **1**, p. 158].

Roughly speaking, Axiom 13.12 is the converse of one half of Theorem 13.121. The converse of the other half is

**THEOREM 13.122** *If $A$, $A'$, $B$, $B'$, $C$, $C'$ are six distinct points on three distinct parallel lines $AA'$, $BB'$, $CC'$, so placed that the line $AB$ is parallel to $A'B'$, and $BC$ to $B'C'$, then also $CA$ is parallel to $C'A'$.*

*Proof.* Through $A'$ draw $A'C_1$ parallel to $AC$, to meet $B'C'$ in $C_1$, as in Figure 13.1*c*. By 13.121, applied to the triangles $ABC$ and $A'B'C_1$, since $AA'$ and $BB'$ are parallel, $CC_1$ is parallel to both of them, and therefore also to $CC'$. Hence $C_1$ lies on $CC'$ as well as on $B'C'$. Since the parallel lines $BB'$ and $CC'$ are distinct, $B'$ cannot lie on $CC'$. Therefore $C_1$ coincides with $C'$, and $A'C'$ is parallel to $AC$.

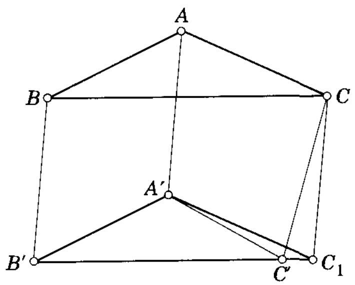



**Figure 13.1c**

**EXERCISES**

1. If a line in the plane of two parallel lines meets one of them, it meets the other also.

2. Can we always say, of three distinct parallel lines, that one lies between the other two?

## 13.2 DILATATIONS

*Dilatations . . . are one-to-one maps of the plane onto itself which move all points of a line into points of a parallel line.*

E. Artin [**1**, p. 51]

Four non-collinear points $A$, $B$, $C$, $D$ are said to form a *parallelogram* $ABCD$ if the line $AB$ is parallel to $DC$, and $BC$ to $AD$. Its *vertices* are the four points; its *sides* are the four segments $AB$, $BC$, $CD$, $DA$, and its *diagonals* are the two segments $AC$, $BD$. Since $B$ and $D$ are on opposite sides of $AC$, the diagonals meet in a point called the *center* [Forder **1**, p. 140].

As in § 5.1, we define a *dilatation* to be a transformation which transforms each line into a parallel line. But now we must discuss more thoroughly the important theorem 5.12, which says that *two given segments, $AB$ and $A'B'$, on parallel lines, determine a unique dilatation $AB \to A'B'$.*

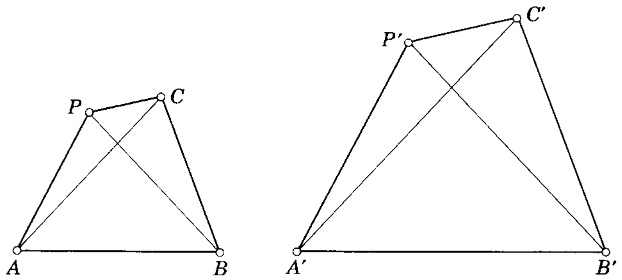



**Figure 13.2a**

For any point $P$, not on $AB$, we can find a corresponding point $P'$ by drawing $A'P'$ parallel to $AP$, and $B'P'$ parallel to $BP$, as in Figure 5.1*a*. (The lines thus drawn through $A'$ and $B'$ cannot be parallel, for, if they were, $AP$ and $BP$ would be parallel.) Similarly, another point $C$ yields $C'$, as in Figure 13.2*a*. By 13.121, the three lines $AA'$, $BB'$, $CC'$ are either concurrent or parallel. So likewise are $AA'$, $BB'$, $PP'$.

If the two parallel lines $AB$ and $A'B'$ do not coincide, it follows that the four lines $AA'$, $BB'$, $CC'$, $PP'$ are all either concurrent or parallel. Then, by 13.12 or 13.122 (respectively), $CP$ and $C'P'$ are parallel, so that the transformation is indeed a dilatation. If the lines $AB$ and $A'B'$ do coincide, we can reach the same conclusion by regarding the transformation as $AC \to A'C'$ instead of $AB \to A'B'$.

We see now that a given dilatation may be specified by its effect on any given segment. The *inverse* of the dilatation $AB \to A'B'$ is the dilatation $A'B' \to AB$. The *product* of two dilatations, $AB \to A'B'$ and $A'B' \to A''B''$, is the dilatation $AB \to A''B''$. In particular, the product of a dilatation with its inverse is the *identity*, $AB \to AB$. Thus all the dilatations together form a (continuous) *group*.

The argument used in proving 5.13 shows that, for a given dilatation, the lines $PP'$ which join pairs of corresponding points are *invariant* lines. The discussion of 5.12 shows that all these lines are either concurrent or parallel.

If the lines $PP'$ are concurrent, their intersection $O$ is an invariant point, and we have a *central* dilatation

$$OA \to OA'$$

(where $A'$ lies on the line $OA$). The invariant point $O$ is unique; for, if $O$ and $O_1$ were two such, the dilatation would be $OO_1 \to OO_1$, which is the identity.

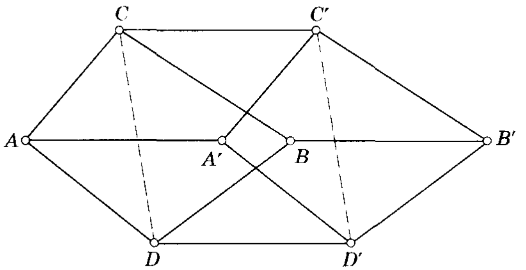



**Figure 13.2b**

If, on the other hand, the lines $PP'$ are parallel, there is no invariant point, and we have a *translation* $AB \to A'B'$, where not only is $AB$ parallel to $A'B'$ but also $AA'$ is parallel to $BB'$. If these two parallel lines are distinct, $AA'B'B$ is a parallelogram. If not, we can use auxiliary parallelograms $AA'C'C$ and $C'CBB'$ (or $AA'D'D$ and $D'DBB'$) as in Figure 13.2*b*. Two applications of 13.122 suffice to prove that, when $A$, $B$, $A'$ are given, $B'$ is independent of the choice of $C$ (or $D$). Hence

**13.21** *Any two points $A$ and $A'$ determine a unique translation $A \to A'$.*

We naturally include, as a degenerate case, the identity, $A \to A$. It follows that a dilatation, other than the identity, is a translation if and only if it has no invariant point. Moreover, a given translation may be specified by its effect on any given point; in fact, the translation $A \to A'$ is the same as $B \to B'$ if $AA'B'B$ is a parallelogram, or if, for any parallelogram $AA'C'C$ based on $AA'$, there is another parallelogram $C'CBB'$.

We next prove that dilatations are "ordered transformations:"

**13.22** *The dilatation $AB \to A'B'$ transforms every point between $A$ and $B$ into a point between $A'$ and $B'$.*

*Proof.* If the lines $AB$ and $A'B'$ are distinct, the fact that $[ACB]$ implies $[A'C'B']$ follows at once from 12.401 (for a translation) or 12.402 (for a central dilatation). To obtain the analogous result for two corresponding triads on an invariant line $CC'$, we draw six parallel lines through the six points, as in Figure 13.2*c*, and use the fact that $[acb]$ implies $[a'c'b']$.

To prove Theorem 3.21, which says that *the product of two translations is a translation,* we can argue thus: since translations are dilatations, the product is certainly a dilatation. If it is not a translation it has a unique invariant point $O$. If the first of the two given translations takes $O$ to $O'$, the second must take $O'$ back to $O$. But the translation $O' \to O$ is the inverse of $O \to O'$. Thus the only case in which the product of two translations has an invariant point is when one of the translations is the inverse of the other. (By our convention, the product is still a translation even then.) Hence

**13.23** *The product of two translations $A \to B$ and $B \to C$ is the translation $A \to C$.*

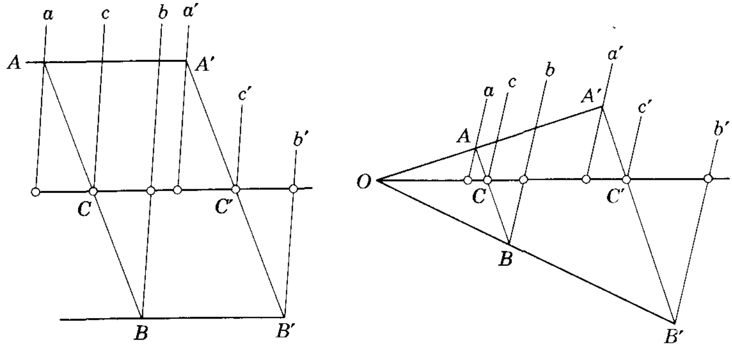



**Figure 13.2c**

To prove that this is a *commutative* product (as in 3.23), we consider first the easy case in which the two translations are along nonparallel lines. Completing the parallelogram $ABCD$, we observe that the translations $A \to B$ and $B \to C$ are the same as $D \to C$ and $A \to D$, respectively. Hence their product in either order is the translation $A \to C$:

$$\begin{aligned}(A \to B)(B \to C) &= (A \to D)(D \to C) \\ &= (B \to C)(A \to B).\end{aligned}$$

To deal with the product of two translations T and X along the same line, let Y be any translation along a nonparallel line, so that X commutes with both Y and TY. Then

$$\text{TXY} = \text{TYX} = \text{XTY}$$

and therefore
$$\text{TX} = \text{XT}$$

[cf. Veblen and Young **2**, p. 76].

As a special case of 5.12, we see that any two distinct points, $A$ and $B$, are interchanged by a unique dilatation $AB \to BA$, or, more concisely,

$$A \leftrightarrow B,$$

which we call a *half-turn*. (Of course, $A \leftrightarrow B$ is the same as $B \leftrightarrow A$.) If $C$ is any point outside the line $AB$, the half-turn transforms $C$ into the point $D$ in which the line through $B$ parallel to $AC$ meets the line through $A$ parallel to $BC$ (Figure 13.2*d*). Therefore $ADBC$ is a parallelogram, and the same half-turn can be expressed as $C \leftrightarrow D$. The invariant lines $AB$ and $CD$, being the diagonals of the parallelogram, intersect in a point $O$, which is the invariant point of the half-turn. It follows that any segment $AB$ has a *midpoint* which can be defined to be the invariant point of the half-turn $A \leftrightarrow B$, and we have proved that the center of a parallelogram is the midpoint of each diagonal, that is, that the two diagonals "bisect" each other. To see how the half-turn transforms an arbitrary point on $AB$, we merely have to join this point to $C$ (or $D$) and then draw a parallel line through $D$ (or $C$).

By considering their effect on an arbitrary point $B$, we may express any two half-turns as $A \leftrightarrow B$ and $B \leftrightarrow C$. If their product has an invariant point $O$, each of them must be expressible in the form $O \leftrightarrow O'$, that is, they must coincide. In every other case, there is no invariant point. Hence

**13.24** *The product of two half-turns $A \leftrightarrow B$ and $B \leftrightarrow C$ is the translation $A \to C$.*

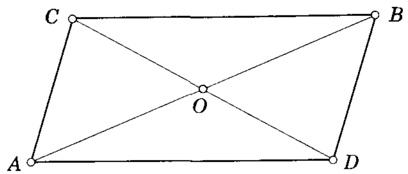



**Figure 13.2d**

We have seen (Figure 13.2*d*) that, if $ADBC$ is a parallelogram, the half-turn $A \leftrightarrow B$ is the same as $C \leftrightarrow D$, and the translation $A \to D$ is the same as $C \to B$. This connection between half-turns and translations remains valid when the parallelogram collapses to form a symmetrical arrangement of four collinear points, as in Figure 13.2*e*:

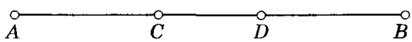



**Figure 13.2e**

**13.25** *The half-turns $A \leftrightarrow B$ and $C \leftrightarrow D$ are equal if and only if the translations $A \to D$ and $C \to B$ are equal.*

In fact, the relation $(A \leftrightarrow B) = (C \leftrightarrow D)$ implies

$$\begin{aligned}(A \to D) &= (A \leftrightarrow B)(B \leftrightarrow D)\\ &= (C \leftrightarrow D)(D \leftrightarrow B) = (C \to B)\end{aligned}$$

and, conversely, the relation $(A \to D) = (C \to B)$ implies

$$\begin{aligned}(A \leftrightarrow B) &= (A \to D)(D \leftrightarrow B)\\ &= (C \to B)(B \leftrightarrow D) = (C \leftrightarrow D).\end{aligned}$$

In the special case when $C$ and $D$ coincide, we call them $C'$ and deduce that $C'$ is the midpoint of $AB$ if and only if the translations $A \to C'$ and $C' \to B$

are equal. This involves the existence of parallelograms $AC'A'B'$ and $A'B'C'B$, as in Figure 13.2*f*. Completing the parallelogram $B'C'A'C$, we obtain a triangle $ABC$ with $A'$, $B'$, $C'$ at the midpoints of its sides. Hence

**13.26** *The line joining the midpoints of two sides of a triangle is parallel to the third side, and the line through the midpoint of one side parallel to another passes through the midpoint of the third.*

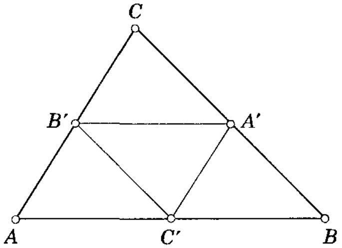



**Figure 13.2f**

Two figures are said to be *homothetic* if they are related by a dilatation, *congruent* if they are related by a translation or a half-turn. In particular, a directed segment $AB$ is congruent to its "opposite" segment $BA$ by the half-turn $A \leftrightarrow B$. Thus, in Figure 13.2*f*, the four small triangles

$$AC'B', \quad C'BA', \quad B'A'C, \quad A'B'C'$$

are all congruent, and each of them is homothetic to the large triangle $ABC$.

## EXERCISES

1. Such equations as those used in proving 13.25 are easily written down if we remember that each must involve an even number of double-headed arrows (indicating half-turns). Explain this rule.

2. The translations $A \to C$ and $D \to B$ are equal if the translations $A \to D$ and $C \to B$ are equal. (This is obvious when $ADBC$ is a parallelogram, but remarkable when all the points are collinear.)

3. Setting $A = C$ in the equation

$$(A \leftrightarrow B)(B \to C) = (A \leftrightarrow C),$$

deduce that any given point $C$ is the invariant point of a half-turn $(C \leftrightarrow B)(B \to C)$ which, by a natural extension of the symbolism, may be written as

$$C \leftrightarrow C.$$

4. If the three diagonals of a hexagon (not necessarily convex) all have the same midpoint, any two opposite sides are parallel (as in Figure 4.1*e*).

5. From any point $A_1$ on the side $BC$ of a triangle $ABC$, draw $A_1B_1$ parallel to $BA$ to meet $CA$ in $B_1$, then $B_1C_1$ parallel to $CB$ to meet $AB$ in $C_1$, and then $C_1A_2$ parallel to $AC$ to meet $BC$ in $A_2$. If $A_1$ is the midpoint of $BC$, $A_2$ coincides with it. If not, continue the process, drawing $A_2B_2$ parallel to $BA$, $B_2C_2$ parallel to $CB$, and $C_2A_3$ parallel to $AC$. The path is now closed: $A_3$ coincides with $A_1$. (This is called Thom-

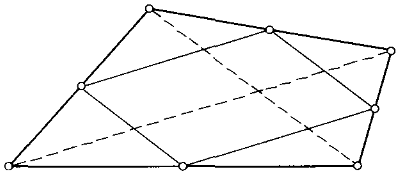


**Figure 13.2g**

sen's figure. See Geometrical Magic, by Nev R. Mind, *Scripta Mathematica,* **19** (1953), pp. 198–200.)

**6.** The midpoints of the four sides of any simple quadrangle are the vertices of a parallelogram (Figure 13.2*g*; cf. Figure 4.2*c*). This theorem was discovered by Pierre Varignon (1654–1722). It shows that the *bimedians,* which join the midpoints of opposite sides of the quadrangle, bisect each other. Thus the corollary to Hjelmslev's theorem (§ 3.6) becomes an affine theorem when we replace the hypotheses 3.61 by

$$AB = BC, \quad A'B' = B'C'.$$

**7.** The midpoints of the six sides of any complete quadrangle are the vertices of a centrally symmetrical hexagon (of the kind considered in Ex. 4, above).

## 13.3 AFFINITIES

*"Yes, indeed," said the Unicorn, . . . "What can we measure? . . . We are experts in the theory of measurement, not its practice."*

J. L. Synge [**2,** p. 51]

The results of § 13.2 may be summarized in the statement that all the translations of the affine plane form a continuous Abelian group, which is a subgroup of index 2 in the group of translations and half-turns; and the latter is a subgroup (of infinite index) in the group of dilatations [Veblen and Young **2,** pp. 79, 93].

Moreover, the group of translations is a *normal* subgroup (or "self-conjugate" subgroup)* in the group of dilatations, that is, if T is a translation while S is a dilatation, then $S^{-1}TS$ is a translation [Artin **1,** p. 57]. To prove this, suppose if possible that the dilatation $S^{-1}TS$ has an invariant point. Since this invariant point could have been derived from a suitable point $O$ by applying S, we may denote it by $O^S$. Thus $S^{-1}TS$ leaves $O^S$ invariant. But $S^{-1}TS$ transforms $O^S$ into $O^{TS}$. Hence $O^{TS} = O^S$. Applying $S^{-1}$, we deduce $O^T = O$, which is absurd (since T has no invariant point).

If T is $A \to B$ and S is $AB \to A^S B^S$, then $S^{-1}TS$ is $A^S \to B^S$. Accordingly, it is sometimes convenient to write $T^S$ for $S^{-1}TS$ [see, e.g., Coxeter **1,** p. 39] and to say that the dilatation S *transforms* the translation T into the translation $T^S$. (Since $A^S B^S$ is parallel to $AB$, $T^S$ has the same direction as T.) In other words, a dilatation transforms the group of translations into

* Birkhoff and MacLane **1,** p. 141; Coxeter **1,** p. 42.

itself in the manner of an *automorphism:* if it transforms T into $T^S$ and another translation U into $U^S$, it transforms the product TU into $(TU)^S = T^S U^S$ and any power of T into the same power of $T^S$.

It is convenient to use the italic letter $T$ for the point into which the translation T transforms an arbitrarily chosen initial point (or origin) $1$. Then, if a central dilatation S has $1$ as its invariant point, it not only transforms T into $T^S$ but also transforms $T$ into $T^S$.

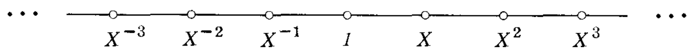



**Figure 13.3a**

Applying to the arbitrary point $1$ all the integral powers of a given translation X, we obtain a *one-dimensional lattice* consisting of infinitely many points "evenly spaced" along a line, as in Figure 13.3*a*. We may regard every such point $X^\mu$ as being derived from the point $X$ by a dilatation $1X \to 1X^\mu$ (which leaves the point $1$ invariant). At first we take $\mu$ to be an integer; but since the same dilatation transforms each $X^n$ into

$$(X^\mu)^n = X^{\mu n},$$

we can consistently extend the meaning of $X^\mu$ so as to allow $\mu$ to have any rational value, and finally any real value. In other words, we can interpolate new points between the points of the one-dimensional lattice and then define $X^\mu$, for any real $\mu$, to mean the translation $1 \to X^\mu$. The details are as follows.

For each rational number $\mu = a/b$ (where $a$ is an integer and $b$ is a positive integer) we derive from the point $X$ a new point $X^\mu$ by means of the dilatation $1X^b \to 1X^a$. A convenient way to construct this point $X^\mu$ is to use the lattice of powers of an arbitrary translation Y along another line through the initial point $1$, drawing a line through the point $Y$ parallel to the join of the points $Y^b$ and $X^a$, as in Figure 13.3*b* (cf. Figure 9.1*c*).

To verify that the order of such points $X^\mu$ agrees with the order of the rational numbers $\mu$, we take three of them and reduce their $\mu$'s to a common denominator so as to express them as $X^{a_1/b}$, $X^{a_2/b}$, $X^{a_3/b}$. If $a_1 < a_2 < a_3$, so that $[X^{a_1} X^{a_2} X^{a_3}]$, we can apply 13.22 to the dilatation $1X^b \to 1X$, with the conclusion that

$$[X^{a_1/b}\, X^{a_2/b}\, X^{a_3/b}].$$

If $\mu$ is irrational, we define $X^\mu$ to be the Dedekind section between all the rational points $X^{a/b}$ for which $a/b < \mu$ and all those for which $a/b > \mu$. More precisely, supposing for definiteness that $\mu$ is positive, we apply the "ray" version of 12.51 to two sets of points, one consisting of all the points

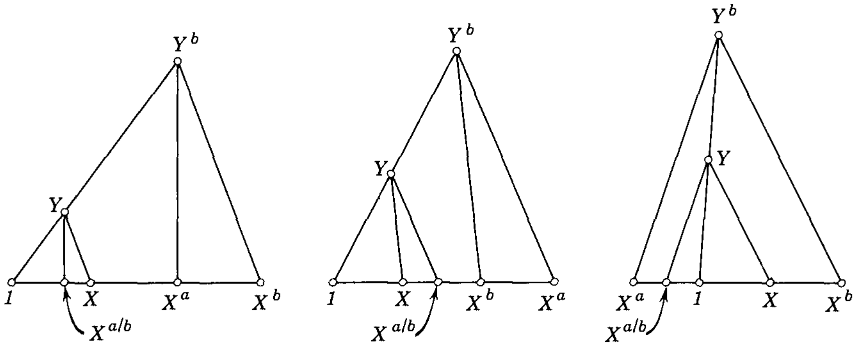



**Figure 13.3b**

whose exponents are positive rational numbers less than $\mu$, and all the points between pairs of these, whereas the other set consists of the rest of the "positive" ray $1X$. (If $\mu$ is negative, we make the same kind of partition of the "negative" ray $1/X$.) Finally $X^\mu$ is, by definition, the translation $1 \to X^\mu$.

We have now interpreted the symbol $X^\mu$ for all real values of $\mu$ (including 0 and 1, which yield $X^0 = 1$ and $X^1 = X$). Conversely, *every point on the line 1X can be expressed in the form* $X^\mu$.

This is obvious for any point of the interval from $X^{-1}$ to $X$. Any other point $T$ satisfies either $[1\ X\ T]$ or $[1\ X^{-1}\ T]$. If $[1\ X\ T]$, the dilatation $1T \to 1X$ transforms $X$ into a point between $1$ and $X$, say $X^\lambda$. The inverse dilatation $1X^\lambda \to 1X$ transforms $X$ into $X^{1/\lambda}$; therefore $T = X^{1/\lambda}$. If, on the other hand, $[1\ X^{-1}\ T]$, we make the analogous use of $1T \to 1X^{-1}$. In either case we obtain an expression for T as a power of X.

Thus, assuming Dedekind's axiom, we have proved the "axiom of Archimedes":

**13.31** *For any point T (except 1) on the line of a translation* X, *there is an integer n such that T lies between the points 1 and* $X^n$.

The exponent $\mu$ provides a measure of distance along the line $1X$. In fact, the segment $X^\nu X^\mu$ ($\nu < \mu$) is said to have *length* $\mu - \nu$ in terms of the segment $1X$ as unit:

$$\frac{X^\nu X^\mu}{1X} = \mu - \nu.$$

Along another line $1Y$ (Figure 13.3*c*) we have an independent unit. Since the dilatation $1X \to 1X^\mu$ transforms the point $Y$ into $Y^\mu$, where the line $X^\mu Y^\mu$ is parallel to $XY$, we have

$$\frac{1X^\mu}{1X} = \frac{1Y^\mu}{1Y}$$

in agreement with Euclid VI.2 (see § 1.3). Thus we can define ratios of the lengths on one line, or on parallel lines, and we can compare such ratios on different lines. But affine geometry contains no machinery for comparing lengths in different directions: it is a meaningless question whether the translation Y is longer or shorter than X.

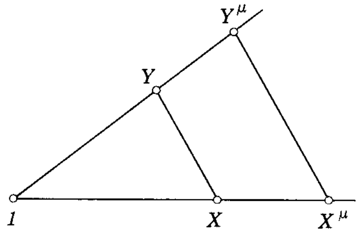



**Figure 13.3c**

The above definition for the length of the segment $X^\nu X^\mu$ ($\nu < \mu$) suggests the propriety of allowing the oppositely directed segment $X^\mu X^\nu$ to have the negative length $\nu - \mu$. This convention enables us to write $\mu = 1X^\mu/1X$ for negative as well as positive values of $\mu$, and to add lengths of collinear segments according to such formulas as

$$AB + BC = AC, \qquad BC + CA + AB = 0,$$

regardless of the order of their end points $A$, $B$, $C$.

Now, to set up a system of *affine coordinates* in the plane, we let $(x, y)$ denote the point into which the origin $1$ is transformed by the translation $X^x Y^y$. This simple device establishes a one-to-one correspondence between points in the plane and ordered pairs of real numbers. In particular, the point $X^x$ is $(x, 0)$, $Y^y$ is $(0, y)$, and the origin itself is $(0, 0)$. When $x$ and $y$ are integers, the points $(x, y)$ form a *two-dimensional lattice,* as in Figure 4.1*b*. The remaining points $(x, y)$ are distributed between the lattice points in the obvious manner.

In affine coordinates (as in Cartesian coordinates) a line has a linear equation. The powers of the translation $X^{-b}Y^a$ transform the origin into the points $(-\mu b, \mu a)$ whose locus is the line $ax + by = 0$. The same powers transform $(x_1, y_1)$ into the points

$$(x_1 - \mu b, y_1 + \mu a)$$

whose locus is

$$a(x - x_1) + b(y - y_1) = 0.$$

We can thus express a line in any of the standard forms 8.11, 8.12, 8.13.

A dilatation is a special case of an *affinity,* which is any transformation (of the whole affine plane onto itself) preserving collinearity. Thus, an affinity transforms parallel lines into parallel lines, and preserves ratios of distances along parallel lines. It also preserves intermediacy (compare 13.22).

**13.32** *An affinity is uniquely determined by its effect on any one triangle.*

For, if it transforms a triangle $IXY$ into $I'X'Y'$, it transforms the point $(x, y)$ referred to the former triangle into the point having the *same* coordinates referred to the latter. Here $IXY$ and $I'X'Y'$ may be *any* two triangles [Veblen and Young **2**, p. 72], and we naturally speak of "the affinity $IXY \to I'X'Y'$." In particular, if $ABC$ and $ABC'$ are two triangles with a common side, $ABC \to ABC'$ is called a *shear* or a *strain* according as the line $CC'$ is or is not parallel to $AB$. One kind of strain is sufficiently important to deserve a special name and a special symbol: the *affine reflection* $A(CC')$ or $B(CC')$, which arises when the midpoint of $CC'$ lies on $AB$. In other words, any triangle $ACC'$ determines an affine reflection $A(CC')$ whose *mirror* (or "axis") is the median through $A$ and whose *direction* is the direction of all lines parallel to $CC'$.

In the language of the Erlangen program (see page 67), the principal group for affine geometry is the group of all affinities.

#### EXERCISES

1. The shear or strain $ABC \to ABC'$ leaves invariant every point on the line $AB$ What is its effect on a point $P$ of general position?

2. Every affinity of period 2 is either a half-turn or an affine reflection.

3. If, for a given affinity, every noninvariant point lies on at least one invariant line, then the affinity is either a dilatation or a shear or a strain.

4. In terms of affine coordinates, affinities are "affine transformations"

**13.33**
$$x' = ax + by + l, \quad y' = cx + dy + m, \qquad ad \neq bc.$$

5. Describe the transformations

(i) $x' = x + 1,\ y' = y$; (ii) $x' = ax, y' = ay$;
(iii) $x' = x + by, y' = y$; (iv) $x' = ax, y' = y$.

## 13.4 EQUIAFFINITIES

*For he, by Geometrick scale,*
*Could take the size of Pots of Ale.*

Samuel Butler (1600-1680)
(*Hudibras*, I.1)

We are now ready to show how the comparison of lengths on parallel lines can be extended to yield a comparison of areas in any position [cf. Forder **1**, pp. 259–265; Coxeter **2**, pp. 125–128]. For simplicity, we restrict consideration to *polygonal* regions. (Other shapes may be included by a suitable limiting process of the kind used in integral calculus.) Clearly, any polygonal region can be dissected into a finite number of triangles.* Following H. Hadwiger and P. Glur [*Elemente der Mathematik,* **6** (1951), pp. 97–120], we declare two such regions to be *equivalent* if they can be dissected into a finite number of pieces that are congruent in pairs (by translations or by half-turns). In other words, two polygonal regions are equivalent if they can be derived from each other by dissection and rearrangement. Superposing two different dissections, we see that this kind of equivalence, which is obviously reflexive and symmetric, is also transitive; two polygons that are equivalent to the same polygon are equivalent to each other.

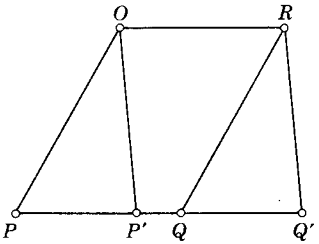


**Figure 13.4a**

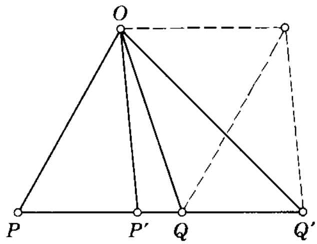


**Figure 13.4b**

The parallelograms $OPQR$ and $OP'Q'R$ of Figure 13.4*a* are equivalent, since each of them consists of the trapezoid $OP'QR$ plus one of the two congruent triangles $OPP'$, $RQQ'$. In some such cases, more than two pieces may be needed, but we find eventually:

**13.41** *Two parallelograms are equivalent if they have one pair of opposite sides of the same length lying on the same pair of parallel lines.*

Since a parallelogram can be dissected along a diagonal to make two triangles that are congruent by a half-turn, it follows that two triangles (such as $OPQ$ and $OP'Q'$ in Figure 13.4*b*) are equivalent if they have a common vertex while their sides opposite to this vertex are congruent segments on one line. In particular, if points $P_0, P_1, \ldots, P_n$ are evenly spaced along a line (not through $O$), so that the segments $P_0P_1$, $P_1P_2$, . . . are all congruent, as in Figure 13.4*c*, then the triangles $OP_0P_1$, $OP_1P_2$, . . . are all equivalent, and we naturally say that the *area* of $OP_0P_n$ is $n$ times the area of $OP_0P_1$. By interpolation of further points on the same line, we can extend this idea to all real values of $n$, with the conclusion that, if $Q$ is on the side $PQ'$ of a triangle $OPQ'$, as in Figure 13.4*d*, the *Cevian* $OQ$ divides the area of the triangle in the same ratio that the point $Q$ divides the side:

**13.42**
$$\frac{OPQ}{OPQ'} = \frac{PQ}{PQ'}.$$

* N. J. Lennes, *American Journal of Mathematics,* **33** (1911), p. 46.

We naturally regard this ratio as being negative if $P$ lies between $Q$ and $Q'$, that is, if the two triangles are oppositely oriented.

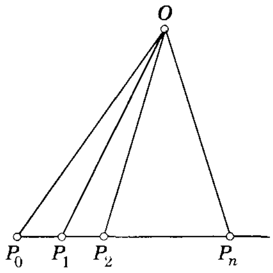



**Figure 13.4c**

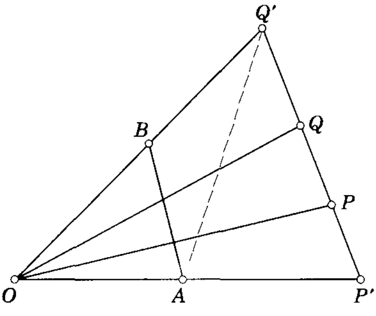



**Figure 13.4d**

These ideas enable us to define the area of any polygon in such a way that *equivalent polygons have the same area*, and when two polygons are stuck together to make a larger polygon, the areas are added. To compute the area of a given polygon in terms of a standard triangle $OAB$ as unit of measurement, we dissect the polygon into triangles and add the areas of the pieces, each computed as follows.

By applying a suitable translation, any given triangle can be shifted so that one vertex coincides with the vertex $O$ of the standard triangle $OAB$. Accordingly, we consider a triangle $OPQ$. Let the line $PQ$ meet $OA$ in $P'$, and $OB$ in $Q'$, as in Figure 13.4*d*. Multiplying together the three ratios

$$\frac{OPQ}{OP'Q'} = \frac{PQ}{P'Q'}, \quad \frac{OP'Q'}{OAQ'} = \frac{OP'}{OA}, \quad \frac{OAQ'}{OAB} = \frac{OQ'}{OB}$$

we obtain the desired ratio

**13.43**
$$\frac{OPQ}{OAB} = \frac{PQ}{P'Q'}\frac{OP'}{OA}\frac{OQ'}{OB}.$$

To obtain an analytic expression for the area of a triangle $OPQ$, referred to axes through the vertex $O$, we take the coordinates of the points

$$O, \quad A, \quad B, \quad P, \quad Q, \quad P', \quad Q'$$

to be

$$(0, 0), (1, 0), (0, 1), (x_1, y_1), (x_2, y_2), (p, 0), (0, q),$$

respectively. Since the equation

$$\frac{x}{p} + \frac{y}{q} = 1$$

for the line $PQ$ is satisfied by $(x_1, y_1)$ and $(x_2, y_2)$, we have

$$\frac{1 - y_1/q}{x_1} = \frac{1}{p} = \frac{1 - y_2/q}{x_2},$$

whence

$$q = \frac{x_1y_2 - x_2y_1}{x_1 - x_2}.$$

Taking the product of

$$\frac{PQ}{P'Q'} = \frac{x_1 - x_2}{p}, \quad \frac{OP'}{OA} = p, \quad \frac{OQ'}{OB} = \frac{x_1y_2 - x_2y_1}{x_1 - x_2},$$

we obtain

**13.44** $$\frac{OPQ}{OAB} = x_1y_2 - x_2y_1 = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}.$$

We deduce, as in § 8.2, that a triangle

$$(x_1, y_1)\ (x_2, y_2)\ (x_3, y_3),$$

of general position, has area $PQR$, where

**13.45** $$\frac{PQR}{OAB} = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}.$$

Since the homogeneous linear transformation

$$x' = ax + by, \qquad y' = cx + dy$$

takes the triangle $OAB$ to

$$(0, 0)(a, c)(b, d),$$

we conclude that the affinity 13.33 *preserves area* if and only if

$$ad - bc = 1.$$

An area-preserving affinity is called an *equiaffinity* (or "equiaffine collineation" [Veblen and Young **2**, pp. 105–113]). The group of all equiaffinities, like the group of all dilatations, includes the group of all translations and half-turns as a normal subgroup, and is itself a normal subgroup in the group of all affinities. Equiaffinities are of many kinds. Here are some examples:

The *hyperbolic rotation* ("Lorentz transformation" or "Procrustean stretch")

**13.46** $$x' = \mu^{-1}x, \quad y' = \mu y \qquad (\mu > 0, \quad \mu \neq 1),$$

for which $x'y' = xy$, leaves invariant each branch of the hyperbola $xy = 1$. Th c ossed hvperbolic rotation

**13.47** $$x' = -\mu^{-1}x, \quad y' = -\mu y \qquad (\mu > 0, \quad \mu \neq 1)$$

interchanges the two branches. The *parabolic rotation*

**13.48** $$x' = x + 1, \quad y' = 2x + y + 1,$$

for which $x'^2 - y' = x^2 - y$, leaves invariant the parabola $y = x^2$. The *elliptic rotation*

**13.49** $$x' = x\cos\theta - y\sin\theta, \; y' = x\sin\theta + y\cos\theta$$

leaves invariant the ellipse $x^2 + y^2 = 1$, and is periodic if $\theta$ is commensurable with $\pi$.

In §2.8 (page 36) we derived a regular polygon $P_0P_1P_2\ldots$ from a point $P_0$ (other than the center) by repeated application of a rotation through $2\pi/n$. (The rotation takes $P_0$ to $P_1$, $P_1$ to $P_2$, and so on.) Although measurement of angles has no meaning in affine geometry, we can define an *affinely regular polygon* whose vertices $P_j$ are derived from suitable point $P_0$ by repeated application of an equiaffinity. The polygon is said to be of type $\{n\}$ if the equiaffinity is an elliptic rotation 13.49, where $\theta = 2\pi/n$ and $n$ is rational, so that $P_j$ has affine coordinates

$$(\cos j\theta, \sin j\theta) \qquad (\theta = 2\pi/n).$$

Figure 13.4*e* shows an affinely regular pentagram ($n = 5/2$) and pentagon ($n = 5$).

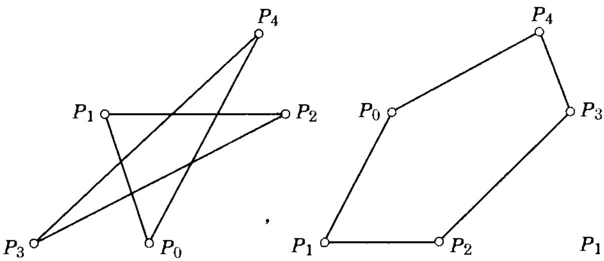



**Figure 13.4e**

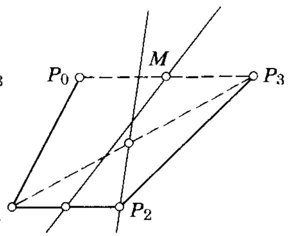



**Figure 13.4f**

## EXERCISES

**1.** Two triangles with a common side (such as $ABC$ and $BCD$ in Figure 13.2*d*) have the same area if and only if the line joining their remaining vertices is parallel to the common side (that is, $AD$ parallel to $BC$).

**2.** If a pentagon has four of its diagonals parallel to four of its sides, the remaining diagonal is parallel to the remaining side.

**3.** When is a dilatation an equiaffinity?

**4.** When is a shear an equiaffinity?

**5.** When is a strain an equiaffinity?

**6.** The product of any even number of affine reflections is an equiaffinity.

**7.** Any translation or half-turn or shear can be expressed as the product of two affine reflections.

**8.** If an equiaffinity is neither a translation nor a half-turn nor a shear, it can be expressed as $P_0P_1P_2 \to P_1P_2P_3$ where $P_0P_3$ is parallel to $P_1P_2$. (See Figure 13.4*f*.)

**9.** Every equiaffinity can be expressed as the product of two affine reflections. (Veblen.)

**10.** In an affinely regular polygon $P_0P_1P_2 \ldots$, the lines $P_iP_j$ and $P_hP_k$ are parallel whenever $i + j = h + k$.

**11.** Why did we call $x^2 + y^2 = 1$ an ellipse rather than a circle (just below 13.49)?

**12.** What triangles and quadrangles are affinely regular?

**13.** Construct an affinely regular hexagon.

**14.** Compute the ratio $P_0P_3/P_1P_2$ for an affinely regular polygon of type $\{n\}$.

**15.** For which values of $n$ can an affinely regular polygon of type $\{n\}$ be constructed with a parallel-ruler?

## 13.5 TWO-DIMENSIONAL LATTICES

*Farey has a notice of twenty lines in the Dictionary of National Biography. . . . His biographer does not mention the one thing in his life which survives.*

G. H. Hardy

[Hardy and Wright **1**, p. 37]

Our treatment of lattices in § 4.1 (as far as the description of Figure 4.1*d*) is purely affine. In fact, a lattice is the set of points whose affine coordinates are integers. Any one of the points will serve as the origin $O$.

Let $A'$ be any lattice point, and $A$ the first lattice point along the ray $OA'$. Following Hardy and Wright [**1**, p. 29], we call $A$ a *visible* point, because there is no lattice point between $O$ and $A$ to hide $A$ from an observer at $O$. In terms of affine coordinates, a necessary and sufficient condition for $(x, y)$ to be visible is that the integers $x$ and $y$ be coprime, that is, that they have no common divisor greater than 1. The three visible points

$$(1, 0), \quad (1, 1), \quad (0, 1)$$

form with the origin a parallelogram. This is called a *unit cell* (or "typical parallelogram") of the lattice, because the translations transform it into infinitely many such cells filling the plane without overlapping and without interstices: it is a fundamental region for the group of translations. Thus it serves as a convenient unit for computing the area of a region.

According to Steinhaus [**2**, pp. 76–77, 260] it was G. Pick, in 1899, who discovered the following theorem:*

* For an extension to three dimensions, see J. E. Reeve, On the volume of lattice polyhedra, *Proceedings of the London Mathematical Society* (3), **7** (1957), pp. 378–395.

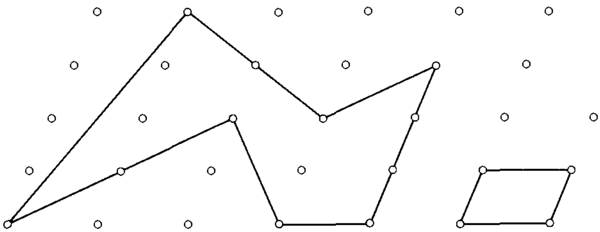


**Figure 13.5a**

**13.51** *The area of any simple polygon whose vertices are lattice points is given by the formula*

$$\tfrac{1}{2}b + c - 1,$$

*where b is the number of lattice points on the boundary while c is the number of lattice points inside.*

(By a "simple" polygon we mean one whose sides do not cross one another. Figure 13.5*a* shows an example in which $b = 11$, $c = 3$.)

*Proof.* We first observe that the expression $\tfrac{1}{2}b + c - 1$ is additive when two polygons are juxtaposed. In fact, if two polygons, involving $b_1 + c_1$ and $b_2 + c_2$ lattice points respectively, have a common side containing $n$ ($\geqslant 0$) lattice points in addition to the two vertices at its ends, then the values of $b$ and $c$ for the combined polygon are

$$b = b_1 + b_2 - 2n - 2, \qquad c = c_1 + c_2 + n,$$

so that

$$\tfrac{1}{2}b + c - 1 = (\tfrac{1}{2}b_1 + c_1 - 1) + (\tfrac{1}{2}b_2 + c_2 - 1).$$

Next, the formula holds for a parallelogram having no lattice points on its sides (so that $b = 4$ and the expression reduces to $c + 1$). For, when $N$ such parallelograms are fitted together, four at each vertex, to fill a large region, the number of lattice points involved (apart from a negligible peripheral error) is $N(c + 1)$, and this must be the same as the number of unit cells needed to fill the same region.

Splitting the parallelogram into two congruent triangles by means of a diagonal, we see that the formula holds also for a triangle having no lattice points on its sides. A triangle that does have lattice points on a side can be dealt with by joining such points to the opposite vertex so as to split the triangle into smaller triangles. This procedure may have to be repeated, but obviously only a finite number of times. Finally, as we remarked on page 204, any given polygon can be dissected into triangles; then the expressions for those pieces can be added to give the desired result.

In particular, any parallelogram for which $b = 4$ and $c = 0$ has area 1 and can serve as a unit cell. If the vertices of such a parallelogram (in counterclockwise order) are

$$(0, 0), \quad (x, y), \quad (x + x_1, y + y_1), \quad (x_1, y_1),$$

we see from 13.44 that

**13.52** $$xy_1 - yx_1 = 1.$$

In other words, this is the condition for the points

**13.53** $$(0, 0), \quad (x, y), \quad (x_1, y_1)$$

to form a positively oriented "empty" triangle of area $\frac{1}{2}$, which could be used just as well as (0, 0) (1, 0) (0, 1) to generate the lattice. Thus a lattice is completely determined, apart from its position, by the area of its unit cell. Moreover, although there are infinitely many visible points in a given lattice, they all play the same role. (These properties of affine geometry are in marked contrast to Euclidean geometry, where the shape of a lattice admits unlimited variation and each lattice contains visible points at infinitely many different distances.)

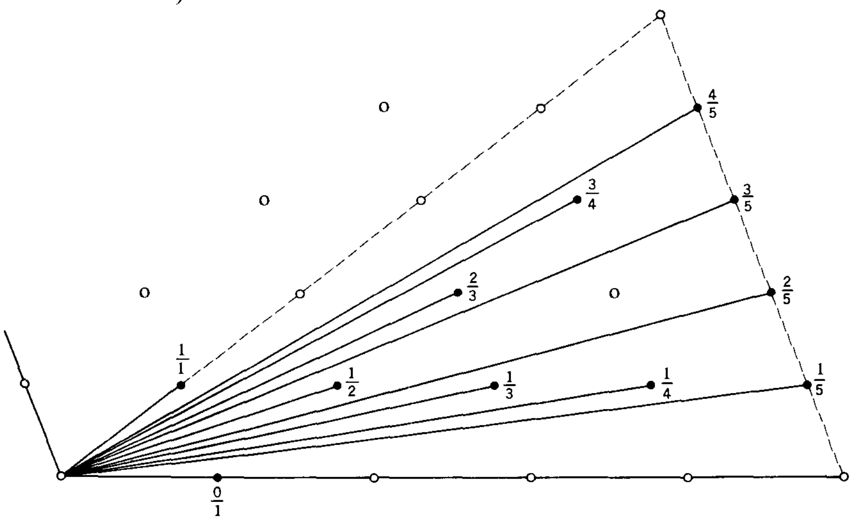



**Figure 13.5b**

George Pólya* has applied 13.52 to a useful lemma in the theory of numbers. The *Farey series* $F_n$ of order $n$ is the ascending sequence of fractions from 0 to 1 whose denominators do not exceed $n$. Thus $y/x$ belongs to $F_n$ if $x$ and $y$ are coprime and

**13.54** $$0 \leqslant y \leqslant x \leqslant n.$$

For instance, $F_5$ is

$$\tfrac{0}{1}, \tfrac{1}{5}, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{2}{5}, \tfrac{1}{2}, \tfrac{3}{5}, \tfrac{2}{3}, \tfrac{3}{4}, \tfrac{4}{5}, \tfrac{1}{1}.$$

* *Acta Litterarum ac Scientiarum Regia Universitatis Hungaricae Francisco-Josephinae*, Sectio

The essential property of such a sequence, from which many other properties follow by simple algebra, is that 13.52 holds for any two adjacent fractions

$$\frac{y}{x} \text{ and } \frac{y_1}{x_1}.$$

To prove this, we represent each term $y/x$ of the sequence by the point $(x, y)$ of a lattice. For example, the terms of $F_5$ are the lattice points emphasized in Figure 13.5*b* (where, for convenience, the angle between the axes is obtuse). Since the fractions are in their "lowest terms," the points are visible. By 13.54, they belong to the triangle (0, 0) ($n$, 0) ($n$, $n$). A ray from the origin, rotated counterclockwise, passes through the representative points in their proper order. If $y/x$ and $y_1/x_1$ are consecutive terms of the sequence, then $(x, y)$ and $(x_1, y_1)$ are visible points such that the triangle joining them to the origin contains no lattice point in its interior. Hence this triangle is one half of a unit cell, and 13.52 holds, as required.

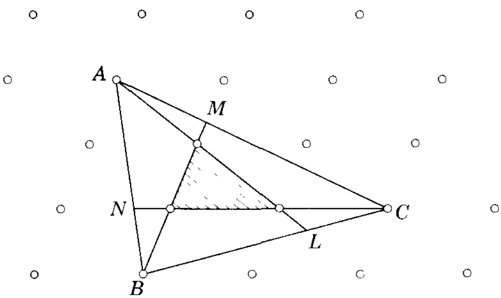



**Figure 13.5c**

Another result belonging to affine geometry is

**13.55** *If the sides BC, CA, AB of a triangle ABC are divided at L, M, N in the respective ratios* $\lambda : 1$, $\mu : 1$, $\nu : 1$, *the Cevians AL, BM, CN form a triangle whose area is*

$$\frac{(\lambda\mu\nu - 1)^2}{(\lambda\mu + \lambda + 1)(\mu\nu + \mu + 1)(\nu\lambda + \nu + 1)}$$

*times that of ABC.*

This was discovered by Routh [**1**, p. 82; see also Dörrie **1**, pp. 41–42]. We shall give a general proof in § 13.7, but it is interesting to observe that, when $\lambda = \mu = \nu$, so that the ratio of areas is $(\lambda - 1)^3/(\lambda^3 - 1)$, the result can be deduced from 13.51. For instance, when $\lambda = \mu = \nu = 2$, so that each side is trisected [Steinhaus **2**, p. 8], the central triangle is one-seventh of the whole, and we can see this immediately by embedding the figure in a lattice, as in Figure 13.5*c*. Since the central triangle has $b = 3$, $c = 0$ while $ABC$ has $b = 3$, $c = 3$, the ratio of areas is $\frac{1}{2}/\frac{7}{2} = \frac{1}{7}$.

**EXERCISES**

**1.** If $y/x$ and $y_1/x_1$ are two consecutive terms of a Farey series, $x$ and $x_1$ are coprime.

**2.** If $y_0/x_0$, $y/x$, $y_1/x_1$ are three consecutive terms of a Farey series,

$$\frac{y_0 + y_1}{x_0 + x_1} = \frac{y}{x}.$$

(C. Haros, 1802.)

**3.** The points $A$, $B$, $C$ in Figure 13.5*c* belong to a lattice whose unit cell has seven times the area of that of the basic lattice. (For the Euclidean theory of such compound lattices, see Coxeter, Configurations and maps, *Reports of a Mathematical Colloquium* (2), **8** (1948), pp. 18–38, especially Figs. i, v, vii.)

**4.** Use lattices to verify 13.55 when (*a*) $\lambda = \mu = \nu = 3$, (*b*) $\lambda = \mu = \nu = \frac{1}{2}$.

**5.** Join the vertices $A$, $B$, $C$, $D$ of a parallelogram to the midpoints of the respective sides $BC$, $CD$, $DA$, $AB$ so as to form a smaller parallelogram in the middle. Its area is one-fifth that of $ABCD$. Another such parallelogram is obtained by joining $A$, $B$, $C$, $D$ to the midpoints of $CD$, $DA$, $AB$, $BC$. The common part of these two small parallelograms is a centrally symmetrical octagon whose area is one-sixth that of $ABCD$ [Dörrie **1**, p. 40].

**6.** In the notation of 13.55, the area of the triangle $LMN$ is

$$\frac{\lambda\mu\nu + 1}{(\lambda + 1)(\mu + 1)(\nu + 1)}$$

times that of $ABC$. (*Hint*: Use 13.42 to compute the relative area of $CLM$, etc.)

**7.** Of the four triangles $ANM$, $BLN$, $CML$, $LMN$, the last cannot have the smallest area unless $L$, $M$, $N$ are the midpoints of $BC$, $CA$, $AB$. (H. Debrunner.*)

## 13.6 VECTORS AND CENTROIDS

*A vector is really the same thing as a translation, although one uses different phraseologies for vectors and translations. Instead of speaking of the translation $A \to A'$ which carries the point $A$ into $A'$ one speaks of the vector $\overrightarrow{AA'}$. . . . The same vector laid off from $B$ ends in $B'$ if the translation carrying $A$ into $A'$ carries $B$ into $B'$.*

H. Weyl [**1**, p. 45]

As we saw in § 2.5, a *group* is an associative system containing an identity and, for each element, an inverse. Arithmetical instances are provided by the positive rational numbers, the positive real numbers, the complex numbers of modulus 1, and all the complex numbers except 0, combined, in each case, by ordinary multiplication. Such instances make it natural to adopt a multiplicative notation for all groups, so that the combination of S and T is ST, the inverse of S is $S^{-1}$, and the identity is 1. However, it is often convenient, especially in the case of Abelian (i.e., commutative)

* *Elemente der Mathematik*, **12** (1957), p. 43, Aufgabe 260.

groups, to use instead the additive notation, in which the combination of S and T is S + T, the inverse of S is −S, and the identity is 0. To see that this other notation has equally simple arithmetical instances, we merely have to consider in turn the integers, the rational numbers, the real numbers, and the complex numbers, combined, in each case, by ordinary addition.

The transition from a multiplicative group to the corresponding additive group is the foundation of the theory of logarithms [Infeld **1**, pp. 97–100].

When we go outside the domain of arithmetic, the choice between multiplication and addition is merely a matter of notation. In particular, the Abelian group of translations, which we have expressed as a multiplicative group, becomes the additive group of *vectors*.

In this notation, 13.21 asserts that any two points $A$ and $A'$ determine a unique vector $\overrightarrow{AA'}$ (going from $A$ to $A'$), Figure 13.2*b* illustrates a situation in which

$$\overrightarrow{AA'} = \overrightarrow{CC'} = \overrightarrow{BB'},$$

13.23 asserts that

$$\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC},$$

and 3.23 asserts that, for any two vectors $\mathbf{a}$ and $\mathbf{b}$,

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}.$$

In the same spirit the "origin" will henceforth be called $O$ instead of $I$, and the zero vector will be denoted by $\mathbf{0}$. The integral multiples of any non-zero vector proceed from the origin to the points of a one-dimensional lattice. Two vectors $\mathbf{e}$ and $\mathbf{f}$ are said to be *independent* if neither is a (real) multiple of the other, that is, if the only numbers that satisfy the vector equation

$$x\mathbf{e} + y\mathbf{f} = \mathbf{0}$$

are $x = 0$ and $y = 0$. Two such vectors (corresponding to the translations $X$ and $Y$ in Figure 4.1*c*) provide a basis for a system of affine coordinates: they enable us to define the coordinates of any point to be the coefficients in the expression

$$x\mathbf{e} + y\mathbf{f}$$

for the *position vector* which goes from the origin to the given point. In other words, with reference to a triangle $OAB$, the affine coordinates of a point $P$ are the coefficients in the expression

$$\overrightarrow{OP} = x\,\overrightarrow{OA} + y\,\overrightarrow{OB}.$$

We shall find it useful to borrow from statics the notion of the centroid (or "center of gravity") of a set of "weighted" points, that is, of points to each of which a real number is attached in a special way. For convenience, we shall call these numbers masses, although, when some of them are negative, electric charges provide a more appropriate illustration.

Let masses $t_1, \ldots, t_k$ be assigned to $k$ distinct points $A_1, \ldots, A_k$, let $O$ be any point (possibly coincident with one of the $A$'s), and consider the vector

$$t_1 \overrightarrow{OA_1} + \ldots + t_k \overrightarrow{OA_k}.$$

If $t_1 + \ldots + t_k = 0$, this vector is independent of the choice of $O$. For, if we subtract from it the result of using $O'$ instead, we obtain

$$t_1 (\overrightarrow{OA_1} - \overrightarrow{O'A_1}) + \ldots + t_k (\overrightarrow{OA_k} - \overrightarrow{O'A_k})$$

$$= (t_1 + \ldots + t_k) \overrightarrow{OO'} = \mathbf{0}.$$

More interestingly, if

$$t_1 + \ldots + t_k \neq 0,$$

we have

$$t_1 \overrightarrow{OA_1} + \ldots + t_k \overrightarrow{OA_k} = (t_1 + \ldots + t_k) \overrightarrow{OP},$$

where the point $P$ is independent of the choice of $O$. For, if the same procedure with $O'$ instead of $O$ yields $P'$ instead of $P$, we have, by subtraction,

$$(t_1 + \ldots + t_k) \overrightarrow{OO'} = (t_1 + \ldots + t_k)(\overrightarrow{OP} - \overrightarrow{O'P'})$$

whence $$\overrightarrow{OP'} = \overrightarrow{OO'} + \overrightarrow{O'P'} = \overrightarrow{OP},$$

so that $P'$ coincides with $P$. This point $P$, given by

$$\Sigma t_i \overrightarrow{OP} = \Sigma t_i \overrightarrow{OA_i},$$

is called the *centroid* (or "barycenter") of the $k$ masses $t_i$ at $A_i$.

Since, having found $P$, we may choose this position for $O$, we have

$$\Sigma t_i \overrightarrow{PA_i} = \mathbf{0}.$$

If there are only two points,

$$t_1 \overrightarrow{PA_1} = -t_2 \overrightarrow{PA_2},$$

so that $P$ lies on the line $A_1A_2$ and divides the segment $A_1A_2$ in the ratio $t_2 : t_1$. In particular, if $t_1 = t_2$, $P$ is the midpoint of $A_1A_2$.

For a triangle $A_1A_2A_3$, we have

$$(t_1 + t_2 + t_3)\,\overrightarrow{OP} = t_1\overrightarrow{OA_1} + t_2\overrightarrow{OA_2} + t_3\overrightarrow{OA_3}$$
$$= t_1\overrightarrow{OA_1} + (t_2 + t_3)\,\overrightarrow{OQ},$$

where $Q$ is the centroid of $t_2$ at $A_2$ and $t_3$ at $A_3$. Thus, in seeking the centroid of three masses, we may replace two of them by their combined mass at their own centroid. (There is an obvious generalization to more than three masses.) In particular, when $t_1 = t_2 = t_3$ (= 1, say), $Q$ is the midpoint of $A_2A_3$, and $P$ divides $A_1Q$ in the ratio 2 : 1. Thus the "centroid" $G$ of a triangle (§ 1.4) is the centroid of equal masses at its three vertices.

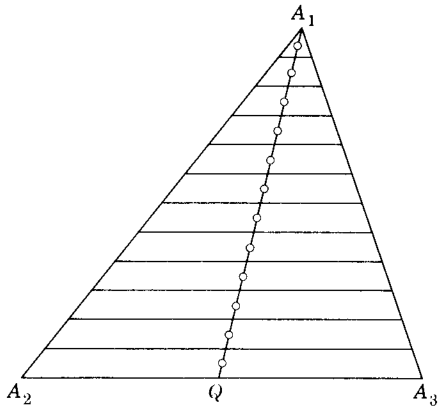



**Figure 13.6a**

This same point $G$, where the medians concur, is also the centroid of a triangular *lamina* or "plate" of uniform density. (Strictly speaking, this notion requires integral calculus.) For we may divide the triangle into thin strips parallel to the side $A_2A_3$, as in Figure 13.6*a*. The centroids of these strips evidently lie on the median $A_1Q$. Hence the centroid of the whole lamina lies on this median, and similarly on the others. (This argument was used by Archimedes in the third century B.C.)

## EXERCISES

**1.** Verify in detail that
(i) the positive rational numbers,
(ii) the positive real numbers,
(iii) the complex numbers of modulus 1,
(iv) all the complex numbers except 0
form multiplicative groups; and that
(v) the integers,
(vi) the rational numbers,
(vii) the real numbers,
(viii) the complex numbers

form additive groups. Explain why the first four sets do not form additive groups, and why the last four do not form multiplicative groups.

**2.** If $A$, $B$, $C$ are on one line and $A'$, $B'$, $C'$ on another with

$$\frac{AB}{A'B'} = \frac{BC}{B'C'},$$

then points dividing all the segments $AA'$, $BB'$, $CC'$ in the same ratio are either collinear or coincident (cf. § 3.6). (*Hint:* Consider the centroid of suitable masses at $A$, $C$, $A'$, $C'$.)

**3.** The centroid of equal masses at the vertices of a quadrangle is the center of the Varignon parallelogram (Figure 13.2*g*).

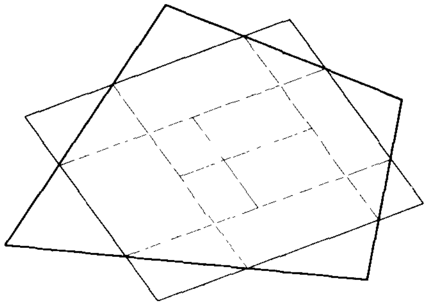


**Figure 13.6b**

**4.** The centroid of a quadrangular lamina is the center of the Wittenbauer parallelogram, whose sides join adjacent points of trisection of the sides, as in Figure 13.6*b*. This theorem, due to F. Wittenbauer (1857–1922) [Blaschke **2,** p. 13], was rediscovered by J. J. Welch and V. W. Foss.*

**5.** For what kind of quadrangle will the centroids described in the two preceding exercises coincide?

## 13.7 BARYCENTRIC COORDINATES

If $t_1 + t_2 \neq 0$, masses $t_1$ and $t_2$ at two fixed points $A_1$ and $A_2$ determine a unique centroid $P$, as in Figure 13.7*a*. This point is $A_1$ itself if $t_2 = 0$, $A_2$ if $t_1 = 0$. It is on the segment $A_1A_2$ if the $t$'s are both positive (or both negative), on the ray $A_1/A_2$ if

$$t_1 > -t_2 > 0,$$

and on the ray $A_2/A_1$ if $\qquad t_2 > -t_1 > 0.$

* *Mathematical Gazette*, **42** (1958), p. 55; **43** (1959), p. 46.

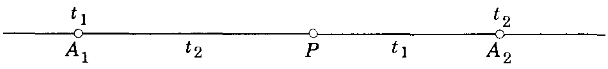



**Figure 13.7a**

Conversely, given a point $P$ on the line $A_1A_2$, we can find numbers $t_1$ and $t_2$ such that

$$\frac{t_2}{t_1} = \frac{A_1P}{PA_2} \quad \text{or} \quad \frac{t_1}{t_2} = \frac{PA_2}{A_1P};$$

then $P$ will be the centroid of masses $t_1$ and $t_2$ at $A_1$ and $A_2$. Since masses $\mu t_1$ and $\mu t_2$ (where $\mu \neq 0$) determine the same point as $t_1$ and $t_2$, these *barycentric coordinates* are homogeneous:

$$(t_1, t_2) = (\mu t_1, \mu t_2) \qquad (\mu \neq 0).$$

Similarly, as Möbius observed in 1827, we may set up barycentric coordinates in the plane of a *triangle of reference* $A_1A_2A_3$. If $t_1 + t_2 + t_3 \neq 0$, masses $t_1, t_2, t_3$ at the three vertices determine a point $P$ (the centroid) whose coordinates are $(t_1, t_2, t_3)$. In particular, $(1, 0, 0)$ is $A_1$, $(0, 1, 0)$ is $A_2$, $(0, 0, 1)$ is $A_3$, and $(0, t_2, t_3)$ is the point on $A_2A_3$ whose one-dimensional coordinates with respect to $A_2$ and $A_3$ are $(t_2, t_3)$. To find coordinates for a given point $P$ of general position, we find $t_2$ and $t_3$ from such a point $Q$ on the line $A_1P$, as in Figure 13.7*b*, and then determine $t_1$ as the mass at $A_1$ that will balance a mass $t_2 + t_3$ at $Q$ so as to make $P$ the centroid. Again, as in the one-dimensional case, these coordinates are homogeneous:

$$(t_1, t_2, t_3) = (\mu t_1, \mu t_2, \mu t_3) \qquad (\mu \neq 0).$$

Joining $P$ to $A_1, A_2, A_3$, we decompose $A_1A_2A_3$ into three triangles having a common vertex $P$. *The areas of these triangles are proportional to the barycentric coordinates of P*, as in Figure 13.7*c*. This fact follows at once from 13.42, since

$$\frac{t_3}{t_2} = \frac{A_2Q}{QA_3} = \frac{A_1A_2Q}{A_1QA_3} = \frac{PA_2Q}{PQA_3} = \frac{A_1A_2Q - PA_2Q}{A_1QA_3 - PQA_3} = \frac{PA_1A_2}{PA_3A_1},$$

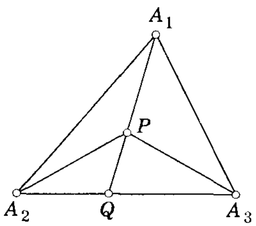



**Figure 13.7b**

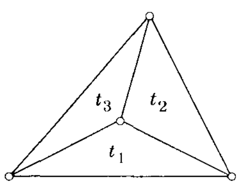



**Figure 13.7c**

and similarly for $t_1/t_3$, $t_2/t_1$. Positions of $P$ outside the triangle are covered by means of our convention for the sign of the area of a directed triangle.

The inequality

$$t_1 + t_2 + t_3 \neq 0$$

enables us to normalize the coordinates so that

**13.71** $$t_1 + t_2 + t_3 = 1.$$

(We merely have to divide each coordinate by the sum of all three.) These normalized barycentric coordinates are called *areal* coordinates, because they are just the areas of the triangles $PA_2A_3$, $PA_3A_1$, $PA_1A_2$, expressed in terms of the area of the whole triangle $A_1A_2A_3$ as unit of measurement. Areal coordinates are not homogeneous but "redundant": the position of a point is determined by two of the three, and the third is retained for the sake of symmetry. However, any expression involving them can be made homogeneous by inserting suitable powers of $t_1 + t_2 + t_3$ in appropriate places.

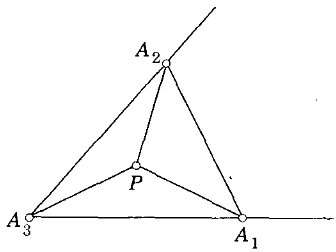



**Figure 13.7d**

In affine coordinates, as we have seen, a line has a linear equation. In barycentric coordinates, as we shall soon see, *a line has a linear homogeneous equation.* For this purpose we use the segments $A_3A_1$ and $A_3A_2$ as axes for affine coordinates, as in Figure 13.7*d*, so that the coordinates of $P$, $A_1$, $A_2$, $A_3$, which were formerly

$$(t_1, t_2, t_3),\quad (1, 0, 0),\quad (0, 1, 0),\quad (0, 0, 1),$$

are now

$$(x, y), \qquad (1, 0), \qquad (0, 1), \qquad (0, 0).$$

By 13.44, the areas of $PA_2A_3$ and $PA_3A_1$, as fractions of the "unit" triangle $A_1A_2A_3$, are just

$$\begin{vmatrix} x & y \\ 0 & 1 \end{vmatrix} = x \quad \text{and} \quad \begin{vmatrix} 1 & 0 \\ x & y \end{vmatrix} = y.$$

By subtraction, the area of $PA_1A_2$ is $1 - x - y$. Hence the *areal* coordinates of $P$ are related to the affine coordinates by the very simple formulas

$$t_1 = x, \quad t_2 = y, \quad t_3 = 1 - x - y.$$

The general line, having the affine equation 8.11, has the areal equation

$$at_1 + bt_2 + c = 0.$$

Making this homogeneous by the insertion of $t_1 + t_2 + t_3$, we deduce the barycentric equation

$$at_1 + bt_2 + c(t_1 + t_2 + t_3) = 0$$

or

$$(a + c)t_1 + (b + c)t_2 + ct_3 = 0$$

or, in a more symmetrical notation,

**13.72** $$T_1t_1 + T_2t_2 + T_3t_3 = 0.$$

Thus every line has a linear homogeneous equation. In particular, the lines $A_2A_3, A_3A_1, A_1A_2$ have the equations

**13.73** $$t_1 = 0, \quad t_2 = 0, \quad t_3 = 0.$$

The line joining two given points $(r)$ and $(s)$, meaning

$$(r_1, r_2, r_3) \quad \text{and} \quad (s_1, s_2, s_3),$$

has the equation

**13.74** $$\begin{vmatrix} r_1 & r_2 & r_3 \\ s_1 & s_2 & s_3 \\ t_1 & t_2 & t_3 \end{vmatrix} = 0.$$

For, this equation is linear in the $t$'s and is satisfied when the $t$'s are replaced by the $r$'s or the $s$'s. Another way to obtain this result is to ask for the fixed points $(r)$ and $(s)$ to form with the variable point $(t)$ a "triangle" whose area is zero. In terms of areal coordinates, with the triangle of reference as unit, the area of the triangle $(r)(s)(t)$ is, by 13.45 and 13.71,

$$\begin{vmatrix} r_1 & r_2 & 1 \\ s_1 & s_2 & 1 \\ t_1 & t_2 & 1 \end{vmatrix} = \begin{vmatrix} r_1 & r_2 & r_1 + r_2 + r_3 \\ s_1 & s_2 & s_1 + s_2 + s_3 \\ t_1 & t_2 & t_1 + t_2 + t_3 \end{vmatrix} = \begin{vmatrix} r_1 & r_2 & r_3 \\ s_1 & s_2 & s_3 \\ t_1 & t_2 & t_3 \end{vmatrix}.$$

Hence the area in general barycentric coordinates is this last determinant divided by

$$(r_1 + r_2 + r_3)(s_1 + s_2 + s_3)(t_1 + t_2 + t_3).$$

We are now ready to prove Routh's theorem 13.55 in its full generality. Identifying $ABC$ with $A_1A_2A_3$, so that the points $L$, $M$, $N$ are

$$(0, 1, \lambda), \quad (\mu, 0, 1), \quad (1, \nu, 0),$$

we can express the lines $AL$, $BM$, $CN$ as

$$\lambda t_2 = t_3, \quad \mu t_3 = t_1, \quad \nu t_1 = t_2.$$

They intersect in pairs in the three points

$$(\mu, \mu\nu, 1), \quad (1, \nu, \nu\lambda), \quad (\lambda\mu, 1, \lambda),$$

forming a triangle whose area, in terms of that of the triangle of reference, is the result of dividing the determinant

$$\begin{vmatrix} \mu & \mu\nu & 1 \\ 1 & \nu & \nu\lambda \\ \lambda\mu & 1 & \lambda \end{vmatrix} = (\lambda\mu\nu - 1)^2$$

by $(\mu + \mu\nu + 1)(1 + \nu + \nu\lambda)(\lambda\mu + 1 + \lambda)$, in agreement with the statement of 13.55.

As an important special case we have

**CEVA'S THEOREM.** *Let the sides of a triangle ABC be divided at L, M, N in the respective ratios* $\lambda : 1$, $\mu : 1$, $\nu : 1$. *Then the three lines AL, BM, CN are concurrent if and only if* $\lambda\mu\nu = 1$.

The general line 13.72 meets the sides 13.73 of the triangle of reference in the points

$$(0, T_3, -T_2), \quad (-T_3, 0, T_1), \quad (T_2, -T_1, 0),$$

which divide them in the ratios

$$-\frac{T_2}{T_3}, \quad -\frac{T_3}{T_1}, \quad -\frac{T_1}{T_2},$$

whose product is $-1$. Conversely, any three numbers whose product is $-1$ can be expressed in this way for suitable values of $T_1$, $T_2$, $T_3$. Hence

**MENELAUS'S THEOREM.** *Let the sides of a triangle be divided at L, M, N in the respective ratios* $\lambda : 1$, $\mu : 1$, $\nu : 1$. *Then the three points L, M, N are collinear if and only if* $\lambda\mu\nu = -1$.

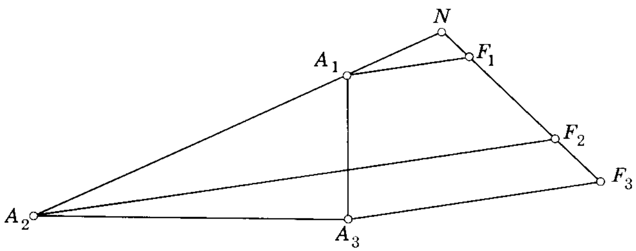



**Figure 13.7e**

The coefficients $T_1$, $T_2$, $T_3$ in the equation 13.72 for a line are sometimes called the *tangential coordinates* of the line. These homogeneous "coordinates" have a simple geometric interpretation [Salmon **1**, p. 11]: they may be regarded as *the distances from* $A_1$, $A_2$, $A_3$ *to the line,* measured in any direction (the same for all). To prove this, let $A_1F_1$, $A_2F_2$, $A_3F_3$ be these distances, as in Figure 13.7*e*. Since

$$\frac{A_1N}{NA_2} = -\frac{T_1}{T_2},$$

the homothetic triangles $NA_1F_1$ and $NA_2F_2$ yield

$$\frac{A_1F_1}{A_2F_2} = \frac{A_1N}{A_2N} = \frac{T_1}{T_2}.$$

Hence

$$\frac{A_1F_1}{T_1} = \frac{A_2F_2}{T_2},$$

and similarly each of these expressions is equal to $\frac{A_3F_3}{T_3}$.

Möbius's invention of homogeneous coordinates was one of the most far-reaching ideas in the history of mathematics: comparable to Leibniz's invention of differentials, which enabled him to express the equation

$$\frac{d}{dx} f(x) = f'(x)$$

in the homogeneous form

$$df(x) = f'(x)\,dx$$

(for instance, $d \sin x = \cos x\, dx$).

## EXERCISES

**1.** Sketch the seven regions into which the lines $A_2A_3$, $A_3A_1$, $A_1A_2$ decompose the plane, marking each according to the signs of the three areal coordinates.

**2.** Verify that 13.45 yields $1 - x - y$ as the area of the triangle $PA_1A_2$ in Figure 13.7*d*.

**3.** In areal coordinates, the midpoint of $(s_1, s_2, s_3)(t_1, t_2, t_3)$ is

$$\left(\frac{s_1 + t_1}{2}, \frac{s_2 + t_2}{2}, \frac{s_3 + t_3}{2}\right).$$

**4.** The centroid of masses $\sigma$ and $\tau$ at points whose areal coordinates are $(s_1, s_2, s_3)$ and $(t_1, t_2, t_3)$ is the point whose barycentric coordinates are

$$(\sigma s_1 + \tau t_1,\ \sigma s_2 + \tau t_2,\ \sigma s_3 + \tau t_3).$$

**5.** In barycentric coordinates, any point on the line $(s)(t)$ may be expressed in the form

$$(\sigma s_1 + \tau t_1,\ \sigma s_2 + \tau t_2,\ \sigma s_3 + \tau t_3).$$

**6.** Apply barycentric coordinates to Ex. 6 at the end of § 13.5. What becomes of this result when $L$, $M$, $N$ are collinear?

**7.** In what way do the signs of $T_1$, $T_2$, $T_3$ depend on the position of the line 13.72 in relation to the triangle of reference? When $T_2$ and $T_3$ are positive, describe the cases $T_2 < T_3$, $T_2 = T_3$, $T_2 > T_3$.

## 13.8 AFFINE SPACE

*Give me something to construct and I shall become God for the time being, pushing aside all obstacles, winning all the hard knowledge I need for the construction . . . advancing Godlike to my goal!*

J. L. Synge [**2,** p. 162]

Affine geometry can be extended from two dimensions to three by using Axioms 12.42 and 12.43 instead of 12.41. The total number of axioms is not really increased, as 13.12 now becomes a provable theorem [Forder **1,** pp. 155–157]. A line and a plane, or two planes, are said to be *parallel* if they have no common point (or if the line lies in the plane, or if the two planes coincide). Thus any plane that meets two parallel planes meets them in parallel lines; if two planes are parallel, any line in either plane is parallel to the other plane; if two lines are parallel, any plane through either line is parallel to the other line.

The existence of parallel planes is ensured by the following theorem (cf. Axiom 13.11):

**13.81** *For any point $A$ and any plane $\gamma$, not through $A$, there is just one plane through $A$ parallel to $\gamma$.*

*Proof.* Let $q$ and $r$ be two intersecting lines in $\gamma$. Let $q'$ and $r'$ be the respectively parallel lines through $A$. Then the plane $q'r'$ is parallel to $\gamma$. For otherwise, by 12.431, the two planes would meet in a line $l$. Since $q'$ and $r'$ are parallel to $\gamma$, they cannot meet $l$. Thus $q'$ and $r'$ are two parallels to $l$ through $A$, contradicting 13.11. This proves that $q'r'$ is parallel to $\gamma$. Moreover, $q'r'$ is the only plane through $A$ parallel to $\gamma$. For, two such would meet in a line $s'$ through $A$, and we could obtain a contradiction by considering their section by the plane $As$, where $s$ is a line in $\gamma$ not parallel to $s'$.

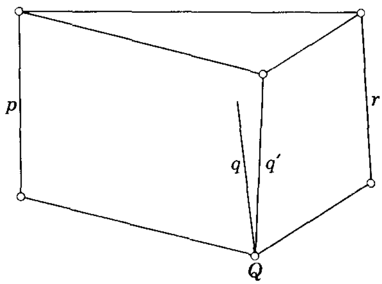



**Figure 13.8a**

Parallelism for lines is transitive in space as well as in a plane:

**13.82** *If $p$ and $q$ are both parallel to $r$, they are parallel to each other.*

*Proof* [Forder **1**, p. 140]. When all three lines are in one plane, this follows at once from 13.11, so let us assume that they are not. For any point $Q$ on $q$, the planes $Qp$ and $Qr$ meet in a line, say $q'$ (Figure 13.8*a*). Any common point of $q'$ and $r$ would lie in both the planes $Qp$, $pr$, and therefore on their common line $p$; this is impossible, since $p$ is parallel to $r$. Hence $q'$ is parallel to $r$. But the only line through $Q$ parallel to $r$ is $q$. Hence $q$ coincides with $q'$, and is coplanar with $p$. Any common point of $p$ and $q$ would lie also on $r$. Hence $p$ and $q$ are parallel.

The transitivity of parallelism provides an alternative proof for 13.81. To establish the impossibility of a point $O$ lying on both planes $\gamma$ and $q'r'$, we imagine two lines through $O$, parallel to $q$ (and $q'$), $r$ (and $r'$). The planes $\gamma$ and $q'r'$, each containing both these lines, would coincide, contradicting our assumption that $A$ does not lie in $\gamma$.

The three face planes $OBC$, $OCA$, $OAB$ of a tetrahedron $OABC$ form with the respectively parallel planes through $A$, $B$, $C$ a *parallelepiped* whose faces are six parallelograms, as in Figure 13.8*b* [Forder **1**, p. 155].

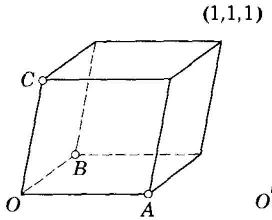



**Figure 13.8b**

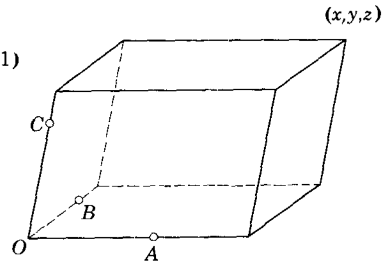



**Figure 13.8c**

It is now easy to build up a three-dimensional theory of dilatations, translations, and vectors. Three vectors $\mathbf{d}$, $\mathbf{e}$, $\mathbf{f}$ are said to be *dependent* if they are coplanar, in which case each is expressible as a linear combination of the other two. Three vectors $\mathbf{e}$, $\mathbf{f}$, $\mathbf{g}$ are said to be *independent* if the only solution of the vector equation

$$x\mathbf{e} + y\mathbf{f} + z\mathbf{g} = \mathbf{0}$$

is $x = y = z = 0$. Three such vectors provide a basis for a system of three-dimensional *affine coordinates*. In fact, if

$$\mathbf{e} = \overrightarrow{OA},\quad \mathbf{f} = \overrightarrow{OB},\quad \mathbf{g} = \overrightarrow{OC},$$

as in Figure 13.8*c*, the general vector $\overrightarrow{OP}$ may be exhibited as a diagonal of the parallelepiped formed by drawing through $P$ three planes parallel to $OBC$, $OCA$, $OAB$. Then

$$\overrightarrow{OP} = x\mathbf{e} + y\mathbf{f} + z\mathbf{g},$$

where the terms of this sum are vectors along three edges of the parallelepiped.

In space, as in a plane, the centroid $P$ of masses $t_i$ at points $A_i$ is determined by a vector $\overrightarrow{OP}$ such that

$$\Sigma t_i \overrightarrow{OP} = \Sigma t_i \overrightarrow{OA_i} \qquad (\Sigma t_i \neq 0).$$

If $\overrightarrow{OA_i} = x_i\mathbf{e} + y_i\mathbf{f} + z_i\mathbf{g}$, we deduce

$$\Sigma t_i \overrightarrow{OP} = \Sigma t_i x_i \mathbf{e} + \Sigma t_i y_i \mathbf{f} + \Sigma t_i z_i \mathbf{g}.$$

Hence, in terms of affine coordinates,

**13.83** *The centroid of $k$ masses $t_i$ $(\Sigma t_i \neq 0)$ at points $(x_i, y_i, z_i)$ $(i = 1, \ldots, k)$ is*

$$\left(\frac{\Sigma t_i x_i}{\Sigma t_i}, \frac{\Sigma t_i y_i}{\Sigma t_i}, \frac{\Sigma t_i z_i}{\Sigma t_i}\right).$$

In particular, if $t_1 + t_2 + t_3 = 1$, the centroid of three masses $t_1, t_2, t_3$ at the points

$$(1, 0, 0), \quad (0, 1, 0), \quad (0, 0, 1)$$

is $(t_1, t_2, t_3)$. Hence

**13.84** *The affine coordinates of any point in the plane $x + y + z = 1$ are the same as its areal coordinates referred to the triangle cut out from this plane by the coordinate planes $x = 0$, $y = 0$, $z = 0$.*

It follows that there is a line

$$\frac{x}{t_1} = \frac{y}{t_2} = \frac{z}{t_3}$$

through the origin (in affine space) for each point with barycentric coordinates $(t_1, t_2, t_3)$. On the other hand, lines lying in the plane $x + y + z = 0$ yield no corresponding points in the parallel plane $x + y + z = 1$, unless we agree to extend the affine plane by postulating a line at infinity

$$t_1 + t_2 + t_3 = 0$$

so as to form the projective plane. This possibility has already been mentioned in § 6.9; we shall explore it more systematically in Chapter 14.

## EXERCISES

1. If a line $a$ is parallel to a plane $\alpha$, and a plane through $a$ meets $\alpha$ in $b$, then $a$ and $b$ are parallel lines. If another plane through $a$ meets $\alpha$ in $c$, then $b$ and $c$ are parallel lines.

2. If $\alpha, \beta, \gamma$ are planes intersecting in lines $\beta \cdot \gamma = a$, $\gamma \cdot \alpha = b$, $\alpha \cdot \beta = c$, and $a$ is parallel to $b$, then $a, b, c$ are all parallel.

3. All the lines through $A$ parallel to $\alpha$ are in a plane parallel to $\alpha$ [Forder **1**, p. 155].

**4.** Each of the six edges of a tetrahedron lies on a plane joining this edge to the midpoint of the opposite edge. The six planes so constructed all pass through one point: the centroid of equal masses at the four vertices.

**5.** Develop the theory of three-dimensional barycentric coordinates referred to a tetrahedron $A_1A_2A_3A_4$.

## 13.9 THREE-DIMENSIONAL LATTICES

*The small parallelepiped built upon the three translations selected as unit translations . . . is known as the unit cell. . . . The entire crystal structure is generated through the periodic repetition, by the three unit translations, of the matter contained within the volume of the unit cell.*

M. J. Buerger (1903 - )
[Buerger **1**, p. 5]

The theory of volume in affine space is more difficult than that of area in the affine plane, because of the complication introduced by M. Dehn's observation that two polyhedra of equal volume are not necessarily derivable from each other by dissection and rearrangement. A valid treatment, suggested by Mrs. Sally Ruth Struik, may be described very briefly as follows. It is found that any two tetrahedra are related by a unique *affinity* $ABCD \to A'B'C'D'$, which transforms the whole space into itself in such a way as to preserve collinearity. In particular, a tetrahedron $ABCC'$ is transformed into $ABC'C$ by the *affine reflection*

$$AB(CC'),$$

which interchanges $C$ and $C'$ while leaving invariant every point in the plane that joins $AB$ to the midpoint of $CC'$. Two tetrahedra are said to have the same *volume* if one can be transformed into the other by an *equiaffinity*: the product of an even number of affine reflections. Such a comparison is easily extended from tetrahedra to parallelepipeds, since a parallelepiped can be dissected into six tetrahedra all having the same volume.

In three dimensions, as in two, a *lattice* may be regarded as the set of points whose affine coordinates are integers. However, as it is independent of the chosen coordinate system, it is more symmetrically described as a discrete set of points whose set of position vectors is *closed under subtraction,* that is, along with any two of the vectors the set includes also their difference. Subtracting any one of the vectors from itself, we obtain the zero vector

$$\mathbf{c} - \mathbf{c} = \mathbf{0}$$

and hence also $\mathbf{0} - \mathbf{b} = -\mathbf{b}$, $\mathbf{a} - (-\mathbf{b}) = \mathbf{a} + \mathbf{b}$, $\mathbf{a} + \mathbf{a} = 2\mathbf{a}$, and so on: the set of vectors, containing the difference of any two, also contains the sum of any two, and all the integral multiples of any one. The lattice is one-, two-, or three-dimensional according to the number of independent vectors. In the three-dimensional case, a set of three independent vectors $\mathbf{e}$, $\mathbf{f}$, $\mathbf{g}$ is called a *basis* for the lattice if all the vectors are expressible in the form

**13.91** $$x\mathbf{e} + y\mathbf{f} + z\mathbf{g},$$

where $x$, $y$, $z$ are integers. If three of these vectors, say $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ form another basis for the same lattice, there must exist 18 integers

$$a_\alpha,\ b_\alpha,\ c_\alpha,\ A_\alpha,\ B_\alpha,\ C_\alpha \qquad (\alpha = 1, 2, 3)$$

such that

$$\mathbf{r}_\alpha = a_\alpha\,\mathbf{e} + b_\alpha\,\mathbf{f} + c_\alpha\,\mathbf{g}, \quad \mathbf{e} = \Sigma A_\alpha\,\mathbf{r}_\alpha, \quad \mathbf{f} = \Sigma B_\alpha\,\mathbf{r}_\alpha, \quad \mathbf{g} = \Sigma C_\alpha\,\mathbf{r}_\alpha$$

and therefore

$$\mathbf{r}_\alpha = a_\alpha \Sigma A_\beta\,\mathbf{r}_\beta + b_\alpha \Sigma B_\beta\,\mathbf{r}_\beta + c_\alpha \Sigma C_\beta\,\mathbf{r}_\beta,$$

whence

$$a_\alpha A_\beta + b_\alpha B_\beta + c_\alpha C_\beta = \begin{cases} 1 & \text{if } \alpha = \beta, \\ 0 & \text{if } \alpha \neq \beta. \end{cases}$$

Since the product of two determinants is obtained by combining the rows of one with the columns of the other, we have

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1.$$

Since the two determinants on the left are integers whose product is 1, each must be $\pm 1$. Conversely, if $a_\alpha$, $b_\alpha$, $c_\alpha$ are given so that their determinant is $\pm 1$, we can derive $A_\alpha$, $B_\alpha$, $C_\alpha$ by "inverting the matrix," and the given basis $\mathbf{e}$, $\mathbf{f}$, $\mathbf{g}$ yields the equally effective basis $\mathbf{r}_\alpha$. Hence

*A necessary and sufficient condition for two triads of independent vectors*

$$\mathbf{e}, \mathbf{f}, \mathbf{g} \quad \textit{and} \quad a_\alpha\,\mathbf{e} + b_\alpha\,\mathbf{f} + c_\alpha\,\mathbf{g} \qquad (\alpha = 1, 2, 3)$$

*to be alternative bases for the same lattice is*

**13.92** $$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \pm 1$$

[cf. Hardy and Wright **1**, p. 28; Neville **1**, p. 5].

In other words, a lattice is derived from any one of its points by applying a *discrete group of translations:* one-, two-, or three-dimensional according as the translations are collinear, coplanar but not collinear, or not coplanar. In the one-dimensional case the generating translation is unique (except that it may be reversed), but in the other cases the two or three generators, that is, the basic vectors, may be chosen in infinitely many ways. When they have been chosen, we can use them to set up a system of affine coordinates so that, in the three-dimensional case, the vector 13.91 goes from the origin (0, 0, 0) to the point $(x, y, z)$, and the lattice consists of the points whose coordinates are integers. The eight points

$$(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1),$$

derived from the eight vectors

$$\mathbf{0}, \quad \mathbf{e}, \quad \mathbf{f}, \quad \mathbf{g}, \quad \mathbf{f} + \mathbf{g}, \quad \mathbf{g} + \mathbf{e}, \quad \mathbf{e} - \mathbf{f}, \quad \mathbf{e} + \mathbf{f} + \mathbf{g},$$

evidently form a parallelepiped, which is a *unit cell* of the lattice. By an argument analogous to that used for a two-dimensional lattice in § 4.1, *any two unit cells for the same lattice have the same volume.*

Any line joining two of the lattice points contains infinitely many of them, forming a one-dimensional sublattice of the three-dimensional lattice. In fact, the line joining (0, 0, 0) and $(x, y, z)$ contains also $(nx, ny, nz)$ for every integer $n$. If $x, y, z$ have the greatest common divisor $d$, the lattice point

$$(x/d, y/d, z/d)$$

lies on this same line, and the corresponding translation generates the group of the one-dimensional lattice. The lattice point $(x, y, z)$ is *visible* if and only if the three integers $x, y, z$ have no common divisor greater than 1.

Any triangle of lattice points determines a plane containing a two-dimensional sublattice. For, if vectors

$$\mathbf{r}_1 = x_1\mathbf{e} + y_1\mathbf{f} + z_1\mathbf{g} \quad \text{and} \quad \mathbf{r}_2 = x_2\mathbf{e} + y_2\mathbf{f} + z_2\mathbf{g}$$

have integral components, so also does $t_1\mathbf{r}_1 + t_2\mathbf{r}_2$ for any integers $t_1$ and $t_2$. The parallel plane through any other lattice point will contain a congruent sublattice. Thus we may regard all the lattice points as being distributed among an infinite sequence of parallel planes, called *rational planes* [Buerger **1**, p. 7].

Any such plane, being the join of three points whose coordinates are integers, has an equation of the form

**13.93** $$Xx + Yy + Zz = N,$$

where the coefficients $X, Y, Z, N$ are integers, so that the intercepts on the coordinate axes have the rational values $N/X$, $N/Y$, $N/Z$. (This is the reason for the name "rational" planes.) We may assume that the greatest common divisor of $X, Y, Z$ is 1; for, any common factor of $X, Y, Z$ would be a factor of $N$ too, and then we could divide both sides of the equation by this number, obtaining a simpler and equally effective equation for the same plane.

Conversely, any such equation (in which the greatest common divisor of $X, Y, Z$ is 1) represents a plane containing a two-dimensional sublattice. This is obvious when $X = 1$, since then we can assign arbitrar integral values to $y$, $z$, and solve 13.93 for $x$. When $X$, $Y$, $Z$ are all greater than 1, we consider the set of numbers

$$xX + yY + zZ,$$

where $x$, $y$, $z$ are variable integers while $X$, $Y$, $Z$ remain constant. This set (like the set of lattice vectors) is an *ideal:* it contains the difference of any two of its members and (therefore) all the multiples of any one. Let $d$ denote its smallest positive member, and $N$ any other member. Then $N$ is a multiple of $d$: for otherwise we could divide $N$ by $d$ and obtain a remainder $N - qd$, which would be a member smaller than $d$. Thus every member of the set is a multiple of $d$. But $X$, $Y$, $Z$ are members. Therefore $d$, being a common divisor, must be equal to 1, and the set simply consists of all the integers. In other words, the equation 13.93 has one integral solution (and therefore infinitely many) [cf. Uspensky and Heaslet **1**, p. 54].

For each triad of integers $X$, $Y$, $Z$, coprime in the above sense (but not necessarily coprime in pairs), we have a sequence of parallel planes 13.93, evenly spaced, one plane for each integer $N$. Since every lattice point lies in one of the planes, the infinite region between any two consecutive planes is completely empty. One of the planes, namely that for which $N = 0$, passes through the origin. The nearest others, given by $N = \pm 1$, are appropriately called *first* rational planes [Buerger **1**, p. 9]. We shall have occasion to consider them again in § 18.3.

## EXERCISES

**1.** How can a parallelepiped be dissected into six tetrahedra all having the same volume?

**2.** Identify the transformation $(x, y, z) \to (x, y, -z)$ with the affine reflection that leaves invariant the plane $z = 0$ while interchanging the points $(0, 0, \pm 1)$.

**3.** A lattice is transformed into itself by the central inversion that interchanges two of its points.

**4.** Every lattice point in a first rational plane is visible.

**5.** Is every rational plane through a visible point a first rational plane?

**6.** Find a triangle of lattice points in the first rational plane

$$6x + 10y + 15z = 1.$$

**7.** Obtain a formula for all the lattice points in this plane.

**8.** The origin is the only lattice point in the plane

$$x + \sqrt{2}y + \sqrt{3}z = 0.$$

# 14
# *Projective geometry*

In affine geometry, as we have seen, parallelism plays a leading role. In projective geometry, on the other hand, there is no parallelism: every pair of coplanar lines is a pair of intersecting lines. The conflict with 12.61 is explained by the fact that the projective plane is not an "ordered" plane. The set of points on a line, like the set of lines through a point, is closed: given three, we cannot pick out one as lying "between" the other two. At first sight we might expect a geometry having no circles, no distances, no angles, no intermediacy, and no parallelism, to be somewhat meagre. But, in fact, a very beautiful and intricate collection of propositions emerges: propositions of which Euclid never dreamed, because his interest in measurement led him in a different direction. A few of these nonmetrical propositions were discovered by Pappus of Alexandria in the fourth century A.D. Others are associated with the names of two Frenchmen: the architect Girard Desargues (1591–1661) and the philosopher Blaise Pascal (1623–1662). Meanwhile, the related subject of perspective [Yaglom **2,** p. 31] had been studied by artists such as Leonardo da Vinci (1452–1519) and Albrecht Dürer (1471–1528).

Kepler's invention of points at infinity made it possible to regard the projective plane as the affine plane plus the line at infinity. A converse relationship was suggested by Poncelet's *Traité des propriétés projectives des figures* (1822) and von Staudt's *Geometrie der Lage* (1847), in which projective geometry appeared as an independent science, making it possible to regard the affine plane as the projective plane minus an arbitrary line $o$, and then to regard the Euclidean plane as the affine plane with a special rule for associating pairs of points on $o$ (in "perpendicular directions") [Coxeter **2,** pp. 115, 138]. This standpoint became still clearer in 1899, when Mario Pieri placed the subject on an axiomatic foundation. Other systems of axioms, slightly different from Pieri's, have been proposed by subsequent authors. The particular system that we shall give in § 14.1 was suggested by Bachmann [**1,** pp. 76–77]. To test the consistency of a system of axioms, we apply it to a "model," in which the primitive concepts are represented by familiar concepts whose properties we are prepared to accept [Coxeter **2,** pp.

186–187]. In the present case a convenient model for the projective plane is provided by the affine plane plus the line at infinity (§ 6.9). We shall extend the barycentric coordinates of § 13.7 to general projective coordinates, so as to eliminate the special role of the line at infinity. The result may be regarded as a purely algebraic model in which a *point* is an ordered triad of numbers $(x_1, x_2, x_3)$, not all zero, with the rule that $(\mu x_1, \mu x_2, \mu x_3)$ is the same point for any $\mu \neq 0$, and a *line* is a homogeneous linear equation. One advantage of this model is that the numbers $x_\alpha$ and $\mu$ are not necessarily real. The chosen axioms are sufficiently general to allow the coordinates to belong to any *field:* instead of real numbers we may use rational numbers, complex numbers, or even a finite field such as the residue classes modulo a prime number. Accordingly we speak of the real projective plane, the rational projective plane, the complex projective plane, or a finite projective plane.

## 14.1 AXIOMS FOR THE GENERAL PROJECTIVE PLANE

*The more systematic course in the present introductory memoir . . . would have been to ignore altogether the notions of distance and metrical geometry. . . . Metrical geometry is a part of descriptive geometry, and descriptive geometry is all geometry.*

Arthur Cayley *(1821 -1895)

The projective plane has already been mentioned in § 6.9. As primitive concepts we take *point, line,* and the relation of *incidence.* If a point and a line are incident, we say that the point lies *on* the line and the line passes *through* the point. The related words *join, meet* (or "intersect"), *concurrent* and *collinear* have their usual meanings. Three non-collinear points are the vertices of a *triangle* whose sides are complete lines. ("Segments" are not defined.) A *complete quadrangle,* its four vertices, its six sides, and its three diagonal points, are defined as in § 1.7. A *hexagon* $A_1B_2C_1A_2B_1C_2$ has six vertices $A_1, B_2, \ldots, C_2$ and six sides

$$A_1B_2,\ B_2C_1,\ C_1A_2,\ A_2B_1,\ B_1C_2,\ C_2A_1.$$

*Opposite* sides are defined in the obvious manner; for example, $A_2B_1$ is opposite to $A_1B_2$. After these preliminary definitions, we are ready for the five axioms.

**AXIOM 14.11** *Any two distinct points are incident with just one line.*

**NOTATION.** The line joining points $A$ and $B$ is denoted by $AB$.

* *Collected Mathematical Papers,* **2** (Cambridge, 1889), p. 592. Cayley, in 1859, used the word "descriptive" where today we would say "projective." His idea of the supremacy of projective geometry must now be regarded as a slight exaggeration. It is true that projective geometry includes the affine, Euclidean and non-Euclidean geometries; but it does not include the general Riemannian geometry, nor topology.

**AXIOM 14.12** *Any two lines are incident with at least one point.*

**THEOREM 14.121** *Any two distinct lines are incident with just one point.*

**NOTATION.** The point of intersection of lines $a$ and $b$ is denoted by $a \cdot b$; that of $AB$ and $CD$ by $AB \cdot CD$. The line joining $a \cdot b$ and $c \cdot d$ is denoted by $(a \cdot b)(c \cdot d)$.

**AXIOM 14.13** *There exist four points of which no three are collinear.*

**AXIOM 14.14** (Fano's axiom) *The three diagonal points of a complete quadrangle are never collinear.*

**AXIOM 14.15** (Pappus's theorem) *If the six vertices of a hexagon lie alternately on two lines, the three points of intersection of pairs of opposite sides are collinear.*

One of the most elegant properties of projective geometry is the *principle of duality,* which asserts (in a projective plane) that every definition remains significant, and every theorem remains true, when we consistently interchange the words *point* and *line* (and consequently interchange *lie on* and *pass through, join* and *intersection, collinear* and *concurrent,* etc.). To establish this principle it will suffice to verify that *the axioms imply their own duals.* Then, given a theorem and its proof, we can immediately assert the dual theorem; for a proof of the latter could be written down mechanically by dualizing every step in the proof of the original theorem.

The dual of Axiom 14.11 is Theorem 14.121, which the reader will have no difficulty in proving (with the help of 14.12). The dual of Axiom 14.12 is one-half of 14.11. The dual of Axiom 14.13 asserts the existence of a *complete quadrilateral,* which is a set of four lines (called *sides*) intersecting in pairs in six distinct points (called *vertices*). Two vertices are said to be *opposite* if they are not joined by a side. The three joins of pairs of opposite vertices are called *diagonals.* If $PQRS$ is a quadrangle with sides

$$p = PQ, \quad q = PS, \quad r = RS, \quad s = QR, \quad w = PR, \quad u = QS,$$

as in Figure 14.1$a$, then $pqrs$ is a quadrilateral with vertices

$$P = p \cdot q, \quad Q = p \cdot s, \quad R = r \cdot s, \quad S = q \cdot r, \quad W = p \cdot r, \quad U = q \cdot s.$$

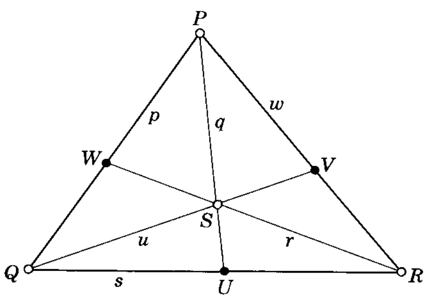



**Figure 14.1a**

Axiom 14.14 tells us that the three diagonal points

$$U = q \cdot s, \quad V = w \cdot u, \quad W = p \cdot r$$

are not collinear. Its dual asserts that the three diagonals of a complete quadrilateral are never concurrent. If this is false, there must exist a particular quadrilateral whose diagonals are concurrent. Let it be *pqrs*, with diagonals

$$u = QS, \quad v = WU, \quad w = PR.$$

Since these are concurrent, the point $w \cdot u = V$ must lie on $v$, contradicting the statement that $U$, $V$, $W$ are not collinear.

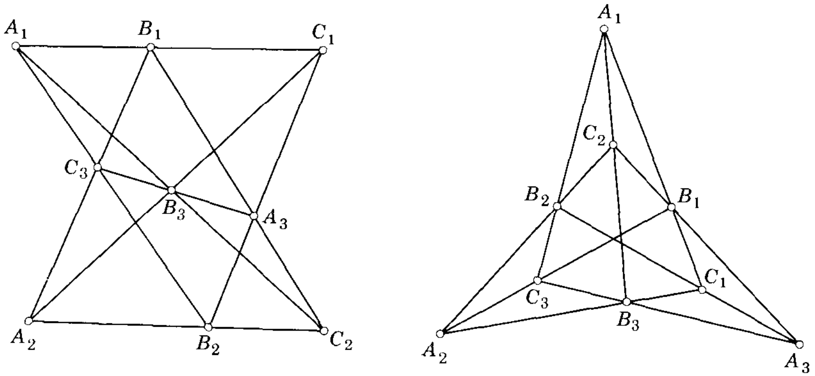



**Figure 14.1b**

Axiom 14.15 involves nine points and nine lines, which can be drawn in many ways (apparently different though projectively equivalent), such as the two shown in Figure 14.1*b*. $A_1B_2C_1A_2B_1C_2$ is a hexagon whose vertices lie alternately on the two lines $A_1B_1C_1$, $A_2B_2C_2$. The points of intersection of pairs of opposite sides are

$$A_3 = B_1C_2 \cdot B_2C_1, \; B_3 = C_1A_2 \cdot C_2A_1, \; C_3 = A_1B_2 \cdot A_2B_1.$$

The axiom asserts that these three points are collinear. Our notation has been devised in such a way that the three points $A_i$, $B_j$, $C_k$ are collinear whenever

$$i + j + k \equiv 0 \qquad (\text{mod } 3).^*$$

Another way to express the same result is to arrange the 9 points in the form of a *matrix*

**14.151**
$$\begin{Vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{Vmatrix}.$$

* Coxeter, Self-dual configurations and regular graphs, *Bulletin of the American Mathematical Society,* **56** (1950), p. 432.

If this were a determinant that we wished to evaluate, we would proceed to multiply the elements in triads. These six "diagonal" triads, as well as the first two rows of the matrix, indicate triads of collinear points. The axiom asserts that the points in the bottom row are likewise collinear. Its inherent self-duality is seen from an analogous matrix of lines

$$\begin{Vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{Vmatrix}.$$

These lines can be picked out in many ways, one of which is

$$\begin{aligned} a_1 &= A_3B_1C_2, & b_1 &= A_1B_3C_2, & c_1 &= A_2B_2C_2, \\ a_2 &= A_2B_3C_1, & b_2 &= A_3B_2C_1, & c_2 &= A_1B_1C_1, \\ a_3 &= A_1B_2C_3, & b_3 &= A_2B_1C_3, & c_3 &= A_3B_3C_3. \end{aligned}$$

This completes our proof of the principle of duality.

## EXERCISES

1. Every line is incident with at least three distinct points. (This statement, and the existence of a nonincident point and line, are sometimes used as axioms instead of 14.13 [Robinson **1**, p. 10; Coxeter **2**, p. 13].)

2. A set of $m$ points and $n$ lines is called a *configuration* $(m_c, n_d)$ if $c$ of the $n$ lines pass through each of the points while $d$ of the $m$ points lie on each of the lines. The four numbers are not independent but satisfy $cm = dn$. The dual of $(m_c, n_d)$ is $(n_d, m_c)$.

In the case of a self-dual configuration, we have $m = n$, $c = d$, and the symbol $(n_d, n_d)$ is conveniently abbreviated to $n_d$. Simple instances are the triangle $3_2$, the complete quadrangle $(4_3, 6_2)$ and the complete quadrilateral $(6_2, 4_3)$. Axiom 14.14 asserts the nonexistence of the *Fano configuration** $7_3$. The points and lines that occur in Axiom 14.15 (Figure 14.1*b*) form the *Pappus configuration* $9_3$, which may be regarded (in how many ways?) as a cycle of three triangles such as

$$A_1B_1C_2, \quad A_2B_2C_3, \quad A_3B_3C_1,$$

each inscribed in the next (cf. Figure 1.8*a*, where $UVW$ is inscribed in $ABC$). The self-duality is evident.

By a suitable change of notation, Axiom 14.15 may be expressed thus: *If AB, CD EF are concurrent, and DE, FA, BC are concurrent, then AD, BE, CF are concurrent.*

3. A particular *finite projective plane,* in which only 13 "points" and 13 "lines" exist, can be defined abstractly by calling the points $P_i$ and the lines $p_i$ ($i = 0, 1, \ldots, 12$) with the rule that $P_i$ and $p_j$ are "incident" if and only if

$$i + j \equiv 0, 1, 3 \text{ or } 9 \pmod{13}.$$

Construct a table to indicate the 4 points on each line and the 4 lines through each point [Veblen and Young **1**, p. 6]. Verify that all the axioms are satisfied; for example, $P_0P_1P_2P_5$ is a complete quadrangle with sides

$$P_0P_1 = p_0, \quad P_0P_2 = p_1, \quad P_1P_5 = p_8, \quad P_0P_5 = p_9, \quad P_2P_5 = p_{11}, \quad P_1P_2 = p_{12}$$

* Coxeter, *Bulletin of the American Mathematical Society,* **56** (1950), pp. 423–425.

and diagonal points $P_3 = p_0 \cdot p_{11}$, $P_4 = p_9 \cdot p_{12}$, $P_8 = p_1 \cdot p_8$. A possible matrix for Axiom 14.15 is

$$\begin{Vmatrix} P_0 & P_2 & P_8 \\ P_3 & P_4 & P_6 \\ P_9 & P_{10} & P_5 \end{Vmatrix}.$$

The first row may be any set of three collinear points. The second row may be any such set on a line not incident with a point in the first row. The last row is then determined; e.g., in the above instance it consists of

$$P_2P_6 \cdot P_4P_8 = P_9, \qquad P_3P_8 \cdot P_0P_6 = P_{10}, \qquad P_0P_4 \cdot P_2P_3 = P_5.$$

This differs from the general "Pappus matrix" 14.151 in that sets of collinear points occur not only in the rows and generalized diagonals but also in the columns. In other words, the 9 points form a configuration which is not merely $9_3$ but $(9_4, 12_3)$. When any one of the 9 points is omitted, the remaining 8 form a self-dual configuration $8_3$ which may be regarded as a pair of mutually inscribed quadrangles (such as $P_0P_9P_5P_8$ and $P_2P_3P_{10}P_6$). [Hilbert and Cohn-Vossen **1**, pp. 101–102.]

**4.** The geometry described in Ex. 3 is known as $PG(2, 3)$. More generally, $PG(2, p)$ is a finite plane in which each line contains $p + 1$ points. Consequently, each point lies on $p + 1$ lines. There are $p^2 + p + 1$ points (and the same number of lines) altogether. In other words, the whole geometry is a configuration $n_d$ with $n = p^2 + p + 1$ and $d = p + 1$. (Actually $p$ is not arbitrary, e.g., although it may be any power of an odd prime, for instance, 5, 7, or 9, it cannot be 6.)* The possibility of such finite planes indicates that the projective geometry defined by Axioms 14.11 to 14.15 is not *categorical:* it is not just one geometry but many geometries, in fact, infinitely many.

**5.** In any finite projective geometry, Sylvester's theorem (§ 4.7) is false.

## 14.2 PROJECTIVE COORDINATES

*Modern algebra does not seem quite so terrifying when expressed in these geometrical terms!*

G. de B. Robinson (1906- )

[Robinson **1**, p. 94]

We saw, in § 13.7, that three real numbers $t_1$, $t_2$, $t_3$ will serve as barycentric coordinates for a point in the affine plane (with respect to any given triangle of reference) if and only if

$$t_1 + t_2 + t_3 \neq 0.$$

Also a linear homogeneous equation 13.72 will serve as the equation for a line if and only if the coefficients $T_1$, $T_2$, $T_3$ are not all equal. The remarks

* By not insisting on Axiom 14.14, we can develop a "geometry of characteristic 2" in which $p$ is a power of 2. By not insisting on Axiom 14.15, we can develop a "non-Desarguesian plane." For the application to mutually orthogonal Latin squares, see Robinson **1**, p. 161, Appendix II.

just after 13.84 indicate that these artificial restrictions will be avoided when we have extended the real affine plane to the real projective plane by adding the line at infinity

**14.21** $$t_1 + t_2 + t_3 = 0$$

and all its points (which are the points at infinity in various directions).

When we interpret $T_1$, $T_2$, $T_3$ as the distances from $A_1$, $A_2$, $A_3$ to the line

$$T_1t_1 + T_2t_2 + T_3t_3 = 0,$$

it is obvious that a parallel line is obtained by adding the same number to all three $T$'s. Hence the point of intersection of two parallel lines satisfies 14.21, that is, it lies on the line at infinity.

To emphasize the fact that, in projective geometry, the line at infinity no longer plays a special role, we shall abandon the barycentric coordinates $(t_1, t_2, t_3)$ in favor of general *projective* coordinates $(x_1, x_2, x_3)$, given by

$$t_1 = \mu_1x_1, \quad t_2 = \mu_2x_2, \quad t_3 = \mu_3x_3,$$

where $\mu_1$, $\mu_2$, $\mu_3$ are constants, $\mu_1\mu_2\mu_3 \neq 0$. Thus $(x_1, x_2, x_3)$ is the centroid of masses $\mu_\alpha x_\alpha$ at $A_\alpha$ ($\alpha = 1, 2, 3$), and the line at infinity has the undistinguished equation

$$\mu_1x_1 + \mu_2x_2 + \mu_3x_3 = 0.$$

The contrast between these two kinds of coordinates may also be expressed as follows. Barycentric coordinates can be referred to any given triangle; the "simplest" points

$$(1, 0, 0), \quad (0, 1, 0), \quad (0, 0, 1)$$

are the vertices, and the *unit point* (1, 1, 1) is the centroid. More usefully, projective coordinates can be referred to any given quadrangle! Taking three of the four vertices to determine a system of barycentric coordinates, suppose the fourth vertex is $(\mu_1, \mu_2, \mu_3)$. By using these $\mu$'s for the transition to projective coordinates, we give this fourth vertex the new coordinates (1, 1, 1). Just as, in affine geometry, all triangles are alike, so in projective geometry *all quadrangles are alike.*

To prove that projective coordinates provide a model (in the augmented affine plane) for the abstract projective plane described in § 14.1, we can take each of our geometric axioms and prove it analytically (i.e., algebraically).

To prove 14.11, we merely have to observe that the line joining points $(y_1, y_2, y_3)$ and $(z_1, z_2, z_3)$ is

**14.22** $$\begin{vmatrix} y_2 & y_3 \\ z_2 & z_3 \end{vmatrix} x_1 + \begin{vmatrix} y_3 & y_1 \\ z_3 & z_1 \end{vmatrix} x_2 + \begin{vmatrix} y_1 & y_2 \\ z_1 & z_2 \end{vmatrix} x_3 = 0$$

(cf. 13.74). Similarly, for 14.12 (or rather, 14.121), the point of intersection of lines $\Sigma Y_\alpha x_\alpha = 0$ and $\Sigma Z_\alpha x_\alpha = 0$ is

$$\left( \begin{vmatrix} Y_2 & Y_3 \\ Z_2 & Z_3 \end{vmatrix}, \begin{vmatrix} Y_3 & Y_1 \\ Z_3 & Z_1 \end{vmatrix}, \begin{vmatrix} Y_1 & Y_2 \\ Z_1 & Z_2 \end{vmatrix} \right).$$

For 14.13, we can use the four points

**14.23** $$(1, 0, 0), \quad (0, 1, 0), \quad (0, 0, 1), \quad (1, 1, 1).$$

The diagonal points of the quadrangle so formed are

$$(0, 1, 1), \quad (1, 0, 1), \quad (1, 1, 0).$$

If these three points lay on a line $\Sigma X_\alpha x_\alpha = 0$, we should have

**14.24** $$X_2 + X_3 = 0, \quad X_3 + X_1 = 0, \quad X_1 + X_2 = 0,$$

whence $X_1 = X_2 = X_3 = 0$, which is absurd. This proves 14.14.

Finally, to prove 14.15 we use the coordinates 14.23 for the four points

$$A_1, \quad A_2, \quad A_3, \quad C_1.$$

On the lines $C_1A_1$, $C_1A_2$, $C_1A_3$, which are

$$x_2 = x_3, \quad x_3 = x_1, \quad x_1 = x_2,$$

we take the points $B_1$, $B_3$, $B_2$ to be

$$(p, 1, 1), \quad (1, q, 1), \quad (1, 1, r).$$

The three lines $A_3B_1$, $A_1B_3$, $A_2B_2$, being

$$x_1 = px_2, \quad x_2 = qx_3, \quad x_3 = rx_1,$$

all pass through the same point $C_2$ if

**14.25** $$pqr = 1.$$

The three lines $A_3B_3$, $A_2B_1$, $A_1B_2$, being

$$x_2 = qx_1, \quad x_1 = px_3, \quad x_3 = rx_2,$$

all pass through the same point $C_3$ if

$$qpr = 1.$$

Since this condition agrees with 14.25, the proof is complete. However, it is important to observe that the above deduction can be carried through in the more general situation where the coordinates belong not to a *field* but to an arbitrary *division ring* [Birkhoff and MacLane **1**, p. 222]. We can still speak of points and lines, but Axiom 14.15 will have to be replaced by a weaker statement if the coordinate ring includes elements $p$ and $q$ such that

$$pq \neq qp.$$

For instance, we might have $p = k$ and $q = j$ in a "quaternion geometry" whose coordinates are based on "units" $i, j, k$ satisfying

$$i^2 = j^2 = k^2 = ijk = -1.$$

When the $A$'s and $B$'s are so chosen, 14.15 is false. We have thus established an important connection between geometry and algebra: Hilbert's discovery that, when homogeneous coordinates are used in a plane satisfying the first four axioms, *Pappus's theorem is equivalent to the commutative law for multiplication.*

### EXERCISES

**1.** Given five points, no three collinear, we can assign the coordinates 14.23 to any four of them, and then the coordinates $(x_1, x_2, x_3)$ of the fifth are definite (apart from the possibility of multiplying all by the same constant). If the mutual ratios of the three $x$'s are rational, we can multiply by a "common denominator" so as to make them all integral. In this case we can derive the fifth point from the first four by a linear construction, involving a finite sequence of operations of joining two known points or taking the point of intersection of two known lines. Devise such a construction for the point (1, 2, 3).

**2.** The four points $(1, \pm 1, \pm 1)$ form a complete quadrangle whose diagonal triangle is the triangle of reference.

**3.** A configuration $8_3$, consisting of two mutually inscribed quadrangles, exists in the complex projective plane, but not in the real projective plane. When it does exist, its eight points appear in four pairs of "opposites" whose joins are concurrent. The complete figure is a $(9_4, 12_3)$. *Hint*: Let the two quadrangles be $P_0P_2P_4P_6$ and $P_1P_3P_5P_7$, so that the sets of three collinear points are

$$P_0P_1P_3,\ P_1P_2P_4,\ P_2P_3P_5,\ P_3P_4P_6,\ P_4P_5P_7,\ P_5P_6P_0,\ P_6P_7P_1,\ P_7P_0P_2.$$

Take $P_0P_1P_2$ as triangle of reference and let $P_3$, $P_4$, $P_7$ be (1, 1, 0), (0, 1, 1), $(1, 0, x)$. Deduce that $P_5$ and $P_6$ are $(1, 1, x + 1)$ and $(1, x + 1, x)$. Obtain an equation for $x$ from the collinearity of $P_0P_5P_6$.

**4.** If $p$ is an odd prime, a finite projective plane $PG(2, p)$ can be obtained by taking the coordinates to belong to the field $GF(p)$ which consists of the $p$ residues (or, strictly, residue classes) modulo $p$ [Ball **1**, pp. 60–61]. For instance, the appropriate "finite arithmetic" for $PG(2, 3)$ consists of symbols 0, 1, 2 which behave like ordinary integers except that

$$1 + 2 = 0 \quad \text{and} \quad 2 \times 2 = 1.$$

In the notation of Ex. 3 at the end of § 14.1, take $P_0P_1P_2$ to be the triangle of reference and $P_5$ the unit point (1, 1, 1). Find coordinates for the remaining points, and equations for the lines.

Finite planes, and the analogous finite $n$-spaces $PG(n, p)$, were discovered by von Staudt* and rediscovered by Fano. Von Staudt took $n$ to be 2 or 3. Fano took $p$ to be a prime. The generalization $PG(n, p^k)$ is credited to Veblen and Bussey.

**5.** Taking the coordinates to belong to $GF(2)$, which consists of the two "numbers" 0 and 1 with the rule for addition

$$1 + 1 = 0,$$

we obtain a finite "geometry" in which the diagonal points of a complete quadrangle are always collinear! Our proof of 14.14 breaks down because now the equations 14.24

* K. G. C. von Staudt, *Beiträge zur Geometrie der Lage*, vol. I (Nürnberg, 1856), pp. 87–88; Gino Fano, *Giornale di Matematiche*, **30** (1892), pp. 114–124; Veblen and Bussey, *Transactions of the American Mathematical Society*, **7** (1906), pp. 241–259.

have not only the inadmissible solution $X_1 = X_2 = X_3 = 0$ but also the significant solution $X_1 = X_2 = X_3 = 1$, which yields the line

$$x_1 + x_2 + x_3 = 0.$$

This $PG(2, 2)$ can be described abstractly by calling its seven points $P_i$ and its seven lines $p_i$ ($i = 0, 1, \ldots, 6$) with the rule that $P_i$ and $p_j$ are incident if and only if

$$i + j \equiv 0, 1 \text{ or } 3 \qquad (\text{mod } 7).$$

## 14.3 DESARGUES'S THEOREM

*The fundamental idea for this pure geometry came from the desire of Renaissance painters to produce a "visual" geometry. How do things really look, and how can they be presented on the plane of the drawing? For example, there will be no parallel lines, since such lines appear to the eye to converge.*

S. H. Gould (1909- )
[Gould **1**, p. 298]

Two triangles, with their vertices named in a particular order, are said to be *perspective from a point* (or briefly, "perspective") if their three pairs of corresponding vertices are joined by concurrent lines. For instance, in Figure 14.1*b*, the triangles $A_1A_2A_3$ and $B_1B_3B_2$ (*sic*) are perspective from $C_1$. By permuting the vertices of $B_1B_3B_2$ cyclically, either forwards or backwards, we see that the same two triangles are also perspective from $C_2$ or $C_3$. In fact, one of the neatest statements of Axiom 14.15 [see Veblen and Young **1**, p. 100] is:

*If two triangles are doubly perspective they are trebly perspective.*

Dually, two triangles are said to be *perspective from a line* if their three pairs of corresponding sides meet in collinear points. It was observed by G. Hessenberg* that our axioms suffice for a proof of

**DESARGUES'S THEOREM.** *If two triangles are perspective from a point they are perspective from a line, and conversely.*

The details are as follows. Let two triangles $PQR$ and $P'Q'R'$ be perspective from $O$, as in Figure 14.3*a*, and let their corresponding sides meet in points

$$D = QR \cdot Q'R', \quad E = RP \cdot R'P', \quad F = PQ \cdot P'Q'.$$

We wish to prove that $D$, $E$, $F$ are collinear. After defining four further points

$$S = PR \cdot Q'R', \qquad T = PQ' \cdot OR,$$
$$U = PQ \cdot OS, \qquad V = P'Q' \cdot OS,$$

we have, in general,† enough triads of collinear points to make three applications of Axiom 14.15. The "matrix" notation enables us to write simply

* *Mathematische Annalen*, **61** (1905), pp. 161–172.

† ... Birkhäuser, Basel, 1963, pp.

$$\begin{Vmatrix} O & Q & Q' \\ P & S & R \\ D & T & U \end{Vmatrix}, \quad \begin{Vmatrix} O & P & P' \\ Q' & R' & S \\ E & V & T \end{Vmatrix}, \quad \begin{Vmatrix} P & Q' & T \\ V & U & S \\ D & E & F \end{Vmatrix}.$$

The last row of the last matrix exhibits the desired collinearity.

The converse follows by the principle of duality.

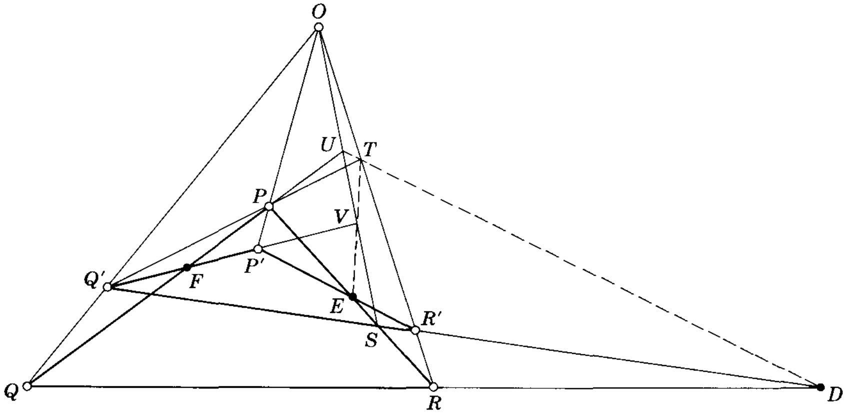



**Figure 14.3a**

## EXERCISES

**1.** The triangle $(p, 1, 1)\ (1, q, 1)\ (1, 1, r)$ is perspective with the triangle of reference from the unit point $(1, 1, 1)$. Pairs of corresponding sides meet in the three collinear points

$$(0, q-1, 1-r), \quad (1-p, 0, r-1), \quad (p-1, 1-q, 0).$$

**2.** Desargues's theorem involves 10 points and 10 lines, forming a configuration $10_3$. To obtain a symmetrical notation, consider triangles $P_{14}P_{24}P_{34}$ and $P_{15}P_{25}P_{35}$, perspective from a point $P_{45}$ and consequently from a line $P_{23}P_{31}P_{12}$. Then three points $P_{ij}$ are collinear if their subscripts involve just three of the numbers 1, 2, 3, 4, 5. If the remaining two of the five numbers are $k$ and $l$, we may call the line $p_{kl}$. Then the same two triangles may be described as $p_{15}\,p_{25}\,p_{35}$ and $p_{14}\,p_{24}\,p_{34}$, perspective from the line $p_{45}$.

**3.** In the finite projective plane $PG(2, 3)$, the two triangles $P_1P_2P_7$ and $P_3P_8P_4$ are perspective from the point $P_0$ and from the line $P_9P_{12}P_{10}$. Identify the remaining points in Figure 14.3*a*. (In this special geometry, $U$ and $V$ both coincide with $F$, which is not surprising in view of the fact that Figure 14.3*a* involves 14 points whereas the whole plane contains only 13.)

## 14.4 QUADRANGULAR AND HARMONIC SETS

Desargues's theorem enables us to prove an important property of a *quadrangular set* of points, which is the section of the six sides of a complete quadrangle by any line that does not pass through a vertex:

**14.41** *Each point of a quadrangular set is uniquely determined by the remaining points.*

*Proof.* Let $PQRS$ be a complete quadrangle whose sides $PS$, $QS$, $RS$, $QR$, $RP$, $PQ$ meet a line $g$ (not through a vertex) in six points $A$, $B$, $C$, $D$, $E$, $F$, certain pairs of which may possibly coincide. (The first three points come from three sides all containing the same vertex $S$; the last three from the respectively opposite sides, which form the triangle $PQR$.) To show that $F$ is uniquely determined by the remaining five points, we set up another quadrangle $P'Q'R'S'$ whose first five sides pass through $A$, $B$, $C$, $D$, $E$, as in Figure 14.4*a*. Since the two triangles $PRS$ and $P'R'S'$ are perspective from the line $g$, the converse of Desargues's theorem tells us that they are also perspective from a point; thus $PP'$ passes through the point $O = RR' \cdot SS'$. Similarly, the perspective triangles $QRS$ and $Q'R'S'$ show that $QQ'$ passes through this same point $O$. In fact, all the four lines $PP'$ $QQ'$, $RR'$, $SS'$ pass through $O$, so that $PQRS$ and $P'Q'R'S'$ are "persp tive quadrangles." By the direct form of Desargues's theorem, the trian $PQR$ and $P'Q'R'$, which are perspective from the point $O$, are also persp tive from the line $DE$, which is $g$; that is, the sides $PQ$ and $P'Q'$ both m $g$ in the same point $F$.

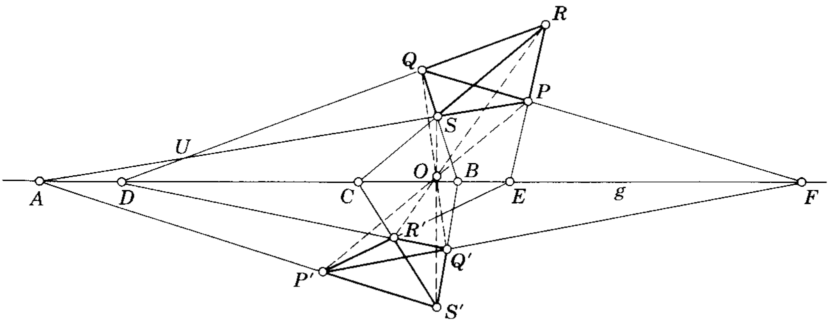



**Figure 14.4a**

We shall find it convenient to use the symbol

$$(AD)\ (BE)\ (CF)$$

to denote the statement that the six points form a quadrangular set in the above manner. This statement is evidently unchanged if we apply any permutation to $ABC$ and the same permutation to $DEF$. It is also equivalent to any of

$$(AD)\ (EB)\ (FC),\quad (DA)\ (BE)\ (FC),\quad (DA)\ (EB)\ (CF).$$

To obtain other permutations we need a new quadrangle. With the ex-

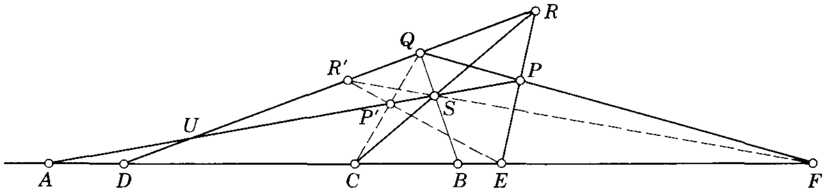



**Figure 14.4b**

ercise of some ingenuity we can retain two of the four old vertices, say $Q$ and $S$. Defining

$$R' = QR \cdot SF, \quad P' = PS \cdot QC,$$

as in Figure 14.4*b*, we apply Axiom 14.15 to the hexagon $PRQCFS$ according to the scheme

$$\left\| \begin{matrix} P & F & Q \\ C & R & S \\ R' & P' & E \end{matrix} \right\|,$$

with the conclusion that $R'P'$ passes through $E$. Just as the quadrangle $PQRS$ yields $(AD)\ (BE)\ (CF)$, the quadrangle $P'QR'S$ yields $(AD)\ (BE)\ (FC)$. In other words, the statement $(AD)\ (BE)\ (CF)$ implies $(AD)\ (BE)\ (FC)$, and hence also

**14.42** $(AD)\ (BE)\ (CF)$ *implies* $(DA)\ (EB)\ (FC)$.

In the important special case $(AA)\ (BB)\ (CF)$, which is abbreviated to

$$\mathrm{H}(AB, CF),$$

we say that the four points form a *harmonic set,* or, more precisely, that $F$ is the *harmonic conjugate* of $C$ with respect to $A$ and $B$. This means that $A$ and $B$ are two of the three diagonal points of a quadrangle while $C$ and $F$ lie respectively on the remaining sides, that is, on the sides that pass through the third diagonal point. Axiom 14.14 tells us that the harmonic conjugates $C$ and $F$ are distinct (except in the degenerate case when they coincide with $A$ or $B$).

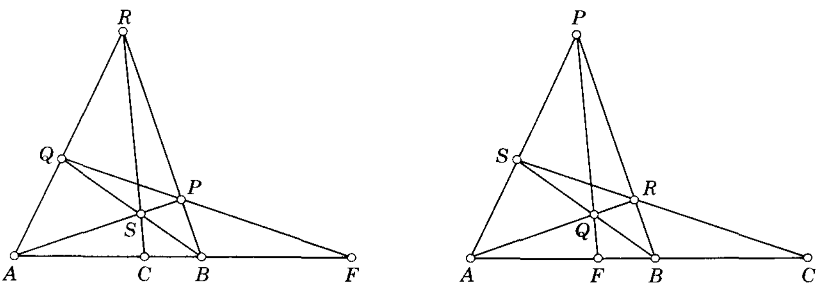



**Figure 14.4c**

**EXERCISES**

1. $\mathsf{H}(AB, CF)$ is equivalent to $\mathsf{H}(BA, CF)$ or $\mathsf{H}(AB, FC)$ or $\mathsf{H}(BA, FC)$.

2. Describe in detail a construction for the harmonic conjugate of $C$ with respect to two given points $A$ and $B$ (on a line through $C$, as in Figure 14.4*c*).

3. The harmonic conjugate of $(0, 1, \lambda)$ with respect to $(0, 1, 0)$ and $(0, 0, 1)$ is $(0, 1, -\lambda)$.

4. In $PG(2, 3)$ (see Ex. 3 at the end of § 14.1), every set of four collinear points is a harmonic set in every order; e.g., $\mathsf{H}(P_0P_1, P_3P_9)$, $\mathsf{H}(P_0P_3, P_9P_1)$, $\mathsf{H}(P_0P_9, P_1P_3)$.

5. In Figure 6.6*a*, $\mathsf{H}(AA', A_1A_2)$. Deduce the metrical definition

$$\frac{AA_1}{A_1A'} = \frac{AA_2}{A'A_2}$$

for a harmonic set. (*Hint*: Defining $E'$ as in Ex. 4 at the end of § 6.6, consider the quadrangle formed by $P$, $E$, $E'$ and the point at infinity on $A_1P$.)

## 14.5 PROJECTIVITIES

A *range* is the set of all points on a line. Dually, a *pencil* is the set of all lines through a point. Ranges and pencils are instances of *one-dimensional forms*. We shall often have occasion to consider a (one-to-one) correspondence between two one-dimensional forms. The simplest possible correspondence between a range and a pencil arises when the lines of the pencil join the points of the range to another point, so that the range is a *section* of the pencil. The correspondence between two ranges that are sections of one pencil by two distinct lines is called a *perspectivity*; in such a case we write

$$X \;\overset{}{\underset{\wedge}{=}}\; X' \quad \text{or} \quad X \;\overset{O}{\underset{\wedge}{=}}\; X',$$

meaning that, if $X$ and $X'$ are corresponding points of the two ranges, their join $XX'$ continually passes through a fixed point $O$, which we call the *center* of the perspectivity. There is naturally also a dual kind of perspectivity relating pencils instead of ranges.

The product of any number of perspectivities is called a *projectivity*. Two ranges (or pencils) related by a projectivity are said to be *projectively related*, and we write

$$X \;\underset{\wedge}{-}\; X'.$$

For instance, in the circumstances illustrated in Figure 14.5*a*,

$$ABCD \;\overset{O}{\underset{\wedge}{=}}\; A_0B_0C_0D_0 \;\overset{O_1}{\underset{\wedge}{=}}\; A'B'C'D', \qquad ABCD \;\underset{\wedge}{-}\; A'B'C'D'.$$

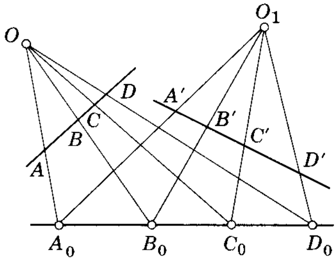



**Figure 14.5a**

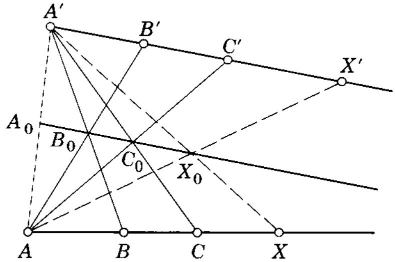



**Figure 14.5b**

Analogously, we can define a projectivity relating a range to a pencil, or vice versa.

Given three distinct points $A$, $B$, $C$ on a line, and three distinct points $A'$, $B'$, $C'$ on another line, we can relate them by a pair of perspectivities in the manner of Figure 14.5*b*, where the *axis* (or "intermediary line") of the projectivity joins the points

$$B_0 = AB' \cdot BA', \quad C_0 = AC' \cdot CA',$$

so that

$$ABC \overset{A'}{\underset{\wedge}{=}} A_0B_0C_0 \overset{A}{\underset{\wedge}{=}} A'B'C'.$$

For each point $X$ on $AB$ we obtain a corresponding point $X'$ on $A'B'$ by joining $A$ to the point $X_0 = A'X \cdot B_0C_0$, so that

$$ABCX \overset{A'}{\underset{\wedge}{=}} A_0B_0C_0X_0 \overset{A}{\underset{\wedge}{=}} A'B'C'X'.$$

By Axiom 14.15, the axis $B_0C_0$, being the "Pappus line" of the hexagon $AB'CA'BC'$, contains the point $BC' \cdot CB'$. Similarly, it contains the point of intersection of the "cross joins" of any two pairs of corresponding points. In particular, we could have derived the same point $X'$ from a given point $X$ by using perspectivities from $B'$ and $B$ (or any other pair of corresponding points) instead of $A'$ and $A$.

It can be proved [Baker **1**, pp. 62–64; Robinson **1**, pp. 24–36] that the product of any number of perspectivities can be reduced to such a product of two, provided the initial and final ranges are not on the same line. In other words,

**14.51** *Any projectivity relating ranges on two distinct lines is expressible as the product of two perspectivities whose centers are corresponding points (in reversed order) of the two related ranges.*

To relate two triads of distinct points $ABC$ and $A'B'C'$ on one line, we may use an arbitrary perspectivity $ABC \underset{\wedge}{=} A_1B_1C_1$ to obtain a triad on another line, and then relate $A_1B_1C_1$ to $A'B'C'$ as in 14.51. Hence

**14.52** *It is possible, by a sequence of not more than three perspectivities, to relate any three distinct collinear points to any other three distinct collinear points.*

A projectivity $X \barwedge X'$ on one line may have one or more *invariant* points (such that $X = X'$). If it has more than two invariant points, it is merely the identity, $X \barwedge X$. In fact, the above construction for a projectivity

$$ABCX \barwedge ABCX'$$

on one line involves four points on another line such that

$$ABCX \doublebarwedge A_1B_1C_1X_1 \barwedge ABCX'.$$

By 14.51, there is essentially only one projectivity $A_1B_1C_1 \barwedge ABC$. We have thus proved

**THE FUNDAMENTAL THEOREM OF PROJECTIVE GEOMETRY.** *A projectivity is determined when three points of one range and the corresponding three points of the other are given.*

If a projectivity relating ranges on two distinct lines has an invariant point $A$, this point, belonging to both ranges, must be the common point of the two lines, as in Figure 14.5*c*. Let $B$ and $C$ be any other points of the first range, $B'$ and $C'$ the corresponding points of the second. The fundamental theorem tells us that the perspectivity

$$ABC \overset{O}{\doublebarwedge} AB'C',$$

where $O = BB' \cdot CC'$, is the same as the given projectivity $ABC \barwedge AB'C'$. Hence

**14.53** *A projectivity between two distinct lines is equivalent to a perspectivity if and only if their point of intersection is invariant.*

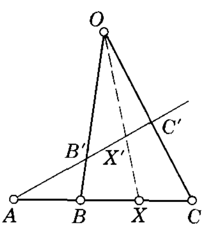



**Figure 14.5c**

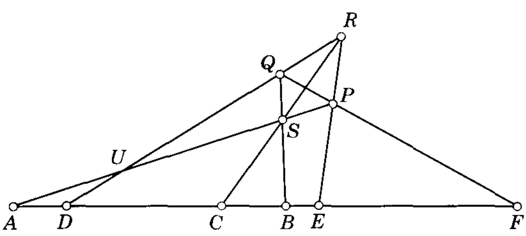



**Figure 14.5d**

Returning to the notion of a projectivity between ranges on one line (i.e., a projective transformation of the line into itself), we recall that, if such a transformation is not merely the identity, it cannot have more than two invariant points. It is said to be *elliptic, parabolic,* or *hyperbolic* according as the number of invariant points is 0, 1, or 2. When coordinates are used, invariant points arise from roots of quadratic equations; thus elliptic projectivities do not occur in complex geometry, but

$$ABC \;\overline{\wedge}\; BCA$$

is elliptic in real geometry [Coxeter **2**, p. 48].

Figure 14.5*d* (cf. 14.4*a*) suggests a simple construction for a hyperbolic projectivity $ABF \;\overline{\wedge}\; ACE$ in which one of the invariant points is given:

$$ABF \overset{Q}{\underset{\wedge}{\equiv}} ASP \overset{R}{\underset{\wedge}{\equiv}} ACE.$$

Here $S$ and $P$ may be any two points collinear with $A$, and then the other two vertices of the quadrangle are

$$Q = BS \cdot FP, \quad R = CS \cdot EP.$$

The second invariant point is evidently $D$, on $QR$. When the same projectivity is expressed in the form $ADB \;\overline{\wedge}\; ADC$ (that is, when both invariant points are given), we have the analogous construction

$$ADB \overset{Q}{\underset{\wedge}{\equiv}} AUS \overset{R}{\underset{\wedge}{\equiv}} ADC,$$

where $U = AS \cdot QD$. This can still be carried out if $A$ and $D$ coincide (i.e., if $g$ passes through the diagonal point $U = PS \cdot QR$ of the quadrangle), in which case we have the parabolic projectivity

$$AAB \;\overline{\wedge}\; AAC$$

[Coxeter **2**, p. 50].

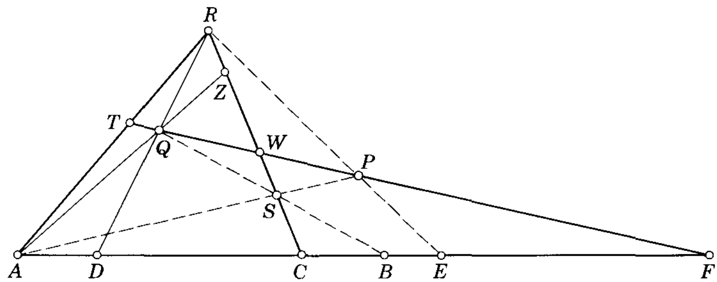



**Figure 14.5e**

An *involution* is a projectivity of period 2, that is, a projectivity which *interchanges* pairs of points. Figure 14.5*e* is derived from Figure 14.5*d* by adding extra points $T$, $W$, $Z$. We may imagine this figure to have been derived from *any* four given collinear points $A$, $C$, $D$, $F$ by taking a point $R$ outside their line, letting the joins $RA$, $RD$, $RC$ meet an arbitrary line through $F$ in $T$, $Q$, $W$, respectively, and then taking $Z = AQ \cdot RC$. Since

$$ADCF \overset{Q}{\underset{\wedge}{=}} ZRCW \overset{A}{\underset{\wedge}{=}} QTFW \overset{R}{\underset{\wedge}{=}} DAFC,$$

we have

**14.54** $$ADCF \;\overline{\wedge}\; DAFC.$$

But, by the fundamental theorem, there is only one projectivity $ADC \;\overline{\wedge}\; DAF$. Hence, if a projectivity interchanging $A$ and $D$ transforms $C$ into $F$, it *interchanges* $C$ *and* $F$. In other words,

**14.55** *Any projectivity that interchanges two points is an involution.*

Applying the same set of three perspectivities to another point $B$, we have

$$B \overset{Q}{\underset{\wedge}{=}} S \overset{A}{\underset{\wedge}{=}} P \overset{R}{\underset{\wedge}{=}} E.$$

Since Q($ABC$, $DEF$), we have now proved the theorem of the quadrangular set:

**14.56** *The three pairs of opposite sides of a quadrangle meet any line (not through a vertex) in three pairs of an involution.*

Combining this with 14.55, we have an alternative proof for 14.42 [Veblen and Young **1,** p. 101].

Since the involution $ACD \;\overline{\wedge}\; DFA$ is determined by its pairs $AD$ and $CF$ (or any other two of its pairs), it is conveniently denoted by

$$(AD)(CF)$$

or $(DA)(CF)$ or $(CF)(AD)$, etc. Thus $(AD)(BE)(CF)$ implies that the pair $BE$ belongs to $(AD)(CF)$, and $CF$ to $(AD)(BE)$, and $AD$ to $(BE)(CF)$. The points in a pair are not necessarily distinct. When $A = D$ and $B = E$, so that H($AB$, $CF$), we have the hyperbolic involution $(AA)(BB)$ which interchanges pairs of harmonic conjugates with respect to $A$ and $B$. Since this same involution is expressible as $(AA)(CF)$,

**14.57** *If an involution has one invariant point, it has another, and consists of the correspondence between harmonic conjugates with respect to these two points.*

It follows that there is no parabolic involution.

## EXERCISES

**1.** Let the lines $OA, OB, \ldots, O_1A', O_1B', \ldots$ and $A_0B_0$ in Figure 14.5$a$ be denoted by $a, b, \ldots, a', b', \ldots$ and $o$. Use the principle of duality to justify the notation

$$abcd \overset{o}{\underset{\wedge}{=}} a'b'c'd'.$$

**2.** The harmonic property is invariant under a projectivity: if H($AB$, $CF$) and $ABCF \;\overline{\wedge}\; A'B'C'F'$, then H($A'B'$, $C'F'$) [Coxeter **2,** p. 23].

**3.** H($AB$, $CF$) implies H($CF$, $AB$). (*Hint*: By 14.54, $ACBF \;\overline{\wedge}\; CAFB$.)

**4.** Draw a quadrangle and its section, as in Figure 14.5*d*. Take an arbitrary point $X$ on $g$ and construct the corresponding point $X'$ in the hyperbolic projectivity

$$ABF \barwedge ACE.$$

Do the same for $ADB \barwedge ADC$, and draw the modified figure that is appropriate for the parabolic projectivity $AAB \barwedge AAC$.

**5.** Two perspectivities cannot suffice for the construction of an elliptic projectivity.

**6.** In the notation of Figure 14.4*b*,

$$ADCF \overset{Q}{\doublebarwedge} AUP'P \overset{E}{\doublebarwedge} DUR'R \overset{S}{\doublebarwedge} DAFC.$$

**7.** Any projectivity may be expressed as the product of two involutions [Coxeter **2,** p. 54].

**8.** The projectivities on the line $x_3 = 0$ are the linear transformations

$$\mu x'_1 = c_{11}x_1 + c_{12}x_2,$$

$$\mu x'_2 = c_{21}x_1 + c_{22}x_2,$$

where $c_{11}c_{22} \neq c_{12}c_{21}$. Under what circumstances is such a projectivity (i) parabolic, (ii) an involution?

## 14.6 COLLINEATIONS AND CORRELATIONS

A *collineation* is a transformation (of the plane) which transforms collinear points into collinear points. Thus it transforms lines into lines, ranges into ranges, pencils into pencils, quadrangles into quadrangles, and so on. A *projective collineation* is a collineation which transforms every one-dimensional form projectively.

**14.61** *Any collineation that transforms one range into a projectively related range is a projective collineation.*

*Proof* [Bachmann **1,** p. 85]. Let the given collineation transform the range of points $X$ on a certain line $a$ into a projectively related range of points $X'$ on the corresponding line $a'$, and let it transform the points $Y$ on another line $b$ into corresponding points $Y'$ on $b'$. Any perspectivity relating $X$ and $Y$ will be transformed into a perspectivity relating $X'$ and $Y'$. Hence

$$Y \doublebarwedge X \barwedge X' \doublebarwedge Y',$$

so that the collineation induces a projectivity $Y \barwedge Y'$ between the points of $b$ and $b'$, as desired.

It follows that a projective collineation is determined when two corresponding quadrangles (or quadrilaterals) are given [Coxeter **2,** p. 60].

A *perspective collineation* with center $O$ and axis $o$ is a collineation which leaves invariant all the lines through $O$ and all the points on $o$. (By 14.61, every perspective collineation is a projective collineation.) Following Sophus Lie (1842–1899), we call a perspective collineation an *elation* or a *homology* according as the center and axis are or are not incident. A *harmonic* homology is the special case when corresponding points $A$ and $A'$, on a line $a$ through $O$, are harmonic conjugates with respect to $O$ and $o \cdot a$. Every projective collineation of period 2 is a harmonic homology [Coxeter **2,** p. 64].

We have seen that a collineation is a point-to-point and line-to-line transformation which preserves incidences. Somewhat analogously, a *correlation* is a point-to-line and line-to-point transformation which dualizes incidences: it transforms points $A$ into lines $a'$, and lines $b$ into points $B'$, in such a way that $a'$ passes through $B'$ if and only if $A$ lies on $b$. Thus a correlation transforms collinear points into concurrent lines (and vice versa), ranges into pencils, quadrangles into quadrilaterals, and so on. A *projective correlation* is a correlation that transforms every one-dimensional form projectively. In a manner resembling the proof of 14.61, we can establish

**14.62** *Any correlation that transforms one range into a projectively related pencil (or vice versa) is a projective correlation.*

It follows that a projective correlation is determined when a quadrangle and the corresponding quadrilateral are given [Coxeter **2,** p. 66].

A *polarity* is a projective correlation of period 2. In general, a correlation transforms a point $A$ into a line $a'$ and transforms this line into a new point $A''$. When the correlation is of period two, $A''$ always coincides with $A$ and we can simplify the notation by omitting the prime ($'$). Thus a polarity relates $A$ to $a$, and vice versa. Following J. D. Gergonne (1771–1859), we call $a$ the *polar* of $A$, and $A$ the *pole* of $a$. Clearly, the polars of all the points on $a$ form a projectively related pencil of lines through $A$.

Since a polarity dualizes incidences, if $A$ lies on $b$, $a$ passes through $B$. In this case we say that $A$ and $B$ are *conjugate points*, $a$ and $b$ are *conjugate lines*. It may happen that $A$ and $a$ are incident, so that each is *self-conjugate*. We can be sure that this does not always happen, for it is easy to prove that the join of two self-conjugate points cannot be a self-conjugate line. It is slightly harder to prove that no line can contain more than two self-conjugate points [Coxeter **2,** p. 68]. The following theorem will be used in § 14.7:

**14.63** *A polarity induces an involution of conjugate points on any line that is not self-conjugate.*

*Proof.* On a non-self-conjugate line $a$, the projectivity $X \barwedge a \cdot x$ (Figure 14.6*a*) transforms any non-self-conjugate point $B$ into another point $C = a \cdot b$, whose polar is $AB$. The same projectivity transforms $C$ into $B$. Since it interchanges $B$ and $C$, it must be an involution.

Dually, $x$ and $AX$ are paired in the involution of conjugate lines through $A$.

Such a triangle $ABC$, in which each vertex is the pole of the opposite side (so that any two vertices are conjugate points, and any two sides are conjugate lines), is said to be *self-polar*. If $P$ is any point not on a side, its polar $p$ does not pass through a vertex, and the polarity may be described as the unique projective correlation that transforms the quadrangle $ABCP$ into the quadrilateral $abcp$. An appropriate symbol, analogous to the symbol $(AB)(PQ)$ for an involution, is

$$(ABC)(Pp).$$

Thus any triangle $ABC$, any point $P$ not on a side, and any line $p$ not through a vertex, determine a definite polarity $(ABC)(Pp)$, in which the polar $x$ of an arbitrary point $X$ can be constructed by simple incidences. As a first step towards this construction we need the following lemma:*

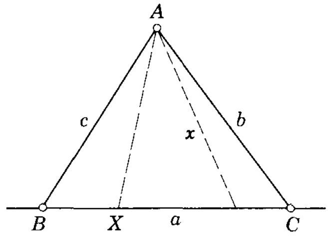



**Figure 14.6a**

**14.64** *If the polars of the vertices of a triangle are distinct from the respectively opposite sides, they meet these sides in three collinear points.*

*Proof.* Let $APX$ be a triangle whose sides $PX$, $XA$, $AP$ meet the polars $a$, $p$, $x$ of its vertices in points $A_1$, $P_1$, $X_1$, as in Figure 14.6*b*. The polar of $X_1 = x \cdot AP$ is, of course, $x_1 = X(a \cdot p)$. Define also the extra points $P' = a \cdot AP$, $X' = a \cdot AX$ and their polars $p' = A(a \cdot p)$, $x' = A(a \cdot x)$. By 14.54 and the polarity, we have

$$AP'PX_1 \barwedge P'AX_1P \barwedge p'ax_1p \barwedge AX'XP_1.$$

By 14.53, $AP'PX_1 \doublebarwedge AX'XP_1$. Since the center of this perspectivity is $P'X' \cdot PX = A_1$, the three points $A_1$, $P_1$, $X_1$ are collinear, as desired.

We are now ready for the construction (Figure 14.6*c*):

**14.65** *In the polarity $(ABC)(Pp)$, the polar of a point $X$ (not on $AP$, $BP$, or $p$) is the line $X_1X_2$ determined by*

* This is known as Chasles's theorem. The proof given in *The Real Projective Plane* [Coxeter **2**, p. 71] suffices for real geometry but not for the more general geometry which is developed here. Lemma 5.54 of that book is false in the finite geometry $PG(2, 3)$, which admits a quadrilateral whose three pairs of opposite vertices $P_1P_2$, $P_3P_6$, $P_5P_9$ are pairs of conjugate points in the polarity $P_i \to p_i$ although the four sides $P_1P_3P_9$, $P_2P_6P_9$, $P_2P_3P_5$, $P_1P_5P_6$ contain their respective poles $P_0$, $P_7$, $P_8$, $P_{11}$. (The remaining three of the thirteen points in this finite plane are the diagonal points of the quadrangle $P_0P_7P_8P_{11}$; their joins in pairs are the diagonals of the quadrilateral $p_0p_7p_8p_{11}$.) See also W. G. Brown, *Canadian Mathematical Bulletin*, **3** (1960), pp. 221–223.

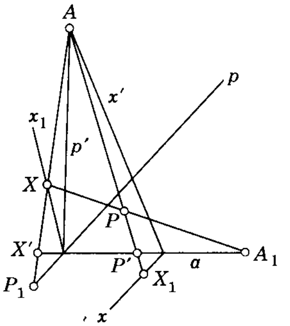



**Figure 14.6b**

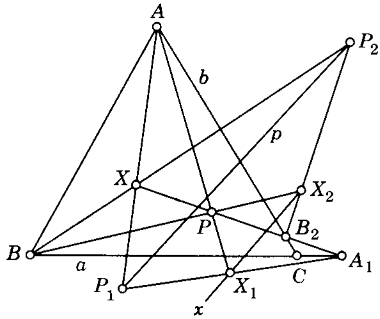



**Figure 14.6c**

$$A_1 = a \cdot PX, \quad P_1 = p \cdot AX, \quad X_1 = AP \cdot A_1P_1,$$
$$B_2 = b \cdot PX, \quad P_2 = p \cdot BX, \quad X_2 = BP \cdot B_2P_2.$$

*Proof.* By 14.64, the polars $a$, $p$, $x$ meet the lines $PX$, $AX$, $AP$ in three collinear points, the first two of which are $A_1$ and $P_1$. Hence $x$ passes through $X_1 = AP \cdot A_1P_1$. Similarly $x$ passes through $X_2 = BP \cdot B_2P_2$.

In terms of coordinates, a projective collineation is a linear homogeneous transformation

**14.66** $$\mu x'_\alpha = \Sigma c_{\alpha\beta} x_\beta, \qquad \det(c_{\alpha\beta}) \neq 0,$$

where the summation is understood to be taken over the repeated index $\beta$ (for each value of $\alpha$). The nonvanishing of the determinant makes it possible to solve the equations for $x_\beta$ in terms of $x'_\alpha$ so as to obtain the inverse collineation. By suitably adjusting the coefficients $c_{\alpha\beta}$, we can transform the particular quadrangle 14.23 into any given quadrangle [Coxeter **2,** p. 197].

Since the product of two correlations (e.g., a polarity and another correlation) is a collineation, any given projective correlation can be exhibited as the product of an arbitrary polarity and a suitable projective collineation. The most convenient polarity for this purpose is that in which the line

$$\Sigma X_\alpha x_\alpha = 0$$

is the polar of the point $(X_1, X_2, X_3)$. Combining this with the general collineation 14.66, we obtain the correlation that transforms each point $(y)$ into the line

**14.661** $$\Sigma(\Sigma c_{\alpha\beta} y_\beta) x_\alpha = 0,$$

where again we must have $\det(c_{\alpha\beta}) \neq 0$. In fact, the correlation is associated with the *bilinear* equation

$$\Sigma\Sigma c_{\alpha\beta} x_\alpha y_\beta = 0$$

[cf. Coxeter **2,** p. 200].

The correlation is a polarity if it is the same as its *inverse*, whose equation, derived by interchanging $(x)$ and $(y)$, is

$$\Sigma\Sigma c_{\alpha\beta} y_\alpha x_\beta = 0, \quad \text{or} \quad \Sigma\Sigma c_{\beta\alpha} x_\alpha y_\beta = 0.$$

Thus a polarity occurs when $c_{\beta\alpha} = \lambda c_{\alpha\beta}$, where $\lambda$ is the same for all $\alpha$ and $\beta$, so that $c_{\alpha\beta} = \lambda c_{\beta\alpha} = \lambda^2 c_{\alpha\beta}, \lambda^2 = 1, \lambda = \pm 1$. But we cannot have $\lambda = -1$, as this would make the determinant

$$\begin{vmatrix} 0 & c_{12} & -c_{31} \\ -c_{12} & 0 & c_{23} \\ c_{31} & -c_{23} & 0 \end{vmatrix} = 0.$$

Hence $\lambda = 1$, and $c_{\beta\alpha} = c_{\alpha\beta}$. In other words,

**14.67** *A projective correlation is a polarity if and only if its matrix of coefficients is symmetric.*

Thus the general polarity is given by

**14.68** $\Sigma\Sigma c_{\alpha\beta} x_\alpha y_\beta = 0, \qquad c_{\beta\alpha} = c_{\alpha\beta}, \quad \det(c_{\alpha\beta}) \neq 0,$

meaning that the polar of $(y_1, y_2, y_3)$ is 14.661, or that 14.68 is the condition for points $(x)$ and $(y)$ to be conjugate. Setting $y_\beta = x_\beta$, we deduce the condition

$$\Sigma\Sigma c_{\alpha\beta} x_\alpha x_\beta = 0,$$

or $c_{11}x_1^2 + c_{22}x_2^2 + c_{33}x_3^2 + 2c_{23}x_2x_3 + 2c_{31}x_3x_1 + 2c_{12}x_1x_2 = 0$, for the point $(x)$ to be self-conjugate. Hence

**14.69** *If a polarity admits self-conjugate points, their locus is given by an equation of the second degree.*

## EXERCISES

**1.** Given the center and axis of a perspective collineation, and one pair of corresponding points (collinear with the center), set up a construction for the transform $X'$ of any point $X$ [Coxeter **2,** p. 62].

**2.** Any two perspective triangles are related by a perspective collineation.

**3.** A collineation which leaves just the points of one line invariant is an elation.

**4.** An elation with axis $o$ may be expressed as the product of two harmonic homologies having this same axis $o$ [Coxeter **2,** p. 63].

**5.** In $PG(2, 3)$, the transformation $P_i \to P_{i+1}$ (with subscripts reduced modulo 13) is evidently a collineation of period 13. Is it a projective collineation? Consider also the transformation $P_i \to P_{3i}$.

**6.** What kind of collineation is

(i) $x'_1 = x_1, \quad x'_2 = x_2, \quad x'_3 = cx_3$;

(ii) $x'_1 = x_1 + c_1x_3, \quad x'_2 = x_2 + c_2x_3, \quad x'_3 = x_3$?

**7.** Use 14.64 to prove *Hesse's theorem*: If two pairs of opposite vertices of a complete quadrilateral are pairs of conjugate points (in a given polarity), then the third pair of opposite vertices is likewise a pair of conjugate points.

**8.** Give an analytic proof of Hesse's theorem. (*Hint*: Apply the condition 14.68 to the pairs of vertices

$$(0, 1, \pm 1), \quad (\pm 1, 0, 1), \quad (1, \pm 1, 0)$$

of the quadrilateral $x_1 \pm x_2 \pm x_3 = 0$.)

**9.** The bilinear equation

$$x_1y_1 + x_2y_2 + x_3y_3 = 0$$

is the condition for $(x)$ and $(y)$ to be conjugate in the polarity $(ABC)(Pp)$, where $ABC$ is the triangle of reference, $P$ is $(1, 1, 1)$, and $p$ is $x_1 + x_2 + x_3 = 0$. Are there any self-conjugate points? Consider, in particular, the case when the coordinates are residues modulo 3.

## 14.7 THE CONIC

*The three familiar curves which we call the "conic sections" have a long history. The reputed discoverer was Menaechmus, who flourished about 350 B.C. They attracted the attention of the best of the Greek geometers down to the time of Pappus of Alexandria. . . . A vivid new interest arose in the seventeenth century. . . . It seems certain that they will always hold a place in the mathematical curriculum.*

J. L. Coolidge (1873-1954)
[Coolidge **1**, Preface]

In the projective plane there is only one kind of conic. The familiar distinction between the hyperbola, parabola, and ellipse belongs to affine geometry. To be precise, it depends on whether the line at infinity is a secant, a tangent, or a nonsecant [Coxeter **2**, p. 129].

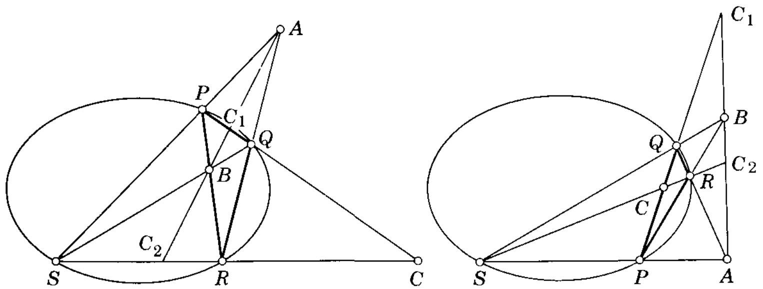



**Figure 14.7a**

A polarity is said to be *hyperbolic* or *elliptic* according as it does or does not admit a self-conjugate point. (In the former case it also admits a self-conjugate line: the polar of the point.) The self-conjugate point $P$, whose existence suffices to make a polarity hyperbolic, is by no means the only self-conjugate point: there is another on every line through $P$ except its polar $p$. To prove this we use 14.63, which tells us that every such line contains an involution of conjugate points. By 14.57, this involution, having one invariant point $P$, has a second invariant point $Q$, which is, of course, another self-conjugate point of the polarity. Thus the presence of one self-conjugate point implies the presence of many (as many as the lines through a point; for example, infinitely many in real or complex geometry). Their locus is a *conic*, and their polars are its *tangents*. This simple definition, due to von Staudt, exhibits the conic as a self-dual figure: the locus of self-conjugate points and also the envelope of self-conjugate lines.

The reader must bear in mind that there are only two kinds of polarity and that there is only one kind of conic. The terminology is perhaps not very well chosen: a hyperbolic polarity has many self-conjugate points, forming a conic; an elliptic polarity has no self-conjugate points at all, but still provides a polar for each point and a pole for each line; there is no such thing as a "parabolic polarity."

A tangent justifies its name by meeting the conic only at its pole, the *point of contact*. Any other line is called a *secant* or a *nonsecant* according as it meets the conic twice or not at all, that is, according as the involution of conjugate points on it is hyperbolic or elliptic. Any two conjugate points on a secant $PQ$, being paired in the involution $(PP)(QQ)$, are harmonic conjugates with respect to $P$ and $Q$.

Let $PQR$ be a triangle inscribed in a conic, as in Figure 14.7*a*. Any line $c$ conjugate to $PQ$ is the polar of some point $C$ on $PQ$. Let $RC$ meet the conic again in $S$. Then $C$ is one of the three diagonal points of the inscribed quadrangle $PQRS$. The other two are

$$A = PS \cdot QR, \quad B = QS \cdot RP.$$

Their join meets the sides $PQ$ and $RS$ in points $C_1$ and $C_2$ such that $\mathsf{H}(PQ, CC_1)$ and $\mathsf{H}(RS, CC_2)$. Since $C_1$ and $C_2$ are conjugate to $C$, the line $AB$, which contains them, is $c$, the polar of $C$. Similarly $BC$ is the polar of $A$. Therefore $A$ and $B$ are conjugate points. These conjugate points are the intersections of $c$ with the sides $QR$ and $RP$ of the given triangle. Hence

**SEYDEWITZ'S THEOREM.** *If a triangle is inscribed in a conic, any line conjugate to one side meets the other two sides in conjugate points.*

From this we shall have no difficulty in deducing

**STEINER'S THEOREM.** *Let lines x and y join a variable point on a conic to two fixed points on the same conic; then* $x \barwedge y$.

*Proof.* The tangents $p$ and $q$, at the fixed points $P$ and $Q$, intersect in $D$, the pole of $PQ$. Let $c$ be a fixed line through $D$ (but not through $P$ or $Q$), meeting $x$ and $y$ in $B$ and $A$, as in Figure 14.7*b*. By Seydewitz's theorem,

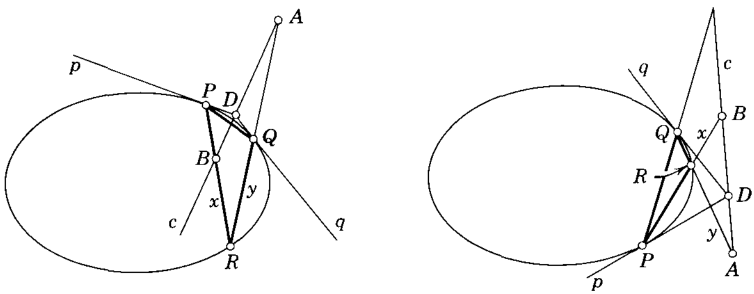



**Figure 14.7b**

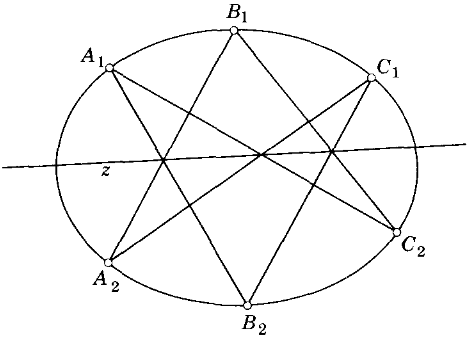



**Figure 14.7c**

$BA$ is a pair of the involution of conjugate points on $c$. Hence, when the point $x \cdot y$ varies on the conic,

$$x \barwedge B \barwedge A \barwedge y.$$

The following construction for a conic through five given points, no three collinear, was discovered by Braikenridge and Maclaurin independently, about 1733 [Coxeter **2**, p. 91]. Let $A_1$, $B_2$, $C_1$, $A_2$, $B_1$ be the five points, as in Figure 14.7$c$; then the conic is the locus of the point

$$C_2 = A_1(z \cdot C_1A_2) \cdot B_1(z \cdot C_1B_2),$$

where $z$ is a variable line through the point $A_1B_2 \cdot B_1A_2$. This is the converse of

**PASCAL'S THEOREM.** *If a hexagon $A_1B_2C_1A_2B_1C_2$ is inscribed in a conic, the points of intersection of pairs of opposite sides, namely,*

$$B_1C_2 \cdot B_2C_1, \quad C_1A_2 \cdot C_2A_1, \quad A_1B_2 \cdot A_2B_1,$$

*are collinear.*

Pascal discovered his famous theorem [Coxeter **2**, p. 103] when he was only sixteen years old. More than 150 years later, it was dualized (see Figure 14.7$d$):

**BRIANCHON'S THEOREM.** *If a hexagon is circumscribed about a conic, its three diagonals are concurrent.*

We saw, in § 8.4, that the familiar conics of Euclidean geometry have equations of the second degree in Cartesian coordinates. The same equations in affine coordinates remain valid in affine geometry, and yield homogeneous equations of the second degree in barycentric coordinates (§ 13.7) and in projective coordinates (§ 14.2). Thus 14.69 serves to reconcile von Staudt's definition of a conic with the classical definitions. In particular,

$$x_1x_3 = x_2^2$$

is a conic touching the lines $x_3 = 0$ and $x_1 = 0$ at the respective points

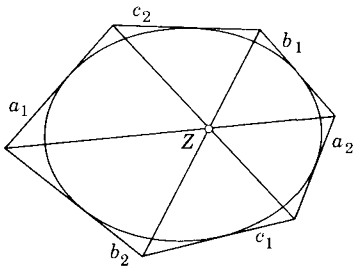



**Figure 14.7d**

(1, 0, 0) and (0, 0, 1). This conic can be parametrized in the form

$$x_1 : x_2 : x_3 = t^2 : t : 1,$$

which exhibits it as the locus of the point of intersection of corresponding members of the projectively related pencils of lines

$$x_1 = tx_2 \quad \text{and} \quad x_2 = tx_3.$$

If det $(c_{\alpha\beta}) = 0$, the quadratic form $\Sigma\Sigma c_{\alpha\beta}x_\alpha x_\beta$ may be expressible as the product of two linear forms $\Sigma a_\alpha x_\alpha$ and $\Sigma b_\beta x_\beta$. Accordingly, a pair of lines is sometimes regarded as a degenerate conic. In this sense, Axiom 14.15 is a special case of Pascal's theorem.

## EXERCISES

**1.** If a quadrangle is inscribed in a conic, its diagonal points form a self-polar triangle. The tangents at the vertices of the quadrangle form a circumscribed quadrilateral whose diagonals are the sides of the same triangle [Coxeter **2,** pp. 85, 86].

**2.** Referring to the projectivity $x \barwedge y$ of Steiner's theorem, investigate the special positions of $x$ and $y$ when $A$ or $B$ coincides with $D$.

**3.** If a projectivity between pencils of lines $x$ and $y$ through $P$ and $Q$ has the effect $xpd \barwedge ydq$, where $d$ is $PQ$, the locus of the point $x \cdot y$ is a conic through $P$ and $Q$ whose tangents at these points are $p$ and $q$. (This construction is often used to *define* a conic; see, e.g., Robinson [**1,** p. 38].)

**4.** Of the conics that touch two given lines at given points, those which meet a third line (not through either of the points) do so in pairs of an involution [Coxeter **2,** p. 90].

**5.** If two triangles are self-polar for a given polarity, their six vertices lie on a conic or on two lines [Coxeter **2,** p. 93].

**6.** If two triangles have six distinct vertices, all lying on a conic, they are self-polar for some polarity [Coxeter **2,** p. 94].

**7.** In $PG(2, 3)$ (Ex. 3 at the end of § 14.1), the polarity $P_i \to p_i$ or $(P_4P_{10}P_{12})(P_0p_0)$ determines a conic consisting of the four points $P_0$, $P_7$, $P_8$, $P_{11}$ and the four lines $p_0$, $p_7, p_8, p_{11}$. (*Hint:* $P_0P_2P_8P_{12} \doublebarwedge P_1P_7P_5P_4 \barwedge p_0p_2p_8p_{12}$.)

**8.** The equation $x_1^2 + x_2^2 - x_3^2 = 0$ represents a conic for which the triangle of reference is self-polar. Verify Pascal's theorem as applied to the inscribed hexagon

$$(0, 1, 1)\ (0, -1, 1)\ (1, 0, 1)\ (-1, 0, 1)\ (3, 4, 5)\ (4, 3, 5).$$

## 14.8 PROJECTIVE SPACE

*Our Geometry is an abstract Geometry. The reasoning could be followed by a disembodied spirit who had no idea of a physical point; just as a man blind from birth could understand the Electromagnetic Theory of Light.*

H. G. Forder [**1**, p. 43]

Axiom 14.12 had the effect of restricting the geometry to a single plane. If we remove this restriction, we must know exactly what we mean by a plane. First we define a *flat pencil* to be the set of lines joining a range of points (on a line) to another point. Then we define a *plane* to be the set of points on the lines of a flat pencil and the set of lines joining pairs of these points. Accordingly *we replace Axiom* 14.12 *by three new axioms.* The first (which may be regarded as a projective version of Pasch's axiom, 12.27) allows us to forget the role of a particular flat pencil in the definition of a plane. The second enables us to speak of more than one plane. The third (cf. 12.431) restricts the number of dimensions to three.

AXIOM **14.81** *If $A$, $B$, $C$, $D$ are four distinct points such that $AB$ meets $CD$, then $AC$ meets $BD$.*

AXIOM **14.82** *There is at least one point not in the plane $ABC$.*

AXIOM **14.83** *Any two planes meet in a line.*

We now have a different principle of duality: points, lines and planes correspond to planes, lines and points (cf. § 10.5). Two intersecting lines, $a$ and $b$, determine a point $a \cdot b$ and a plane $ab$; these are dual concepts. Two lines that do not intersect are said to be *skew*. The theory of collineations and correlations [Coxeter **3**, pp. 63–70] is analogous to the two-dimensional case, except that the number of self-conjugate points on a line is no longer restricted to 0, 1, or 2. In fact, instead of two kinds of polarity we now have four: one "elliptic," having no self-conjugate points, two "hyperbolic," whose self-conjugate points form a quadric (nonruled or ruled), and one, the *null polarity* (or "null system"), in which every point in space is self-conjugate!

The idea of defining a quadric as the locus of self-conjugate points in a three-dimensional polarity (of the second or third kind) is due to von Staudt. Another approach, using a two-dimensional polarity in an arbitrary plane $\omega$, was devised by F. Seydewitz.* The quadric appears as the locus of the point

$$PA \cdot Qa,$$

where $P$ and $Q$ are fixed points (on the quadric) while $A$ is a variable point on $\omega$ and $a$ is its polar. This definition allows the quadric to degenerate to a cone or a pair of planes.

To sample the flavor of solid projective geometry, let us consider a few

* *Archiv für Mathematik und Physik,* **9** (1848), p. 158.

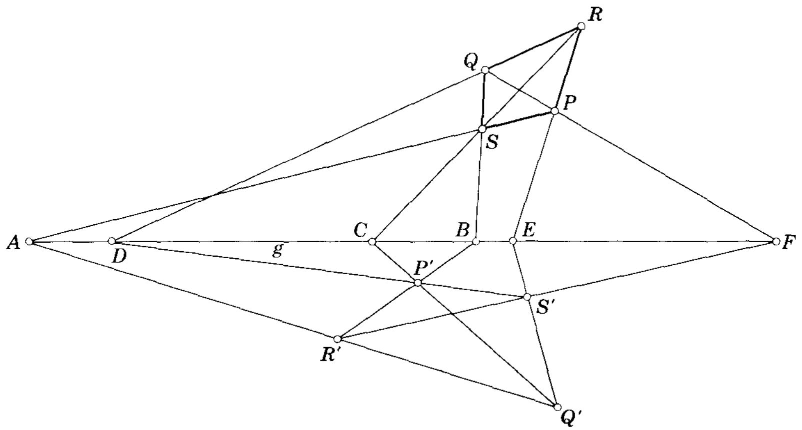



**Figure 14.8a**

typical theorems. Suppose a complete quadrangle $PQRS$ yields a quadrangular set $(AD)(BE)(CF)$ on a line $g$, as in Figure 14.8*a*. In another plane through $g$, let the sides of a triangle $P'Q'R'$ pass through $A$, $B$, $C$, and let $DP'$ meet $EQ'$ in $S'$. Theorem 14.42 tells us that $S'$ lies on $R'F$. This remark yields two interesting configurations: one consisting of eight lines (Figure 14.8*b*), and the other of two mutually inscribed tetrahedra.

**GALLUCCI'S THEOREM.** *If three skew lines all meet three other skew lines, any transversal to the first set of three meets any transversal to the second set.*

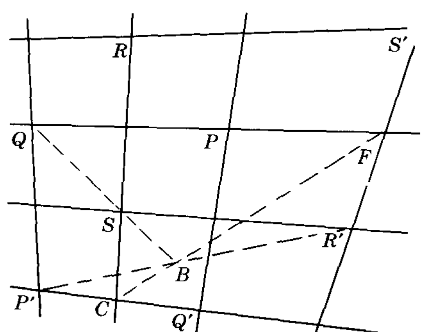
*Proof.* Let the two sets of lines be $PQ'$, $P'Q$, $RS$; $PQ$, $P'Q'$, $R'S$. This notation agrees with Figure 14.8*a*, for, since $PS$ and $Q'R'$ both pass through $A$, $PQ'$ meets $R'S$, and since $QS$ and $R'P'$ both pass through $B$, $P'Q$ meets $R'S$. The transversal from $R$ to $PQ'$ and $P'Q$ is

$$RPQ' \cdot RP'Q = REQ' \cdot RDP' = RS'.$$

The transversal from $R'$ to $PQ$ and $P'Q'$ is

$$R'PQ \cdot R'P'Q' = R'FQ \cdot R'FQ' = R'F.$$

Since $S'$ lies on $R'F$, these transversals meet, as desired.



**MÖBIUS'S THEOREM.** *If the four vertices of one tetrahedron lie respectively in the four face planes of another, while three vertices of the second lie in three face planes of the first, then the remaining vertex of the second lies in the remaining face plane of the first.*

*Proof.* Let $PQRS'$ and $P'Q'R'S$ be the two tetrahedra, with

$$P, \quad Q, \quad R, \quad S', \quad P', \quad Q', \quad S$$

in the respective planes

$$Q'R'S,\ P'R'S,\ P'Q'S,\ P'Q'R',\ QRS',\ PRS',\ PQR,$$

as in Figure 14.8*a*. Since $R'S'$ passes through $F$, on $PQ$, the remaining vertex $R'$ lies in the remaining plane $PQS'$, as desired.

Changing the notation from

$$S, \quad P, \quad Q, \quad R, \quad P', \quad Q', \quad R', \quad S'$$

to

$$S, \quad S_{14}, \quad S_{24}, \quad S_{34}, \quad S_{23}, \quad S_{13}, \quad S_{12}, \quad S_{1234},$$

we deduce the first of a remarkable "chain" of theorems due to Homersham Cox:*

**COX'S FIRST THEOREM.** *Let $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ be four planes of general position through a point $S$. Let $S_{ij}$ be an arbitrary point on the line $\sigma_i \cdot \sigma_j$. Let $\sigma_{ijk}$ denote the plane $S_{ij}S_{ik}S_{jk}$. Then the four planes $\sigma_{234}, \sigma_{134}, \sigma_{124}, \sigma_{123}$ all pass through one point $S_{1234}$.*

Clearly, $\sigma_1$, $\sigma_2$, $\sigma_3$, $\sigma_{123}$ are the face planes of the tetrahedron $P'Q'R'S$, while $\sigma_{234}$, $\sigma_{134}$, $\sigma_{124}$, $\sigma_4$ are those of the inscribed-circumscribed tetrahedron $PQRS'$. Let $\sigma_5$ be a fifth plane through $S$. Then $S_{15}$, $S_{25}$, $S_{35}$, $S_{45}$ are four points in $\sigma_5$; $\sigma_{ij5}$ is a plane through the line $S_{i5}S_{j5}$; and $S_{ijk5}$ is the point $\sigma_{ij5} \cdot \sigma_{ik5} \cdot \sigma_{jk5}$. By the dual of Cox's first theorem, the four points $S_{2345}$, $S_{1345}$, $S_{1245}$, $S_{1235}$ all lie in one plane. Interchanging the roles of $\sigma_4$ and $\sigma_5$, we see that $S_{1234}$ lies in this same plane $S_{2345}S_{1345}S_{1245}$, which we naturally call $\sigma_{12345}$. Hence

**COX'S SECOND THEOREM.** *Let $\sigma_1, \ldots, \sigma_5$ be five planes of general position through $S$. Then the five points $S_{2345}$, $S_{1345}$, $S_{1245}$, $S_{1235}$, $S_{1234}$ all lie in one plane $\sigma_{12345}$.*

Adding the extra digits 56 to all the subscripts in the first theorem, we deduce

**COX'S THIRD THEOREM.** *The six planes $\sigma_{23456}$, $\sigma_{13456}$, $\sigma_{12456}$, $\sigma_{12356}$, $\sigma_{12346}$, $\sigma_{12345}$ all pass through one point $S_{123456}$.*

The pattern is now clear: we can continue indefinitely. "Cox's $(d-3)$rd

* *Quarterly Journal of Mathematics,* **25** (1891), p. 67. See also H. W. Richmond. *Journal of the London Mathematical Society,* **16** (1941), pp. 105–112, and Coxeter, *Bulletin of the American Mathematical Society,* **56** (1950), p. 446. When we describe four planes through a point as being "of general position," we mean that their six lines of intersection are all distinct.

theorem" provides a configuration of $2^{d-1}$ points and $2^{d-1}$ planes, with $d$ of the planes through each point and $d$ of the points in each plane.

Our next result would be difficult to obtain without using coordinates. Since the equation of the general quadric

$$c_{11}x_1^2 + \dots + c_{44}x_4^2 + 2c_{12}x_1x_2 + \dots + 2c_{34}x_3x_4 = 0$$

has $4 + 6 = 10$ terms, a unique quadric $\Sigma = 0$ can be drawn through nine points of general position; for, by substituting each of the nine given sets of $x$'s in $\Sigma = 0$, we obtain nine linear equations to solve for the mutual ratios of the ten $c$'s. Similarly, a "pencil" (or singly infinite system) of quadrics

$$\Sigma + \mu\Sigma' = 0$$

can be drawn through eight points of general position, and a "bundle" (or doubly infinite system) of quadrics

$$\Sigma + \mu\Sigma' + \nu\Sigma'' = 0$$

can be drawn through seven points of general position. But, by solving the simultaneous quadratic equations

$$\Sigma = 0, \quad \Sigma' = 0, \quad \Sigma'' = 0$$

for the mutual ratios of the four $x$'s, we obtain eight points of intersection for these three quadrics. Naturally these eight points lie on every quadric of the bundle. Hence

*Seven points of general position determine a unique eighth point, such that every quadric through the seven passes also through the eighth.*

This idea of the eighth *associated* point provides an alternative proof for Cox's first theorem (and therefore also for the theorems of Möbius and Gallucci). Let $S_{1234}$ be defined as the common point of the three planes $\sigma_{234}$, $\sigma_{134}$, $\sigma_{124}$. (The theorem states that $S_{1234}$ lies also on $\sigma_{123}$.) Since the plane pairs $\sigma_1\sigma_{234}$, $\sigma_2\sigma_{134}$, $\sigma_3\sigma_{124}$ form three degenerate quadrics through the eight points

$$S, S_{14}, S_{24}, S_{34}, S_{23}, S_{13}, S_{12}, S_{1234},$$

these are eight associated points. The first seven belong also to the plane pair $\sigma_4\sigma_{123}$. Since $S_{1234}$ does not lie in $\sigma_4$, it must lie in $\sigma_{123}$, as desired.

The locus of lines meeting three given skew lines is called a *regulus*. Gallucci's theorem shows that the lines meeting three generators of the regulus (including the original three lines) form another "associated" regulus, such that every generator of either regulus meets every generator of the other. The two reguli are the two systems of generators of a *ruled quadric*.

Let $a_1$, $b_1$, $c_1$, $d_1$ be four generators of the first regulus, and $a_2$, $b_2$, $c_2$, $d_2$ four generators of the second, as in Figure 14.8*c*. The three lines

$$a_3 = b_1c_2 \cdot b_2c_1, \quad b_3 = c_1a_2 \cdot c_2a_1, \quad c_3 = a_1b_2 \cdot a_2b_1$$

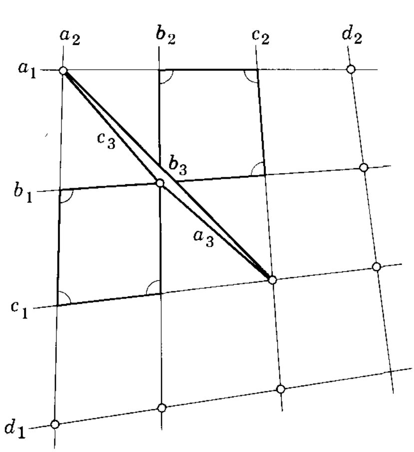



**Figure 14.8c**

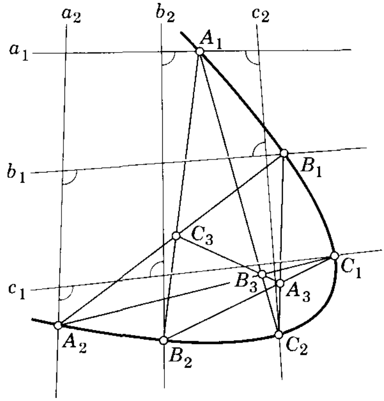



**Figure 14.8d**

evidently form a triangle whose vertices are $a_1 \cdot a_2$, $b_1 \cdot b_2$, $c_1 \cdot c_2$. G. P. Dandelin, in 1824, coined the name *hexagramme mystique* for the skew hexagon $a_1b_2c_1a_2b_1c_2$. Taking the section of its sides by a plane $\delta$ of general position, he obtained a plane hexagon $A_1B_2C_1A_2B_1C_2$ whose sides $A_1B_2$, $B_2C_1$, . . . lie in the planes $a_1b_2$, $b_2c_1$, . . . (Figure 14.8*d*). The points of intersection of pairs of opposite sides, namely,

$$A_3 = B_1C_2 \cdot B_2C_1, \quad B_3 = C_1A_2 \cdot C_2A_1, \quad C_3 = A_1B_2 \cdot A_2B_1,$$

each lying in both the planes $a_3b_3c_3$ and $\delta$, are collinear. By allowing $c_2$ to vary while the remaining sides of the skew hexagon remain fixed, we see from the Braikenridge-Maclaurin construction (which is the converse of Pascal's theorem, Figure 14.7*c*), that

*The section of a ruled quadric, by a plane of general position, is a conic.*

If $\delta$, instead of being a plane of general position, is the plane $d_1d_2$, the vertices of the hexagon $A_1B_2C_1A_2B_1C_2$ line alternately on $d_2$ and $d_1$, as in Axiom 14.15. Thus Pappus's theorem may be regarded as a "degenerate" case of Pascal's theorem. In fact, instead of assuming Pappus's theorem and deducing Gallucci's theorem, we could have taken the latter as an axiom and deduced the former. Bachmann [**1**, p. 254] gives a particularly fine figure to illustrate this deduction.

## EXERCISES

**1.** If $a$ and $b$ are two skew lines and $R$ is a point not on either of them, $Ra \cdot Rb$ is the *only* transversal from $R$ to the two lines.

**2.** Any plane through a generator of a ruled quadric contains another generator. (Such a plane is a *tangent* plane.) Any other plane section of the ruled quadric is a conic.

**3.** If two tetrahedra are trebly perspective they are quadruply perspective (cf. § 14.3). More precisely, if $A_1A_2A_3A_4$ is perspective with each of $B_2B_1B_4B_3$, $B_3B_4B_1B_2$,

$B_4B_3B_2B_1$, it is also perspective with $B_1B_2B_3B_4$. (*Hint*: Since $A_iB_j$ meets $A_jB_i$, $A_iB_i$ must meet $A_jB_j$.)

**4.** The four centers of perspective that were implied in Ex. 3 form a third tetrahedron which is perspective with either of the first two from each vertex of the remaining one.

**5.** In the finite space $PG(3, 3)$, which has 4 points on each line, there are altogether 40 points, 40 planes, and how many lines?

## 14.9 EUCLIDEAN SPACE

*The set of lines drawn from the artist's eye to the various points of the object . . . constitute the projection of the object and are called the Euclidean cone. Then the section of this cone made by the canvas is the desired drawing. . . . Parallel lines in the object converge in the picture to the point where the canvas is pierced by the line from the eye parallel to the given lines.*

S. H. Gould [**1**, p. 299]

The elementary approach to affine space is to regard it as Euclidean space without a metric; the elementary approach to projective space is to regard it as affine space plus the plane at infinity and then to ignore the special role of that plane. It is equally effective to begin with projective space and derive affine space by specializing any one plane, calling it the plane at infinity. (This is still, of course, a projective plane.) Each affine concept has its projective definition: for example, the midpoint of $AB$ is the harmonic conjugate, with respect to $A$ and $B$, of the point at infinity on $AB$ [Coxeter **2**, p. 119]. We then derive Euclidean space by specializing one elliptic polarity in the plane at infinity, calling it the *absolute polarity*. Two lines are orthogonal if their points at infinity are conjugate in the absolute polarity; a line and a plane are orthogonal if the point at infinity on the line is the pole of the line at infinity in the plane. A sphere is the locus of the point of intersection of a line through one fixed point and the perpendicular plane through another; thus it is a special quadric according to Seydewitz's definition. Two segments with a common end are congruent if they are radii of the same sphere [Coxeter **2**, p. 146].

When we use projective coordinates $(x_1, x_2, x_3, x_4)$, referred to an arbitrary tetrahedron

$$(1, 0, 0, 0)\ (0, 1, 0, 0)\ (0, 0, 1, 0)\ (0, 0, 0, 1),$$

it is convenient to take the plane at infinity to be $x_4 = 0$. Any other equation becomes an equation in affine coordinates $x_1, x_2, x_3$ by the simple device of setting $x_4 = 1$. In affine terms, the tetrahedron of reference for the projective coordinates is formed by the origin and the points at infinity on the three axes. Finally, we pass from affine space to Euclidean space by declaring that two points $(x_1, x_2, x_3)$ and $(y_1, y_2, y_3)$ are in perpendicular directions from the origin if they satisfy the bilinear equation

$$x_1y_1 + x_2y_2 + x_3y_3 = 0,$$

that is, if the points at infinity

$$(x_1, x_2, x_3, 0) \quad \text{and} \quad (y_1, y_2, y_3, 0)$$

are conjugate in the absolute polarity.

All the theorems that we proved in § 14.8 remain valid in Euclidean space. An interesting variant of Cox's chain of theorems can be obtained by means of the following specialization. Instead of an *arbitrary* point on the line $\sigma_i \cdot \sigma_j$, we take $S_{ij}$ to be the second intersection of this line with a fixed sphere through $S$. Since the sphere is a quadric through the first seven of the eight associated points

$$S,\ S_{14},\ S_{24},\ S_{34},\ S_{23},\ S_{13},\ S_{12},\ S_{1234},$$

it passes through $S_{1234}$ too, and similarly through $S_{1235}$ and the rest of the $2^{d-1}$ points. The $2^{d-1}$ planes meet the sphere in $2^{d-1}$ circles, which remain circles when we make an arbitrary stereographic projection, as in § 6.9. We thus obtain Clifford's chain of theorems* in the inversive (or Euclidean) plane.

**CLIFFORD'S FIRST THEOREM.** *Let $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ be four circles of general position through a point $S$. Let $S_{ij}$ be the second intersection of the circles $\sigma_i$ and $\sigma_j$. Let $\sigma_{ijk}$ denote the circle $S_{ij}S_{ik}S_{jk}$. Then the four circles $\sigma_{234}, \sigma_{134}, \sigma_{124}, \sigma_{123}$ all pass through one point $S_{1234}$.*

**CLIFFORD'S SECOND THEOREM.** *Let $\sigma_5$ be a fifth circle through $S$. Then the five points $S_{2345}, S_{1345}, S_{1245}, S_{1235}, S_{1234}$ all lie on one circle $\sigma_{12345}$.*

**CLIFFORD'S THIRD THEOREM.** *The six circles $\sigma_{23456}, \sigma_{13456}, \sigma_{12456}, \sigma_{12356}, \sigma_{12346}, \sigma_{12345}$ all pass through one point $S_{123456}$.*

And so on!

## EXERCISES

1. Why is the absolute polarity elliptic?
2. Draw a careful figure for Clifford's first theorem.
3. The circumcircles of the four triangles formed by four general lines all pass through one point (cf. Ex. 2 at the end of § 5.5).
4. The circumcenters of the four triangles of Ex. 3 all lie on a circle which passes also through the point of concurrence of the four circumcircles [Forder **3**, pp. 16–22; Baker **1**, p. 328].

* W. K. Clifford, *Mathematical Papers* (London, 1882), p. 51. Apparently Clifford did not state these theorems in their full generality. Instead of circles through $S$ he took $\sigma_1, \sigma_2, \ldots$ to be straight lines. In other words, he took $S$ to be the point at infinity of the inversive plane. Thus his special form of the theorems could have been derived from the configuration of circles on the sphere by taking the center of the stereographic projection to be the point $S$ on the sphere [Baker **1**, p. 133].

# 15

# Absolute geometry

In the present chapter we shall re-examine the material of some of the earlier chapters in the light of the axiomatic approach outlined in Chapter 12, regarding classical geometry as ordered geometry enriched with the axioms of congruence 15.11–15.15, the last of which is a restatement of 1.26. Except in §§ 15.6 and 15.8, we shall work in the domain of *absolute* geometry, that is, we shall take care not to assume any form of Euclid's fifth postulate. Accordingly, our results will be valid not only in Euclidean geometry but also in the non-Euclidean geometry of Gauss, Lobachevsky, and Bolyai.

In § 15.4 we shall give a simple account of the complete enumeration of finite groups of isometries. According to Weyl [**1**, p. 79], "This is the modern equivalent to the tabulation of the regular polyhedra by the Greeks." The relevance of these kinematical results to crystallography makes it natural, in § 15.6, to reintroduce the full machinery of Euclidean geometry. But in § 15.7 we shall return to absolute geometry for a discussion of finite groups generated by reflections. Many of the methods used remain valid also in spherical geometry.

## 15.1 CONGRUENCE

*Every teacher certainly should know something of non-Euclidean geometry. . . . It forms one of the few parts of mathematics which . . . is talked about in wide circles, so that any teacher may be asked about it at any moment.*

F. Klein [**2**, p. 135]

To give a rigorous approach to absolute geometry, we begin with ordered geometry (Chapter 12) and introduce *congruence* as a third primitive concept: an undefined equivalence relation among point pairs (or segments, or intervals). We use the notation $AB \equiv CD$ to mean "$AB$ is congruent to $CD$." The following axioms are those of Pasch with some refinements due to Hilbert and R. L. Moore [see Kerékjártó **1**, pp. 90–101].

## Axioms of Congruence

**15.11** *If $A$ and $B$ are distinct points, then on any ray going out from $C$ there is just one point $D$ such that $AB \equiv CD$.*

**15.12** *If $AB \equiv CD$ and $CD \equiv EF$, then $AB \equiv EF$.*

**15.13** $AB \equiv BA$.

**15.14** *If $[ABC]$ and $[A'B'C']$ and $AB \equiv A'B'$ and $BC \equiv B'C'$, then $AC \equiv A'C'$.*

**15.15** *If $ABC$ and $A'B'C'$ are two triangles with $BC \equiv B'C'$, $CA \equiv C'A'$, $AB \equiv A'B'$, while $D$ and $D'$ are two further points such that $[BCD]$ and $[B'C'D']$ and $BD \equiv B'D'$, then $AD \equiv A'D'$.*

By two applications of 15.13, we have $AB \equiv AB$; that is, congruence is *reflexive.* From 15.11 and 15.12 we easily deduce that the relation $AB \equiv CD$ implies $CD \equiv AB$; that is, congruence is *symmetric.* Axiom 15.12 itself says that congruence is *transitive.* Hence congruence is an *equivalence* relation. This result, along with the *additive* property of 15.14, provides the basis for a theory of *length* [Forder **1**, p. 95]. Axiom 15.15 enables us to extend the relation of congruence from point pairs or segments to *angles* [Forder **1**, p. 132].

We follow Euclid in defining a *right angle* to be an angle that is congruent to its supplement; and we agree to measure angles on such a scale that the magnitude of a right angle is $\frac{1}{2}\pi$.

The statement $AB \equiv CD$ for segments is clearly equivalent to the statement $AB = CD$ for lengths, so no confusion arises from using the same symbol $AB$ for a segment and its length. A similar remark applies to angles.

The *circle* with center $O$ and radius $r$ is defined as the locus of a variable point $P$ such that $OP = r$. A point $Q$ such that $OQ > r$ is said to be *outside* the circle. Points neither on nor outside the circle are said to be *inside.* It can be proved [Forder **1**, p. 131] that if a circle with center $A$ has a point inside and a point outside a circle with center $C$, then the two circles meet in just one point on each side of the line $AC$. Euclid's first four postulates may now be treated as theorems, and we can prove all his propositions as far as I.26; also I.27 and 28 with the word "parallel" replaced by "nonintersecting." We can define reflection as in § 1.3, and derive its simple consequences such as *pons asinorum* (Euclid I.5) and the symmetry of a circle about its diameters (III.3; see § 1.5). But we must be careful to avoid any appeal to our usual idea about the sum of the angles of a triangle; for example, we can no longer assert that angles in the same segment of a circle are equal (Euclid III.21). Lacking such theorems as VI.2–4, which depend on the affine properties of parallelism, we have to look for some quite different way to prove the concurrence of the medians of a triangle.* On the other hand, the concurrence of the *altitudes* (of an acute-angled triangle) arises as a by-product of Fagnano's problem, which can still be treated as in § 1.8. (Fermat's problem would require a different treatment because we can no longer assume the angles of an equilateral triangle to be $\pi/3$.)

* Bachmann **1**, pp. 74–75.

**EXERCISES**

1. Complete the proof that congruence is symmetric: if $AB \equiv CD$ then $CD \equiv AB$.

2. How much of § 1.5 remains valid in absolute geometry? [Kerékjártó **1**, pp. 161–163.] (See especially Exercises 1 and 3.)

3. For any simple quadrangle inscribed in a circle, the sum of two opposite angles is equal to the sum of the remaining two angles [Sommerville **1**, p. 84].

## 15.2 PARALLELISM

*I have resolved to publish a work on the theory of parallels as soon as I have put the material in order. . . . The goal is not yet reached, but I have made such wonderful discoveries that I have been almost overwhelmed by them. . . . I have created a new universe from nothing.*

Jonos Bolyai (1802-1860)
(From a letter to his father in 1823)

Following Gauss, Bolyai, and Lobachevsky, we say that two lines are *parallel* if they "almost meet." For the precise meaning of this phrase, see § 12.6. (We use the notation $p_1$ for one of the two rays into which the line $p$ is decomposed by a point that lies on it.)

The idea of the *incenter* (§ 1.5) may be extended from a triangle to the figure formed by two parallel lines and a transversal, enabling us to prove that parallelism is symmetric:

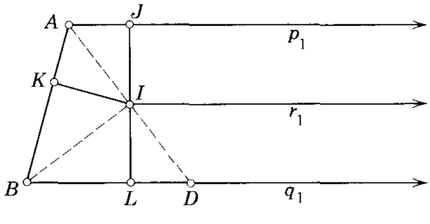



**Figure 15.2a**

**15.21** *If $p_1$ is parallel to $q_1$, then $q_1$ is parallel to $p_1$.*

*Proof.* [Sommerville **1**, p. 32]. If $p_1$, through $A$, is parallel to $q_1$, through $B$, as in Figure 15.2$a$, the internal bisector $AD$ of the angle at $A$ completes a triangle $ABD$. Let the internal bisector of $B$ meet $AD$ in $I$. Draw perpendiculars $IJ$, $IK$, $IL$, to $p_1$, $AB$, $q_1$. Reflecting in $IA$ and $IB$, we see that $IJ = IK = IL$. Let $r_1$ be the internal bisector of $\angle LIJ$. Reflection in the line $r$ interchanges $J$ and $L$, and therefore interchanges $p$ and $q$. Since $p$ is parallel to $q$, it follows that $q$ is parallel to $p$ in the same sense, that is, $q_1$ is parallel to $p_1$. (In the terminology of Gauss, $J$ and $L$ are *corresponding* points on the two parallel rays.)

We can now use the methods of ordered geometry to prove that parallelism is transitive:

**15.22** *If $p_1$ is parallel to $q_1$, and $q_1$ is parallel to $r_1$, then $p_1$ is parallel to $r_1$.*

*Proof* [Gauss **1**, vol. 8, pp. 205–206]. We have to show that, if $p_1$ and $r_1$ are both parallel to $q_1$, they are parallel to each other. We see at once that $p_1$ and $r_1$ cannot meet; for if they did, we would have two intersecting lines $p$ and $r$ both parallel to $q$ in the given sense. By Theorem 12.64, we may assume that $p_1$, $q_1$, $r_1$ begin from three collinear points $A$, $B$, $C$. For the rest of the proof we distinguish the case in which $B$ lies between $A$ and $C$ from the case in which it does not.

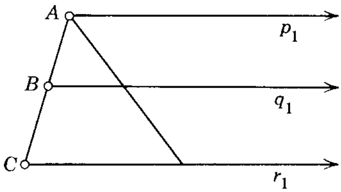



**Figure 15.2b**

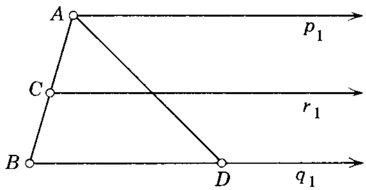



**Figure 15.2c**

If $[ABC]$, as in Figure 15.2*b*, any ray from $A$ within the angle between $AC$ and $p_1$ meets $q_1$ (since $p_1$ is parallel to $q_1$) and then meets $r_1$ (since $q_1$ is parallel to $r_1$). Therefore $p_1$ is parallel to $r_1$.

If $B$ is not between $A$ and $C$, suppose for definiteness that $[ACB]$, as in Figure 15.2*c*. Any ray from $A$ within the angle between $AC$ and $p_1$ meets $q_1$, say in $D$. Since $r$ separates $A$ from $D$, it meets the segment $AD$. Therefore $p_1$ is parallel to $r_1$.

In this second part of the proof we have not used the parallelism of $q_1$ and $r_1$. In fact,

**15.23** *If a ray $r_1$ lies between two parallel rays, it is parallel to both.*

Having proved that parallelism is an equivalence relation, we consider the set of lines parallel to a given ray. We naturally call this a *pencil of parallels*, since it contains a unique line through any given point [Coxeter **2**, p. 5]. Pursuing its analogy with an ordinary pencil (consisting of all the lines through a point), we may also call it a *point at infinity* or, following Hilbert, an *end*. Instead of saying that two rays (or lines) are parallel, or that they belong to a certain pencil of parallels $M$, we say that they have $M$ for a common end. In the same spirit, the ray through $A$ that belongs to the given pencil of parallels is denoted by $AM$, as if it were a segment; the same symbol $AM$ can also be used for the whole line.

Let $AM$, $BM$ be parallel rays, and $\varepsilon$ an arbitrarily small angle. Within the angle $BAM$ (Figure 15.2*d*), take a ray from $A$ making with $AM$ an angle less than $\varepsilon$. This ray cuts $BM$ in some point $C$. On $CM$ (which is $C/B$), take $D$ so that $CD = CA$. The isosceles triangle $CAD$ yields

$$\angle ADC = \angle CAD < \angle CAM < \varepsilon.$$

Hence, when $BD$ tends to infinity, so that $AD$ tends to the position $AM$, $\angle ADB$ tends to zero.

This conclusion motivates the following assertion of Bolyai [**1**, p. 207]:

**15.24** *When two parallel lines are regarded as meeting at infinity, the angle of intersection must be considered as being equal to zero.*

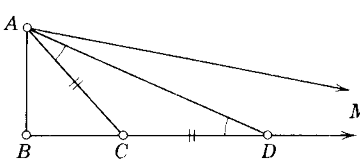


**Figure 15.2d**

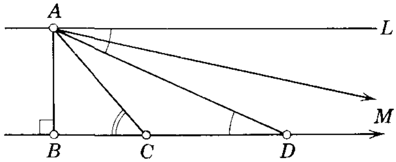


**Figure 15.2e**

When $AM$ and $BM$ are parallel rays, we call the figure $ABM$ an *asymptotic triangle.* Such triangles behave much like finite triangles. In particular, two of them are congruent if they agree in the finite side and one angle [Carslaw **1**, p. 49]:

**15.25** *If two asymptotic triangles $ABM$, $A'B'M'$ have $AB = A'B'$ and $A = A'$, then also $B = B'$.*

It is a consequence of Axiom 15.11 that, if two lines have a common perpendicular, they do not intersect. The following theorem provides a kind of converse for this statement.

**15.26** *If two lines are neither intersecting nor parallel, they have a common perpendicular.*

*Proof.* From $A$ on the first line $AL$, draw $AB$ perpendicular to the second line $BM$, as in Figure 15.2*e*. If $AB$ is perpendicular to $AL$ there is no more to be said. If not, suppose $L$ is on that side of $AB$ for which $\angle BAL$ is acute. Since the two lines are neither intersecting nor parallel, there is a smaller angle $BAM$ such that $AM$ is parallel to $BM$. If $[BCD]$ on $BM$, we can apply Euclid I.16 to the triangle $ACD$, with the conclusion that the internal angle at $D$ is less than the external angle at $C$. Hence, when $BD$ increases from 0 to $\infty$, so that $\angle DAL$ decreases from $\angle BAL$ to $\angle MAL$, $\angle ADB$ decreases from a right angle to zero. At the beginning of this process we have

$$\angle DAL < \angle ADB$$

(since $\angle BAL$ is acute); but at the end the inequality is reversed (since

$\angle MAL$ is positive). Hence there must be some intermediate position for which

$$\angle DAL = \angle ADB.$$

(To be precise, we can apply Dedekind's axiom 12.51 to the points on $BM$ satisfying the two opposite inequalities.) For such a point $D$ (Figure 15.2*f*) we obtain two triangles $OAE$, $ODF$ by drawing $EF$ perpendicular to $BD$ through $O$, the midpoint of $AD$. Since these triangles are congruent, $EF$ is perpendicular not only to $BD$ but also to $AL$.

Nonintersecting lines that are not parallel are said to be *ultraparallel* (or "hyperparallel"). We are not asserting the existence of such lines, but merely showing how they must behave if they do exist.

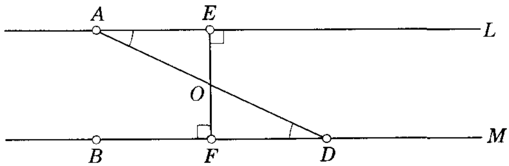



**Figure 15.2f**

**EXERCISES**

1. Prove 15.25 without referring to Carslaw **1**.

2. Give a complete proof that, if two lines have a common perpendicular, they do not intersect.

3. Example 4 on p. 16 remains valid when $A$ is an end so that the triangle is asymptotic.

## 15.3 ISOMETRY

*Beside the actual universe I can set in imagination other universes in which the laws are different.*

J. L. Synge [**2**, p. 21]

The whole theory of finite groups of isometries (§§ 2.3–3.1) belongs to absolute geometry, because it is concerned with isometries having at least one invariant point. The first departure from our previous treatment (§ 3.2) is in the discussion of isometries without invariant points. We must now distinguish between a *translation*, which is the product of half-turns about two distinct points, and a *parallel displacement*, which is the product of reflections in two parallel lines.

The product of half-turns about two distinct points $O$, $O'$ is a translation along a given line (called the *axis* of the translation) in a given sense through a given distance, namely, along $OO'$ in the sense of the ray $O'/O$ through the distance $2OO'$. Since a translation is determined by its axis and directed distance, the product of half-turns about $O$, $O'$ is the same as the product of half-turns about $Q$, $Q'$, provided the directed segment $QQ'$ is congruent to $OO'$ *on the same line* (Figure 3.2*a*). If $P$ is on this line, the distance $PP^{T}$ is just twice $OO'$. (If not, it may be greater!)

By the argument used in proving 3.21, the product of two translations with the same axis, or with intersecting axes, is a translation. (It is only in the former case that we can be sure of commutativity.) More precisely, we have

**15.31** (Donkin's theorem*) *The product of three translations along the directed sides of a triangle, through twice the lengths of these sides, is the identity.*

We shall see later that the product of two translations with nonintersecting axes may be a rotation.

By the argument used in proving 3.22, if two lines have a common perpendicular, the product of reflections in them is a translation along this common perpendicular through twice the distance between them. (Such lines may be either parallel or ultraparallel according to the nature of the geometry.)

Again, as in 3.13, every isometry is the product of at most three reflections. If the isometry is direct, the number of reflections is even, namely 2. It follows from 15.26 that

**15.32** *Every direct isometry* (*of the plane*) *with no invariant point is either a parallel displacement or a translation.*

It is remarkable that absolute geometry includes the whole theory of glide reflection. The only changes needed in the previous treatment (§ 3.3) are where the word "parallel" was used. (In Figure 3.3*b* we must define $m$, $m'$ as being perpendicular to $OO'$; they are not necessarily parallel to each other.) As an immediate application of these ideas we have Hjelmslev's theorem, which is one of the best instances of a genuinely surprising result belonging to absolute geometry. The treatment in § 3.6 remains valid without changing a single word!

Likewise, the one-dimensional groups of § 3.7 belong to absolute geometry, the only change being that again the mirrors $m$, $m'$ (Figure 3.7*b*) should not be said to be "parallel" but both perpendicular to the same (horizontal) line. On the other hand, the whole theory of lattices (Chapter 4) and of similarity (Chapter 5) must be abandoned.

The extension of absolute geometry from two dimensions to three presents no difficulty. In particular, much of the Euclidean theory of isometry (§ 7.1) remains valid in absolute space. It is still true that every direct isometry is the product of two half-turns, and that every opposite isometry with

* W. F. Donkin, On the geometrical theory of rotation, *Philosophical Magazine* (4), **1**, (1851), 187–192. Lamb [**1**, p. 6] used half-turns about the vertices $A$, $B$, $C$ of the given triangle to construct three new triangles which, he said, "are therefore directly equal to one another, and 'symmetrically' equal to $ABC$." This was a mistake: all four triangles are directly congruent!

an invariant point is a rotatory inversion (possibly reducing to a reflection or to a central inversion). Moreover, the classical enumeration of the five Platonic solids (§§ 10.1–10.3) is part of absolute geometry. The few necessary changes are easily supplied; for example, the term *rectangle* must be interpreted as meaning a quadrangle whose angles are all equal (though not necessarily right angles), and a *square* is the special case when also the sides are equal.

### EXERCISES

**1.** If $l$ is a line outside the plane of a triangle $ABC$, what can be said about the three lines in which this plane meets the three planes $Al$, $Bl$, $Cl$? (If two of the three lines intersect, or are parallel, or have a common perpendicular, the same can be said of all three. This property of three lines $m_1$, $m_2$, $m_3$ is equivalent to $R_1R_2R_3 = R_3R_2R_1$ in the notation of § 3.4.)

**2.** The product of reflections in the lines $p$ and $r$ of Figure 15.2$a$ is a parallel displacement which transforms $J$ into $L$.

## 15.4 FINITE GROUPS OF ROTATIONS

*These groups, in particular the last three, are an immensely attractive subject for geometric investigation.*

H. Weyl [**1**, p. 79]

One of the simplest kinds of transformation is a *permutation* (or rearrangement) of a finite number of named objects. For instance, one way to permute the six letters $a, b, c, d, e, f$ is to transpose (or interchange) $a$ and $b$, to change $c$ into $d$, $d$ into $e$, $e$ into $c$, and to leave $f$ unaltered. This permutation is denoted by $(a\ b)(c\ d\ e)$. The two "independent" parts, $(a\ b)$ and $(c\ d\ e)$, are called *cycles* of periods 2 and 3. A permutation that consists of just one cycle is said to be *cyclic*. Clearly, the cyclic group $C_n$ may be represented by the powers of the generating permutation $(a_1a_2 \ldots a_n)$; for instance, the four elements of $C_4$ are

$$1, \quad (a\ b\ c\ d), \quad (a\ c)(b\ d), \quad (a\ d\ c\ b).$$

A cyclic permutation of period 2, such as $(a\ b)$, is called a *transposition*. Since

$$(a_1a_2 \ldots a_n) = (a_1a_n)(a_2a_n)\ldots,$$

any permutation may be expressed as a product of transpositions. A permutation is said to be *even* or *odd* according to the parity of the number of cycles of even period; for instance, $(a\ c)(b\ d)$ is even, but $(a\ b)(c\ d\ e)$ is odd. The identity, 1, has no cycles at all, and is accordingly classified as an even permutation. It is easily proved [see Coxeter **1**, pp. 40–41] that every product of transpositions is even or odd according to the parity of the number of transpositions. It follows that the multiplication of even and odd permutations behaves like the *addition* of even and odd numbers; for example, the product of two odd permutations is even.

It follows also that every group of permutations either consists entirely of even permutations or contains equal numbers of even and odd permutations. The group of all permutations of $n$ objects is called the *symmetric* group of order $n!$ (or of *degree* $n$) and is denoted by $S_n$. The subgroup consisting of all the even permutations is called the *alternating* group of order $\frac{1}{2}n!$ (or of degree $n$) and is denoted by $A_n$. In particular, $S_2$ is the same group as $C_2$, and $A_3$ the same as $C_3$, so we write

$$S_2 \cong C_2, \qquad A_3 \cong C_3.$$

More interestingly, $S_3 \cong D_3$ (see Figure 2.7$a$). For, the six elements of the dihedral group $D_3$, being symmetry operations of an equilateral triangle, may be regarded as permutations of the three sides of the triangle. The even permutations

$$1, \quad (a\ b\ c), \quad (a\ c\ b)$$

(which form the subgroup $A_3 \cong C_3$) are rotations, whereas the odd permutations

$$(b\ c), \quad (c\ a), \quad (a\ b)$$

are reflections in the three medians. If we regard the triangle as lying in three-dimensional (absolute) space, the rotations are about an axis through the center of the triangle, perpendicular to its plane. The reflections may then be interpreted in two alternative ways, yielding two groups which are geometrically distinct but abstractly identical or *isomorphic*: we may either reflect in three planes through the axis or rotate through half-turns about the medians themselves. In the latter representation, all the six elements of $D_3$ appear as rotations. We may describe this as the group of direct symmetry operations of a triangular prism. More generally, the $2n$ direct symmetry operations of an $n$-gonal prism form the dihedral group $D_n$, whereas of course the $n$ direct symmetry operations of an $n$-gonal pyramid form the cyclic group $C_n$. The rotations of $C_n$ all have the same axis, and $D_n$ is derived from $C_n$ by adding half-turns about $n$ lines symmetrically disposed in a plane perpendicular to that axis.

We have thus found two infinite families of finite groups of rotations. Other such groups are the groups of direct symmetry operations of the five Platonic solids $\{p, q\}$. These are only three groups, not five, because any rotation that takes $\{p, q\}$ into itself also takes the reciprocal $\{q, p\}$ into itself: the octahedron has the same group of rotations as the cube, and the icosahedron the same as the dodecahedron.

The regular tetrahedron $\{3, 3\}$ is evidently symmetrical by reflection in the plane that joins any edge to the midpoint of the opposite edge. As a permutation of the four faces $a$, $b$, $c$, $d$ (Figure 15.4$a$), this reflection is just a transposition. Thus the complete symmetry group of the tetrahedron,

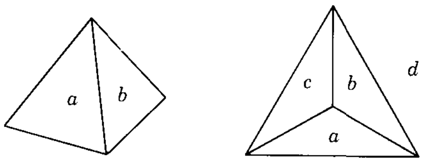



**Figure 15.4a**

being generated by such reflections, is isomorphic to the symmetric group $S_4$, which is generated by transpositions; and the rotation group, being generated by products of pairs of reflections, is isomorphic to the alternating group $A_4$, which is generated by products of pairs of transpositions. The 12 rotations may be counted as follows. The perpendicular from a vertex to the opposite face is the axis of a *trigonal* rotation (i.e., a rotation of period 3); the 4 vertices yield 8 such rotations. The line joining the midpoint of two opposite edges is the axis of a half-turn (or *digonal* rotation); the 3 pairs of opposite edges yield 3 such half-turns. Including the identity, we thus have $8 + 3 + 1 = 12$ rotations. As permutations, the 8 trigonal rotations are

$$(b\ c\ d),\quad (b\ d\ c),\quad (a\ c\ d),\quad (a\ d\ c),\quad (a\ b\ d),\quad (a\ d\ b),\quad (a\ b\ c),\quad (a\ c\ b)$$

and the 3 half-turns are

$$(b\ c)(a\ d),\quad (c\ a)(b\ d),\quad (a\ b)(c\ d).$$

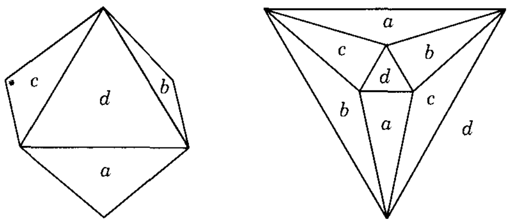



**Figure 15.4b**

The octahedron $\{3, 4\}$ can be derived from the tetrahedron by *truncation*: its eight faces consist of the four vertex figures of the tetrahedron and truncated versions of the four faces. Every symmetry operation of the tetrahedron is retained as a symmetry operation of the octahedron, but the octahedron also has symmetry operations that interchange the two sets of four faces. For instance, the line joining two opposite vertices is the axis of a *tetragonal* rotation (of period 4), and the line joining the midpoints of two opposite edges is the axis of a half-turn. When the four pairs of opposite faces are marked *a*, *b*, *c*, *d*, as in Figure 15.4*b*, such a half-turn appears as a transposition, which is one of the permutations that belong to $S_4$ but not

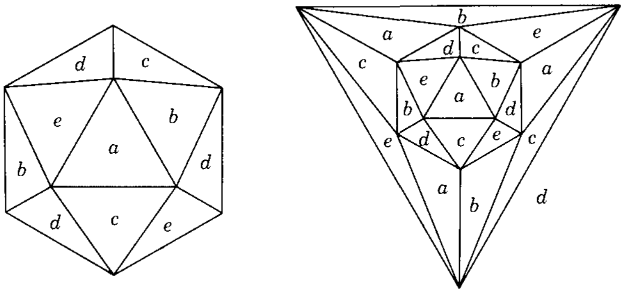



**Figure 15.4c**

to $A_4$. It follows that the rotation group of the octahedron (or of the cube) is isomorphic to the symmetric group $S_4$.

In Figure 15.4*c*, the twenty faces of the icosahedron $\{3, 5\}$ have been marked *a, b, c, d, e* in sets of four, in such a way that two faces marked alike have nothing in common, not even a vertex. In fact, the four *a*'s (for instance) lie in the planes of the faces of a regular tetrahedron, and the respectively opposite faces (marked *b, c, d, e*) form the reciprocal tetrahedron. The twelve rotations of either tetrahedron into itself (represented by the even permutations of *b, c, d, e*) are also symmetry operations of the whole icosahedron. This behavior of the four *a*'s is imitated by the *b*'s, *c*'s, *d*'s and *e*'s, so that altogether we have all the even permutations of the five letters: the rotation group of the icosahedron (or of the dodecahedron) is isomorphic to the alternating group $A_5$. The 60 rotations may be counted as follows: 4 pentagonal rotations about each of 6 axes, 2 trigonal rotations about each of 10 axes, 1 half-turn about each of 15 axes, and the identity [Coxeter **1**, p. 50].

We shall find that the above list exhausts the finite groups of rotations. As a first step in this direction, we observe that all the axes of rotation must pass through a fixed point. In fact, we can just as easily prove a stronger result:

**15.41** *Every finite group of isometries leaves at least one point invariant.*

*Proof.* A finite group of isometries transforms any given point into a finite set of points, and transforms the whole set of points into itself. This, like any finite (or bounded) set of points, determines *a unique smallest sphere that contains all the points on its surface or inside*: unique because, if there were two equal smallest spheres, the points would belong to their common part, which is a "lens"; and the sphere that has the rim of the lens for a great circle is smaller than either of the two equal spheres, contradicting our supposition that these spheres are as small as possible. (The shaded area in Figure 15.4*d* is a section of the lens.) The group transforms this unique sphere into itself. Its surface contains some of the points, and therefore all of them. Its center is the desired invariant point.

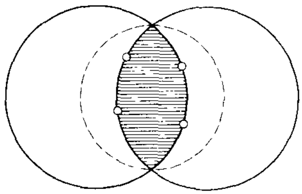


**Figure 15.4d**

It follows that any finite group of rotations may be regarded as operating on the surface of a sphere. In such a group $G$, each rotation, other than the identity, leaves just two points invariant, namely the *poles* where the axis of rotation intersects the sphere. A pole $P$ is said to be $p$-gonal ($p \geqslant 2$) if it belongs to a rotation of period $p$. The $p$ rotations about $P$, through various multiples of the angle $2\pi/p$, are those rotations of $G$ which leave $P$ invariant. Any other rotation of $G$ transforms $P$ into an "equivalent" pole, which is likewise $p$-gonal. Thus all the poles fall into sets of equivalent poles. All the poles in a set have the same period $p$, but two poles of the same period do not necessarily belong to the same set; they belong to the same set only if one is transformed into the other by a rotation that belongs to $G$.

Any set of equivalent $p$-gonal poles consists of exactly $n/p$ poles, where $n$ is the order of $G$. To prove this, take a point $Q$ on the sphere, arbitrarily near to a pole $P$ belonging to the set. The $p$ rotations about $P$ transform $Q$ into a small $p$-gon round $P$. The other rotations of $G$ transform this $p$-gon into congruent $p$-gons round all the other poles in the set. But the $n$ rotations of $G$ transform $Q$ into just $n$ points (including $Q$ itself). Since these $n$ points are distributed into $p$-gons round the poles, the number of poles in the set must be $n/p$.

The $n - 1$ rotations of $G$, other than the identity, consist of $p - 1$ for each $p$-gonal axis, that is, $\frac{1}{2}(p - 1)$ for each $p$-gonal pole, or

$$\tfrac{1}{2}(p - 1)n/p$$

for each set of $n/p$ equivalent poles. Hence

$$n - 1 = \tfrac{1}{2}n \, \Sigma (p - 1)/p,$$

where the summation is over the sets of poles. This equation may be expressed as

$$2 - \frac{2}{n} = \Sigma \left(1 - \frac{1}{p}\right).$$

If $n = 1$, so that $G$ consists of the identity alone, there are no poles, and the sum on the right has no term. In all other cases $n \geqslant 2$, and therefore

$$1 \leqslant 2 - \frac{2}{n} < 2.$$

It follows that the number of sets of poles can only be 2 or 3; for, the single term $1 - 1/p$ would be less than 1, and the sum of 4 or more terms would be

$$\geqslant 4(1 - \tfrac{1}{2}) = 2.$$

If there are 2 sets of poles, we have

$$2 - \frac{2}{n} = 1 - \frac{1}{p_1} + 1 - \frac{1}{p_2},$$

that is,

$$\frac{n}{p_1} + \frac{n}{p_2} = 2.$$

But two positive integers can have the sum 2 only if each equals 1; thus

$$p_1 = p_2 = n,$$

each of the 2 sets of poles consists of one $n$-gonal pole, and we have the cyclic group $C_n$ with a pole at each end of its single axis.

Finally, in the case of 3 sets of poles we have

$$2 - \frac{2}{n} = 1 - \frac{1}{p_1} + 1 - \frac{1}{p_2} + 1 - \frac{1}{p_3},$$

whence

**15.42**
$$\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} = 1 + \frac{2}{n}.$$

Since this is greater than $\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$, the three periods $p_i$ cannot all be 3 or more. Hence at least one of them is 2, say $p_3 = 2$, and we have

$$\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{2} + \frac{2}{n},$$

whence
$$(p_1 - 2)(p_2 - 2) = 4(1 - p_1p_2/n) < 4$$

(cf. 10.33), so that the only possibilities (with $p_1 \leqslant p_2$ for convenience) are:

$$p_1 = 2,\ p_2 = p,\ n = 2p; \qquad p_1 = 3,\ p_2 = 3,\ n = 12;$$
$$p_1 = 3,\ p_2 = 4,\ n = 24; \qquad p_1 = 3,\ p_2 = 5,\ n = 60.$$

We recognize these as the dihedral, tetrahedral, octahedral and icosahedral groups.

This completes our proof [Klein **3,** p. 129] that

**15.43** *The only finite groups of rotations in three dimensions are the cyclic groups $C_p$ ($p = 1, 2, \ldots$), the dihedral groups $D_p$ ($p = 2, 3, \ldots$), the tetrahedral group $A_4$, the octahedral group $S_4$, and the icosahedral group $A_5$.*

(To avoid repetition, we have excluded $D_1$ which, when considered as a group of rotations, is not only abstractly but geometrically identical with $C_2$.)

Any solid having one of these groups for its complete symmetry group

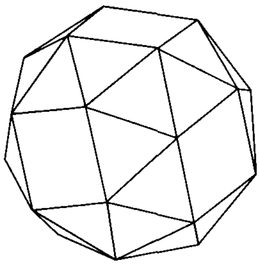


**Figure 15.4e**

(such as the Archimedean *snub cube** shown in Figure 15.4*e*, whose group is $S_4$) can occur in two *enantiomorphous* varieties, *dextro* and *laevo* (i.e., right- and left-handed): mirror images that cannot be superposed by a continuous motion.

### EXERCISES

**1.** Interpret the following permutations as rotations of the octahedron (Figure 15.4*b*):

$$(a\,b\,c\,d), \quad (a\,b\,c), \quad (a\,b), \quad (a\,b)(c\,d).$$

Count the rotations of each type, and check with the known order of $S_4$.

**2.** Using the symbol $(p_1, p_2, p_3)$ for the group having three sets of poles of periods $p_1, p_2, p_3$, consider the possibility of stretching the notation so as to allow $(1, p, p) \cong C_p$ as well as

$$(2, 2, p) \cong D_p, \qquad (2, 3, 3) \cong A_4,$$
$$(2, 3, 4) \cong S_4, \qquad (2, 3, 5) \cong A_5.$$

## 15.5 FINITE GROUPS OF ISOMETRIES

Having enumerated the finite groups of rotations, we can easily solve the wider problem of enumerating the finite groups of isometries (cf. § 2.7). Since every such group leaves one point invariant, we are concerned only with isometries having fixed points. Such an isometry is a rotation or a rotatory inversion according as it is direct or opposite (7.15, 7.41).

If a finite group of isometries consists entirely of rotations, it is one of the groups $G$ considered in § 15.4. If not, it contains such a group $G$ as a subgroup of index 2, that is, it is a group of order $2n$ consisting of $n$ rotations $S_1, S_2, \ldots, S_n$ and an equal number of rotatory inversions $T_1$,

* The vertices of the snub cube constitute a distribution of 24 points on a sphere for which the smallest distance between any 2 is as great as possible. This was conjectured by K. Schütte and B. L. van der Waerden (*Mathematische Annalen,* **123** (1951), pp. 108, 123) and was proved by R. M. Robinson (*ibid.,* **144** (1961), pp. 17–48). The analogous distribution of 6 or 12 points is achieved by the vertices of an octahedron or an icosahedron, respectively. For 8 points the figure is not, as we might at first expect, a cube, but a square antiprism [Fejes Tóth **1**, pp. 162–164].

$T_2, \ldots, T_n$. For, if the group consists of $n$ rotations $S_i$ and (say) $m$ rotatory inversions $T_i$, we can multiply by $T_1$ so as to express the same $n + m$ isometries as $S_iT_1$ and $T_iT_1$. The $n$ isometries $S_iT_1$, being rotatory inversions, are the same as $T_i$ (suitably rearranged if necessary), and the $m$ isometries $T_iT_1$, being rotations, are the same as $S_i$. Therefore $m = n$.

If the central inversion I belongs to the group, the $n$ rotatory inversions are simply

$$S_iI = IS_i \quad (i = 1, 2, \ldots, n),$$

and the group is the direct product $G \times \{I\}$, where $G$ is the subgroup consisting of the S's and $\{I\}$ denotes the group of order 2 generated by I. (As an abstract group, $\{I\}$ is, of course, the same as $C_2$ or $D_1$.)

If I does not belong, the $2n$ transformations $S_i$ and $T_iI$ form a group of rotations of order $2n$ which has the same multiplication table as the given group consisting of $S_i$ and $T_i$. For, if $S_iT_j = T_k$,

$$S_iT_jI = T_kI,$$

and if $T_iT_j = S_k$,

$$T_iIT_jI = T_iI^2T_j = T_iT_j = S_k.$$

In other words, a group of $n$ rotations and $n$ rotatory inversions, not including I, is isomorphic to a rotation group $G'$ of order $2n$ which has a subgroup $G$ of order $n$. To complete our enumeration, we merely have to seek such pairs of related rotation groups. Each pair yields a "mixed" group, say $G'G$, consisting of all the rotations in the smaller group $G$, along with the remaining rotations in $G'$ each multiplied by the central inversion I. Looking back at § 15.4, we see that the possible pairs are

$$C_{2n}C_n, \quad D_nC_n, \quad D_nD_{\frac{1}{2}n}\,(n \text{ even}), \quad S_4A_4.$$

Thus we can complete Table III on p. 413.

## EXERCISES

**1.** Determine the symmetry groups of the following figures: (a) an orthoscheme $O_0O_1O_2O_3$ (Figure 10.4*c*) with $O_0O_1 = O_2O_3$; (b) an $n$-gonal antiprism ($n$ even or odd).

**2.** Designate in the $G'G$ notation the direct product of the group of order 3 generated by a rotation about a vertical axis and the group of order 2 generated by the reflection in a horizontal plane.

## 15.6 GEOMETRICAL CRYSTALLOGRAPHY

> *The sense in which a snail's shell winds is an inheritable character founded in its genetic constitution, as is . . . the winding of the intestinal duct in the species Homo sapiens. . . . Also the deeper chemical constitution of our human body shows that we have a screw, a screw that is turning the same way in every one of us.* . . . A horrid manifestation of this genotypical asymmetry is a metabolic disease called phenylketonuria, leading to insanity, that man contracts when a small quantity of laevo-phenylalanine is added to his food, while the dextro- form has no such disastrous effects.*
>
> H. Weyl [**1**, p. 30]

The discussion of symmetry groups has been phrased in such a way as to be valid not only in Euclidean space but in absolute space. However, it seems appropriate to mention the application of these ideas to the practical science of crystallography. Accordingly, in this digression the geometry is strictly Euclidean.

Crystallographers are interested in those finite groups of isometries which arise as subgroups (and factor groups) of symmetry groups of three-dimensional lattices. By § 4.5, these are the special cases in which the only rotations that occur have periods 2, 3, 4 or 6. This crystallographic restriction reduces the rotation groups to

$$C_1,\ C_2,\ C_3,\ C_4,\ C_6,\ D_2,\ D_3,\ D_4,\ D_6,\ A_4,\ S_4,$$

the direct products to these eleven each multipled by $\{I\}$, and the mixed groups to

$$C_2C_1,\ C_4C_2,\ C_6C_3,\ D_2C_2,\ D_3C_3,\ D_4C_4,\ D_6C_6,\ D_4D_2,\ D_6D_3,\ S_4A_4.$$

(Of course, $C_1 \times \{I\}$ is just $\{I\}$ itself.)

These 32 groups are called the *crystallographic point groups* or "*crystal classes.*" Every crystal has one of them for its symmetry group, and every group except $C_6C_3$ occurs in at least one known mineral. In the more familiar notation of Schoenflies [see, e.g., Burckhardt **1**, p. 71], the groups are respectively

$$C_1,\ C_2,\ C_3,\ C_4,\ C_6,\ D_2,\ D_3,\ D_4,\ D_6,\ T,\ O,$$
$$C_i,\ C_{2h},\ C_{3i},\ C_{4h},\ C_{6h},\ D_{2h},\ D_{3d},\ D_{4h},\ D_{6h},\ T_h,\ O_h,$$
$$C_s,\ S_4,\ C_{3h},\ C_{2v},\ C_{3v},\ C_{4v},\ C_{6v},\ D_{2d},\ D_{3h},\ T_d.$$

To avoid possible confusion, observe that our $C_4C_2$ and $S_4$ ("$S$" for "symmetric") are Schoenflies's $S_4$ and $O$ (for "octahedral"). The 32 groups are customarily divided into seven *crystal systems,* as follows:

| | | |
|---|---|---|
| Triclinic: | $C_1$, $\{I\}$. | |
| Monoclinic: | $C_2$, $C_2 \times \{I\}$, $C_2C_1$. | |
| Orthorhombic: | | $D_2$, $D_2 \times \{I\}$, $D_2C_2$. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Rhombohedral: | $C_3$, | $C_3 \times \{I\}$, | | $D_3$, | $D_3 \times \{I\}$, | $D_3C_3$. | |
| Tetragonal: | $C_4$, | $C_4 \times \{I\}$, | $C_4C_2$, | $D_4$, | $D_4 \times \{I\}$, | $D_4C_4$, | $D_4D_2$. |
| Hexagonal: | $C_6$, | $C_6 \times \{I\}$, | $C_6C_3$, | $D_6$, | $D_6 \times \{I\}$, | $D_6C_6$, | $D_6D_3$. |
| Cubic: | $A_4$, | $A_4 \times \{I\}$, | | $S_4$, | $S_4 \times \{I\}$, | $S_4A_4$. | |

Table I (on p. 413) is a complete list of the 17 discrete groups of isometries in two dimensions involving two independent translations. The analogous groups in three dimensions are the discrete groups of isometries involving three independent translations. The enumeration of these *space groups* is the central problem of mathematical crystallography. The complete list contains $65 + 165 = 230$ groups.

The first 65 are composed entirely of *direct* isometries. Although these were enumerated as long ago as 1869 by C. Jordan [see Hilton **1**, p. 258], they are usually attributed to L. Sohncke who, in 1879, pointed out their application to crystallography. The most obvious group consists of translations alone. The remaining 64 of the 65 contain also rotations and screw displacements; 22 of them occur in 11 enantiomorphous pairs which are mirror images of each other (one containing right-handed screw displacements and the other the reflected left-handed screw displacements). This explains the phenomenon of optical activity [Sayers and Eustace **1**, pp. 238–241, 248–252]. From the standpoint of pure geometry or pure group theory, it would be more natural to ignore this distinction of sense, thus reducing the number 65 to 54, and the total of 230 to 219 [Burckhardt **1**, p. 161].

The remaining 165 groups contain not only direct but also *opposite* isometries: reflections, rotatory reflections (or rotatory inversions), and glide reflections. Their enumeration, by Fedorov in Russia (1890), Schoenflies in Germany (1891), and Barlow in England (1894), provides one of the most striking instances of independent discovery in different places using different methods. Fedorov, who obtained the 230 as $73 + 54 + 103$ instead of $65 + 165$, was probably unaware of the preliminary work of Jordan and Sohncke. It is quite certain that Schoenflies knew nothing of Fedorov, and that Barlow's work was independent of both.

**EXERCISE**

Determine the symmetry groups of the following figures: (a) a rectangular parallelepiped (e.g., a brick), (b) a rhombohedron; (c) a regular dodecahedron with an inscribed cube (whose 8 vertices occur among the 20 vertices of the dodecahedron).

## 15.7 THE POLYHEDRAL KALEIDOSCOPE

*In combining three reflections . . . the effect is highly pleasing*

Sir Dovid Brewster (1781 -1848)

[Brewster **1**, p. 93]

Table III (on p. 413) is a complete list of the finite groups of isometries. In the preceding section, we selected from this list those groups which satisfy the crystallographic restriction. Another significant way to make a selection (partly overlapping with the previous way) is to pick out those groups which are generated by reflections, namely,

$$D_nC_n\ (n \geqslant 1),\quad D_{2n}D_n\ (n \text{ odd}),\quad D_n \times \{\mathrm{I}\}\ (n \text{ even}),$$
$$S_4A_4,\quad S_4 \times \{\mathrm{I}\},\quad A_5 \times \{\mathrm{I}\}.$$

(We have now returned to absolute geometry!)

$D_1C_1$ (Schoenflies's $C_s$, which we previously denoted by $C_2C_1$) is the group of order 2 generated by a single reflection. $D_2C_2$ or $D_2D_1$ (Schoenflies's $C_{2v}$) is the group of order 4 generated by two orthogonal reflections. The remaining groups $D_nC_n$ are the symmetry groups of the $n$-gonal pyramids. In other words, these are the groups $D_n$ of § 2.7 in a different notation. (We now reserve the symbol $D_n$ for the dihedral group of *rotations,* which is, of course, isomorphic to $D_nC_n$. Weyl [**1**, p. 80] makes the distinction by calling the rotation group $D'_n$ and the mixed group $D'_nC_n$.)

$D_2 \times \{\mathrm{I}\}$ is a group of order 8 (abstractly $C_2 \times C_2 \times C_2$) generated by three orthogonal reflections. The remaining groups $D_{2n}D_n$ ($n$ odd) and $D_n \times \{\mathrm{I}\}$ ($n$ even) are the symmetry groups of the $n$-gonal prisms, or of their reciprocals, the dipyramids.

$S_4A_4$, the symmetry group of the regular tetrahedron, is derived from the rotation group $A_4$ by adjoining reflections, such as the reflection in the plane $ABA'B'$ (Figure 10.5$a$) which joins the edge $AB$ to the midpoint of the opposite edge $CD$. (The product of this reflection and the central inversion is the half-turn about the join of the midpoints of the two opposite edges $CD'$, $C'D$ of the cube. This half-turn, which interchanges the two reciprocal tetrahedra $ABCD$, $A'B'C'D'$, is one of the twelve rotations in $S_4$ that do not belong to the subgroup $A_4$; thus it illustrates our special meaning for the "mixed" symbol $S_4A_4$.) Since the remaining Platonic solids are centrally symmetrical, their symmetry groups are simply $S_4 \times \{\mathrm{I}\}$ and $A_5 \times \{\mathrm{I}\}$.

For a practical demonstration in Euclidean space, take the two hinged mirrors of § 2.7, inclined at $180°/n$, which demonstrate the group $D_nC_n$. Standing them upright on a separate horizontal mirror, we obtain the symmetry group of the $n$-gonal prism, i.e., the direct product of $D_nC_n$ and the group of order 2 generated by the horizontal reflection. To demonstrate the three remaining groups, remove the third mirror, and let the first two stand vertically on the table at an angle of 60°, as in the demonstration of $D_3C_3$. Now hold the third mirror obliquely, with its horizontal edge $l$ on the table top at right angles to one of the vertical mirrors and touching the front lower corner of the other. Gradually rotating this third mirror about its edge $l$ from an almost horizontal position (by raising its nearer edge, opposite to $l$), we observe at a certain stage two faces of a regular tetrahedron $\{3, 3\}$. Each face is subdivided into six right-angled triangles, one of which is actually the exposed portion of the table top. At a later stage we see three faces of an octahedron $\{3, 4\}$; still later, four faces of an icosahedron $\{3, 5\}$. Finally, when the adjustable mirror is vertical like the others, we see a theoretically infinite number of faces of the regular tessellation $\{3, 6\}$, subdivided in the manner of Figure 4.6$d$. This device, employing ordinary rectangular mirrors, is a simplified version of Möbius's trihedral kaleidoscope in which the three mirrors are cut in the shape of suitable sectors of a circle [Coxeter **1**, p. 83].

When the $E$ edges of the general Platonic solid $\{p, q\}$ are projected from its center onto a concentric sphere, they become $E$ arcs of great circles, decomposing the surface into $F$ regions which are "spherical $p$-gons." In this manner the polyhedron yields a "spherical tessellation" which closely resembles the plane tessellation of § 4.6. The symmetry group of $\{p, q\}$ is derived from the symmetry group of one face by adding the reflection in a side of that face. Thus it is generated by reflections in the sides of a spherical triangle whose angles are $\pi/p$ (at the center of a face), $\pi/2$ (at the midpoint of an edge), and $\pi/q$ (at a vertex). This spherical triangle is a fundamental region for the group, since it is transformed into neighboring regions by the three generating reflections.

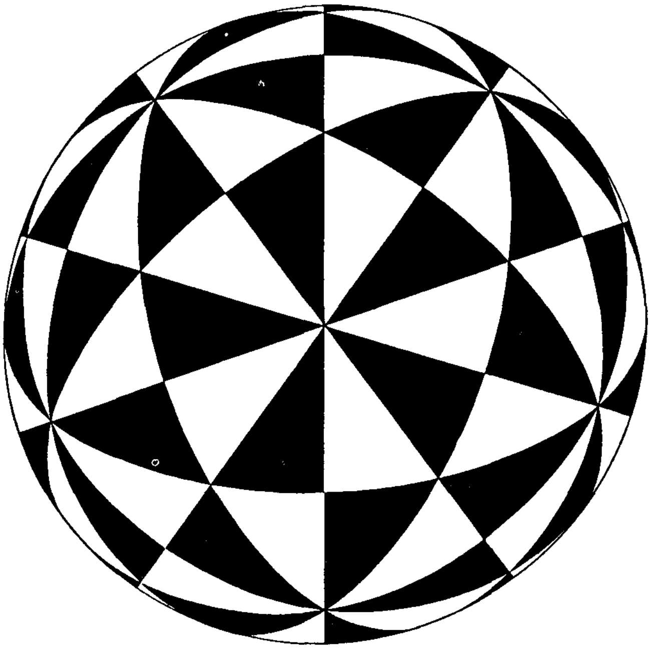


**Figure 15.7a**

The network of such triangles, filling the surface of the sphere, is cut out by all the planes of symmetry of the polyhedron, namely the planes joining the center to the edges of both $\{p, q\}$ and its reciprocal $\{q, p\}$. In Figure 15.7*a* (where $p$ and $q$ are 3 and 5), alternate regions have been blackened so as to exhibit both the complete symmetry group $A_5 \times \{I\}$ and the rotational subgroup $A_5$, which preserves the coloring.

Instead of deriving the network of spherical triangles from the regular polyhedron, we may conversely derive the polyhedron from the network. The ten triangles in the middle of Figure 15.7*a* evidently combine to form a face of the blown-up dodecahedron, and the six triangles surrounding a point where the angles are 60° combine to form a face of the blown-up icosahedron.

### EXERCISES

1. Interpret the symbol $\{p, 2\}$ as a spherical tessellation ("dihedron") whose faces consist of two hemispheres, and $\{2, p\}$ as another whose faces consist of $p$ lunes.

2. How many planes of symmetry does each Platonic solid have? Provided $p$ and $q$ are greater than 2, this number is always a multiple of 3, namely $3c$ in the notation of Ex. 1 at the end of § 10.4.

3. Dividing $4\pi$ by the area of the fundamental region, obtain a formula for the order of the symmetry group of $\{p, q\}$. Reconcile this with the formula for $E$ (the number of edges) in 10.32.

## 15.8 DISCRETE GROUPS GENERATED BY INVERSIONS

In the present section we make one more digression into Euclidean space, so as to be able to talk about inversion. (The absolute theory of inversion presents difficulties that would take us too far afield. [See Sommerville **1**, Chapter VIII.])

Figure 15.7$a$, being an orthogonal projection, represents 10 of the 15 great circles by ellipses. (The difficult task of drawing it was undertaken by J. F. Petrie about 1932). An easier, and perhaps more significant, way to represent such figures is by stereographic projection (§ 6.9), so that the great circles remain circles (or lines) [Burnside **1**, pp. 406–407]. The reader can readily do this for himself, with the aid of the following simple instructions.

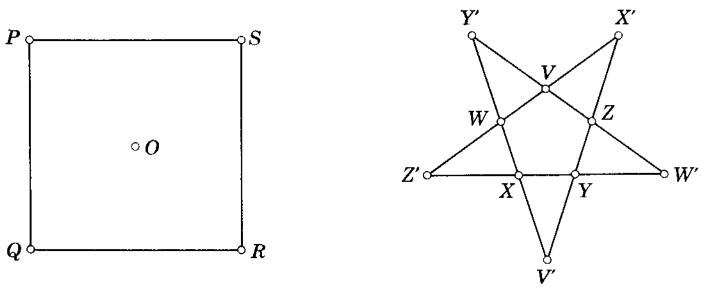



**Figure 15.8a**

Figure 15.8$a$ shows a square $PQRS$ with center $O$, and a regular pentagon $VWXYZ$ with its sides extended to form a pentagram $V'X'Z'W'Y'$. With radius $PQ$ and centers $P$, $Q$, $R$, $S$, draw four circles. These, along with two lines through $O$ parallel to the sides of the square, represent 6 great circles, one in each of the 6 planes of symmetry of the tetrahedron $\{3, 3\}$, which are the planes joining pairs of opposite edges of a cube. Adding the circumcircle and diagonals of the square, we have altogether 9 great circles, one in each of the 9 planes of symmetry of the cube $\{4, 3\}$, which include 3 planes parallel to its faces.

With radius $VX'$ ($= VX$) and centers $V$, $W$, $X$, $Y$, $Z$, draw five circles. With radius $VW'$ ($= V'W'$) and centers $V'$, $W'$, $X'$, $Y'$, $Z'$, draw five more circles. These ten circles, along with the five lines $VV'$, $WW'$, $XX'$, $YY'$, $ZZ'$, represent 15 great circles (Figure 15.7*a*), one in each of the 15 planes of symmetry of the icosahedron $\{3, 5\}$ or of the dodecahedron $\{5, 3\}$. (These planes join pairs of opposite edges of either solid.)

To justify these statements we merely have to examine the curvilinear triangles* and observe that each has angles $\pi/p$, $\pi/q$, $\pi/2$.

Since stereographic projection is an inversion (Figure 6.9*a*), and since an inversion transforms a reflection into an inversion, the figures so constructed are, in fact, representations of the abstract groups $S_4$, $S_4 \times C_2$, and $A_5 \times C_2$ as groups generated by inversions. In other words, they are configurations of circles so arranged that the whole figure is symmetrical by inversion in each circle. (Of course, any straight lines that occur are to be regarded as circles of infinite radius. As we saw in § 6.4, inversion in such a "circle" is simply reflection in the line.) Any one of the regions into which the plane is decomposed will serve as a fundamental region, and the generators of the group may be taken to be the inversions in its sides.

For a group generated by just one inversion, we may invert the circle into a straight line so as to obtain the group $D_1$ of order 2, generated by a single reflection (§ 2.5). The groups generated by inversions in two intersecting circles are essentially the same as the groups $D_n$ of order $2n$, generated by reflections in two intersecting lines (§ 2.7). If the circles of two generating inversions are in contact, they can be inverted into parallel lines, and we have the limiting case $D_\infty$ (Figure 3.7*b*). Two nonintersecting circles can be inverted into concentric circles. Inversions in them generate an infinite sequence of concentric circles whose radii are in geometric progression. Abstractly, the group is again $D_\infty$, but the center is a "point of accumulation" (§ 7.6). So is the point of contact in the case of the group generated by inversions in two touching circles. A group is said to be *discrete* if it has no points of accumulation. Thus, in describing discrete groups generated by inversions, we may insist that every two of the generating circles intersect properly, and do not touch.

For a discrete group generated by three inversions, the fundamental region is a curvilinear triangle whose angles are submultiples of $\pi$: say $\pi/p_1$, $\pi/p_2$, $\pi/p_3$. For instance, two radii of a circle, forming an angle $\pi/p$, cut out a sector which may be regarded as a "triangle" with angles $\pi/p$, $\pi/2$, $\pi/2$; this is a fundamental region for the group $D_p \times D_1$ of order $4p$, generated by reflections in the radii and inversion in the circle. In this case

* For the effect of projecting in a different direction, see Coxeter, *American Mathematical Monthly,* **45** (1938), pp. 523–525, Figs. 4 and 5.

**15.81** $$\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} > 1,$$

so that the angle sum of the triangle is greater than $\pi$: an obvious consequence of the fact that the sector is derived from a spherical triangle (see § 6.9) by stereographic projection, which preserves angles. Every solution of the inequality 15.81 (cf. 15.42) is a triangle that can be drawn with great circles on a sphere. We thus obtain again the symmetry groups

$$S_4, \quad S_4 \times C_2, \qquad A_5 \times C_2$$

of the Platonic solids.

When $1/p_1 + 1/p_2 + 1/p_3 = 1$, so that the angle sum is exactly $\pi$, we have the infinite "Euclidean" groups **p6m**, **p4m**, **p31m** (see Table I and Figure 4.6*d*). We could transform all the straight lines into circles by means of an arbitrary inversion; but then, since the pattern is infinitely extended, the center of inversion would be a point of accumulation.

When $1/p_1 + 1/p_2 + 1/p_3 < 1$, so that the angle sum of the fundamental region is less than $\pi$, we may still take two of the three sides to be straight, but now their point of intersection $A$ is outside the circle $q$ to which the third side belongs, with the result that there is a circle $\Omega$ orthogonal to all three (Figure 15.8*b*); the tangents from $A$ to $q$ are radii of $\Omega$.

Since $\Omega$ is invariant for each of the generating inversions, it is invariant for the whole group. The circle $q$ decomposes the interior of $\Omega$ into two

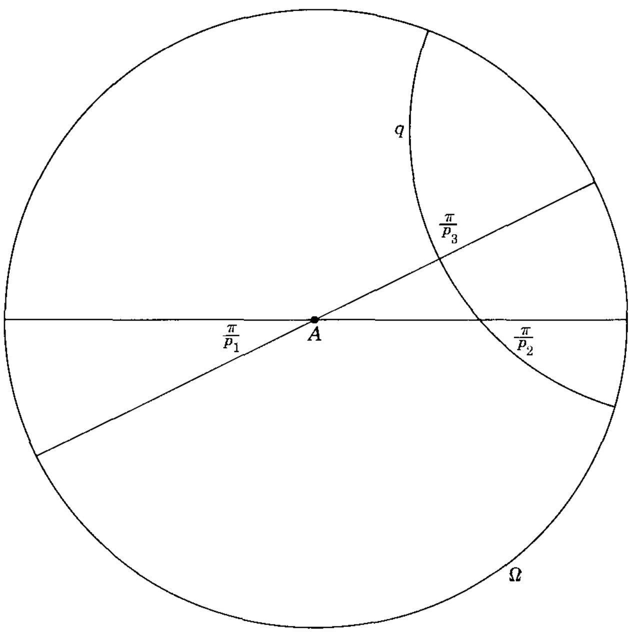



**Figure 15.8b**

unequal regions and inverts each of these regions into the other. Therefore the number of triangles is the same in both regions. But the larger region includes a replica of the smaller. Hence, by Bolzano's definition of an infinite set (namely, a set that has the same power as a proper subset), the number of triangles is infinite; that is, *the group is infinite.*

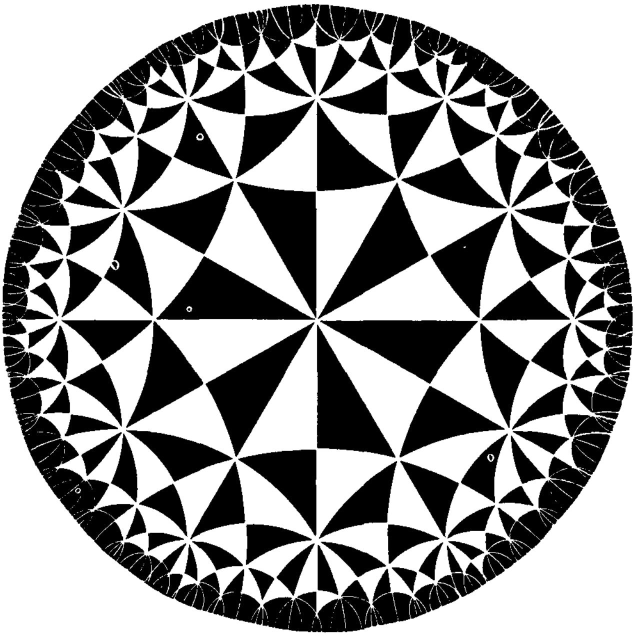


**Figure 15.8c**

The case when $p_1, p_2, p_3$ are 6, 4, 2 is shown in Figure 15.8*c*. Unlike Figure 15.7*a*, this is not a picture of a solid object. Our familiarity with three-dimensional space enables us to accept the idea that the triangles in Figure 15.7*a* are all the same size, even though the peripheral ones are made to look smaller by perspective foreshortening. In the case of Figure 15.8*c*, the smaller peripheral triangles are essentially the same shape as those in the middle (since they have the same angles), but we no longer find it easy to imagine that they are, in some sense, the same *size*. In trying to stretch our imagination to this extent, we are taking a first step towards appreciating hyperbolic geometry, which is the subject of our next chapter.

The reader may wonder why we admit such groups as being worthy of consideration, seeing that the circle $\Omega$ contains infinitely many points of accumulation. However, when we accept the non-Euclidean standpoint, so that the circles and inversions are regarded as lines and reflections, the consequent distortion of distance makes $\Omega$ infinitely far away, so that the points of accumulation disappear.

## EXERCISES

**1.** If a system of concentric circles is transformed into itself by inversion in each circle, the radii are in geometric progression.

**2.** If three circles form a "triangle" with angles $\pi/p_1$, $\pi/p_2$, $\pi/p_3$, the inversions $R_1$, $R_2$, $R_3$ in its sides satisfy the relations

$$R_1^2 = R_2^2 = R_3^2 = (R_2R_3)^{p_1} = (R_3R_1)^{p_2} = (R_1R_2)^{p_3} = 1.$$

These relations suffice to define the abstract group generated by $R_1$, $R_2$, $R_3$ [Coxeter and Moser **1**, pp. 37, 55].

**3.** Given an angle $\pi/p_1$ at the center $A$ of a circle $\Omega$ of unit radius, as in Figure 15.8*b*, find expressions (in terms of $p_1$ and $p_2$) for the radius of the circle $q$ and for the distance from $A$ to its center, in the case when $p_3 = 2$.

**4.** Invert Figure 15.8*c* in a circle whose center lies on $\Omega$; that is, replace the circle $\Omega$ by a straight line, so that all the inverting circles have their centers on this line. (Such an arrangement provides an alternative proof that the group is infinite. For if its order is $g$, the infinite half plane is filled with $g$ curvilinear triangles, each having a finite area!)

**5.** In Figure 15.8*c*, two of the small triangles (one white and one black) with a common hypotenuse form together a "curvilinear kite" having three right angles and one angle of 60°. Trace part of the figure so as to exhibit a network of such kites, alternately white and black. We now have an instance of a group generated by four inversions. Can it happen that more than four inversions are needed to generate a discrete group?

# 16
# *Hyperbolic geometry*

Absolute geometry is not *categorical:* it is two geometries in one. To be precise, it leaves open the question of the existence of ultraparallel lines (see the end of § 15.2). In § 16.1 we shall compare the two possible answers, giving the unfamiliar the same status as the familiar. In § 16.2 we shall justify this action by means of a proof of *relative consistency.* Thereafter, casting aside all scruples, we shall plunge wholeheartedly into the "new universe" which Bolyai "created from nothing."

## 16.1 THE EUCLIDEAN AND HYPERBOLIC AXIOMS OF PARALLELISM

*In the author there lives the perfectly purified conviction (such as he expects too from every thoughtful reader) that by the elucidation of this subject one of the most important and brilliant contributions has been made to the real victory of knowledge, to the education of the intelligence, and consequently to the uplifting of the fortunes of men.*

J. Bolyai (1802-1860)

[Carslaw **1**, p. 31]

In § 12.6, we mentioned the question whether the two rays parallel to a given line $r$ from an outside point $A$ are, or are not, collinear. By applying a suitable isometry, we see that the answer is independent of the position of $r$.

It is true, though less obvious, that, for a given $r$, the answer is independent of the position of $A$. Suppose, if possible, that the rays parallel to $r$ from $A$ are the two halves of a line $q$ while the rays parallel to $r$ from another point $A'$ form an angle, as in Figure 16.1*a*. By the transitivity of parallelism, these rays from $A'$ are parallel to $q$ and also to the infinite sequence of parallel lines derived from $q$ and $r$ by applying the group $D_\infty$ generated by reflections in $q$ and $r$ (Figure 3.7*b*). We obtain a manifest absurdity by considering any one of these lines that lies beyond $A'$ (i.e., in such a position that $A'$ lies between that line and $r$). (Strictly, this argument makes use of the so-called Axiom of Archimedes, 13.31, which is a consequence of 12.51.)

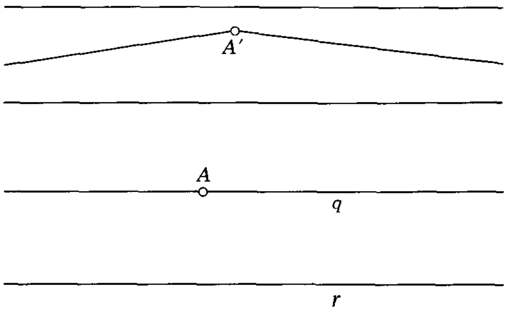



**Figure 16.1a**

Thus we have a clear-cut distinction between two kinds of geometry, called *Euclidean* and *hyperbolic,* which are derived from absolute geometry by adding just one of the following two alternative axioms:

**THE EUCLIDEAN AXIOM.** *For some point A and some line r, not through A, there is not more than one line through A, in the plane Ar, not meeting r.*

**THE HYPERBOLIC AXIOM.** *For some point A and some line r, not through A, there is more than one line through A, in the plane Ar, not meeting r.*

**EXERCISE**

Each of these axioms implies the stronger statement with "some point $A$ and some line $r$" replaced by "any point $A$ and any line $r$." The Euclidean axiom, so amended, is equivalent to the celebrated Postulate V (our 1.25). How does Postulate V break down if we assume the hyperbolic axiom?

## 16.2 THE QUESTION OF CONSISTENCY

*What are we to think of the question: Is Euclidean Geometry true? It has no meaning. We might as well ask . . . if Cartesian coordinates are true and polar coordinates false. One geometry cannot be more true than another; it can only be more convenient.*

H. Poincaré (1854-1912)

(*Science and Hypothesis,* New York, 1952)

We observe that the Euclidean and hyperbolic axioms differ by just one word: the vital word "not." It is meaningless to ask which of the two geometries is *true,* and practically impossible to decide which provides a more *convenient* basis for describing astronomical space. From the standpoint of pure mathematics, a more important question is whether either axiom is logically *consistent* with the remaining axioms of absolute geometry. Even this is difficult to answer; for according to the philosopher Gödel, there is no internal proof of consistency for a system that includes infinite sets. We have to be content with *relative* consistency: if Euclidean geometry is free from contradiction, so is hyperbolic geometry, and vice versa. Relative consistency is established by finding in each geometry a *model* of the other.

One Euclidean model of the hyperbolic plane (due to Poincaré) was mentioned in § 15.8. This uses a circle Ω, as in Figure 16.2*a*. Each pair of inverse points represents a hyperbolic point, and each circle orthogonal to Ω represents a hyperbolic line. The two parallels to *r* from *A* are simply the circles through *A* that touch *r* at its points of intersection with Ω. (These points are the "ends" of *r*.) We call this a *conformal* model because angles retain their proper values though distances are inevitably distorted.

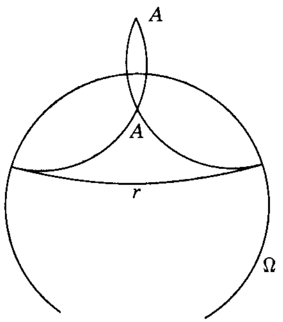



**Figure 16.2a**

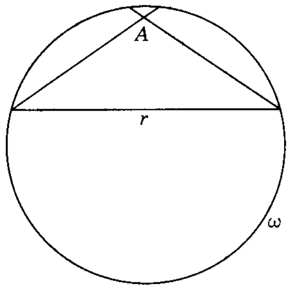



**Figure 16.2b**

A different Euclidean model, suggested by Beltrami (1835–1900), uses another circle ω, as in Figure 16.2*b*. Each point inside ω represents a hyperbolic point. The two parallels to *r* from *A* are the chords joining *A* to the ends of the chord *r*. (Chords whose lines intersect outside ω represent ultraparallel lines.) We call this a *projective* model because straight lines remain straight. Nothing is lost if we replace the circle ω in the Euclidean plane by a conic in the projective plane. In fact, much is gained; for it is possible to extend the hyperbolic plane into a projective plane by means of entities defined in the hyperbolic geometry itself [Coxeter **3**, p. 196]. In this way we can prove that hyperbolic geometry is unique or *categorical* [Borsuk and Szmielew **1**, p. 345], unlike absolute geometry, which includes two contrasting possibilities.

When using models, it is desirable to have two rather than one, so as to avoid the temptation to give either of them undue prominence. Our geometric reasoning should all depend on the axioms. The models, having served their purpose of establishing relative consistency [Pedoe **1**, p. 61; Sommerville **1**, pp. 154–159], are no more essential than diagrams.

Klein [**4**, p. 296] exhibited a connection between the conformal and projective models in the manner of Figure 16.2*c*. A sphere, having the same radius as ω, touches the (horizontal) plane at *S*, the center of both ω and Ω.

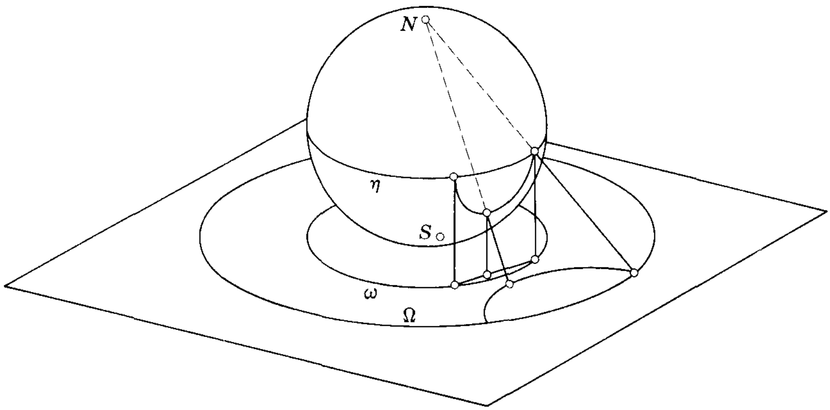



Figure 16.2c

Beginning with the projective model, we use orthogonal (vertical) projection to map $\omega$ on the "equator" $\eta$ of the sphere, and each interior point on two points: one in the southern hemisphere and another (not shown) in the northern hemisphere. Every chord of $\omega$ yields a circle in a vertical plane, that is, a circle orthogonal to $\eta$. We now map the sphere back into the plane by stereographic projection, so that $\eta$ projects into the larger circle $\Omega$, concentric with $\omega$. Because of the angle-preserving and circle-preserving nature of stereographic projection, the vertical circles yield horizontal circles orthogonal to $\Omega$, and we have the conformal model.

Instead of stereographic projection onto the tangent plane at the "south pole" $S$ (i.e., inversion with respect to a sphere of radius $NS$), we could have used stereographic projection (from the same "north pole" $N$) onto the equatorial plane (i.e., inversion with respect to a sphere through $\eta$) so as to make both $\omega$ and $\Omega$ coincide with $\eta$ [Coxeter **3,** p. 260]. Klein's procedure is justified by its property of making the two models agree in the immediate vicinity of $S$. This must have seemed to him more important than making them agree "at infinity."

It must be remembered that both models are in one respect misleading: they give us the impression that the center $S$ should play a special role, whereas, in the abstract hyperbolic plane, all points are alike.

For the sake of completeness, we should mention the problem that the inhabitants of a hyperbolic world would face in trying to visualize the Euclidean plane. One solution [Coxeter **3,** pp. 197–198] is that they could represent the Euclidean points and lines by the lines and planes parallel to a given ray in hyperbolic space!

## EXERCISES

1. Reflection in a line of the hyperbolic plane appears, in the conformal model, as inversion with respect to a circle, and in the projective model as a harmonic homology. What is the corresponding transformation in the space of Klein's sphere?

2. Circles appear as circles (not meeting $\Omega$) in the conformal model, and therefore as circles on the sphere (say, in the southern hemisphere) and as ellipses in the projective model.

## 16.3 THE ANGLE OF PARALLELISM

*. . . a sea-change into something rich and strange.*

W. Shakespeare (1564-1616)
(*The Tempest*, Act I, Scene 2)

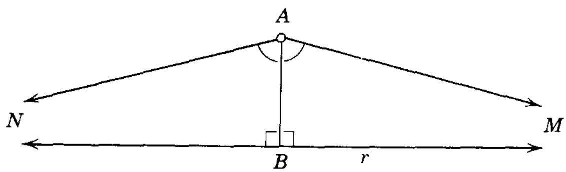



**Figure 16.3a**

For the rest of this chapter the geometry will be hyperbolic, that is, we shall assume the hyperbolic axiom, which implies that, for any point $A$ and line $r$, not through $A$, the two parallels form an angle $NAM$, as in Figure 16.3*a*. From $A$ draw $AB$ perpendicular to $r$. Reflection in $AB$ shows that $\angle BAM$ and $\angle NAB$ are equal acute angles. Following Lobachevsky, we call either of them the *angle of parallelism* corresponding to the distance $AB$, and write

$$\angle BAM = \Pi(AB).$$

Before we can prove that this function is monotonic, we need a few more properties of asymptotic triangles. While proving 15.26 we discovered that, if a transversal ($AD$ in Figure 15.2*f*) meets two lines in such a way that the "alternate" angles are equal, then the two lines are ultraparallel. Hence [Carslaw **1**, p. 48]:

**16.31** *In an asymptotic triangle EFM, the external angle at E (or F) is greater than the internal angle at F (or E).*

In other words, the sum of the angles of an asymptotic triangle is less than $\pi$. This will enable us to prove a kind of converse for Theorem 15.25, to the effect that an asymptotic triangle is determined by its two positive angles:

**16.32** *If two asymptotic triangles $AEM$, $A'E'M'$ have $A = A'$ and $E = E'$, then $AE = A'E'$.*

*Proof* [Carslaw **1**, p. 50]. If $AE$ and $A'E'$ are not equal, one of them must

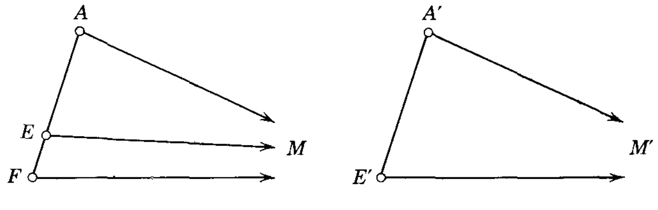



**Figure 16.3b**

be the greater; let it be $A'E'$, as in Figure 16.3*b*. On $E/A$, take $F$ so that $AF = A'E'$, and draw $FM$ parallel to $AM$. By 16.31 and 15.25, we have

$$\angle MEA > \angle MFA = \angle M'E'A' = \angle MEA,$$

which is absurd.

These results will enable us to establish the existence of *a common parallel to two given rays* forming an angle $NOM$, that is, a line $MN$ which is parallel to $OM$ at one end and to $ON$ at the other. From the given rays $OM$, $ON$, cut off any two equal segments $OA$, $OA'$, as in Figure 16.3*c*. Draw $A'M$ parallel to $OM$, and $AN$ parallel to $ON$. Bisect the angles $NAM$ and $NA'M$ by lines $a$ and $a'$. We shall prove that *these lines are ultraparallel,* and that *the desired common parallel $MN$ is perpendicular to both of them.*

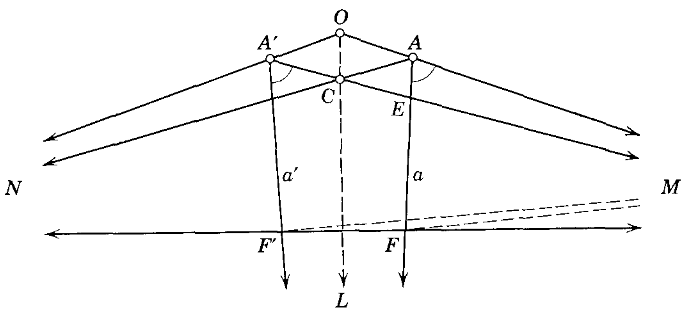



**Figure 16.3c**

Let $A'M$ meet $AN$ in $C$, and $a$ in $E$. Since the whole figure is symmetrical by reflection in $OC$, the two angles at $A$ and the two angles at $A'$ are all equal.

If possible, let $a$ and $a'$ have a common point $L$, which is, of course, equidistant from $A$ and $A'$. Applying 15.25 to the congruent asymptotic triangles $ALM$ and $A'LM$, we deduce that $\angle MLA = \angle MLA'$, which is absurd.

If possible, let $a$ and $a'$ be parallel, with a common end $L$. Applying 16.32 to the congruent asymptotic triangles $AEM$ and $A'EL$, we deduce that $AE = A'E$, whence $E$ coincides with $C$, which is absurd.

We conclude that $a$ and $a'$ are ultraparallel. By 15.26, they have a common perpendicular $FF'$. Applying 15.25 to the congruent asymptotic triangles $AFM$ and $A'F'M$, we conclude that

$$\angle MFA = \angle MF'A'.$$

If $F'F$ were not parallel to $OM$, we would have an asymptotic triangle $FF'M$ whose angle sum is $\pi$, contradicting 16.31. Hence, in fact, $F'F$ is parallel to $OM$, and similarly $FF'$ to $ON$; that is, the line $FF'$ is a common parallel to the two rays as desired.

Moreover, this common parallel is *unique,* since two such would be parallel to each other at both ends, contradicting the "clear-cut distinction" between the Euclidean and hyperbolic properties of parallelism (Figure 16.1*a*). It follows that

**16.33** *Any two ultraparallel lines have a unique common perpendicular.*

For, given $a$ and $a'$, we can reconstruct Figure 16.3*c* as follows: draw any common perpendicular $FF'$, take $O$ on its perpendicular bisector, and let the two parallels through $O$ to the line $FF'$ meet $a$ in $A$, $a'$ in $A'$.

For the sake of brevity, we have been content to assert the *existence* of a line through a given point parallel to a given ray, and of a common perpendicular to two given ultraparallel lines. Actual "ruler and compasses" constructions for these lines have been given by Bolyai and Hilbert, respectively [see Coxeter **3,** pp. 204, 191]. Hilbert apparently failed to notice that his construction for the common parallel to $AM$ and $A'N$ remains valid if these lines meet in a point that is not equidistant from $A$ and $A'$, or even if they do not meet at all. In fact [Carslaw **1,** p. 76],

**16.34** *Any two nonparallel rays have a unique common parallel.*

This result justifies our use of *ends* as if they were ordinary points: any two ends, $M$ and $N$, determine a unique line $MN$.

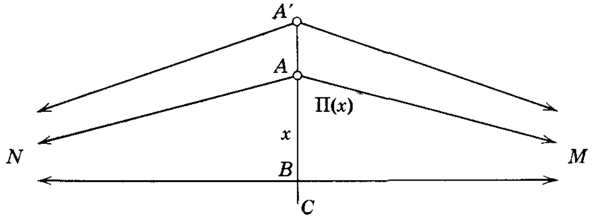



**Figure 16.3d**

The line through $A$ parallel to $BM$ (Figure 16.3*a* or *d*) determines the angle of parallelism $\Pi(AB)$. Conversely, we can now find a distance $x$ whose angle of parallelism $\Pi(x)$ is equal to any given acute angle [Carslaw **1,** p. 77]. In other words, given an acute angle $CAM$, *we can find a line $BM$ which is both perpendicular to $AC$ and parallel to $AM$.* We merely have to reflect $AM$ in $AC$, obtaining $AN$, and then draw the common parallel $MN$, which meets $AC$ in the desired point $B$. Incidentally, since we can draw through any point a ray parallel to a given ray, it follows that

**16.35** *For any two nonperpendicular lines we can find a line which is perpendicular to one and parallel to the other.*

If $A'$ is on the ray $A/B$, so that $A'B > AB$ (as in Figure 16.3$d$), then

$$\Pi(A'B) < \Pi(AB).$$

(This is simply 16.31, applied to the asymptotic triangle $AA'M$.) It follows that the function $\Pi(x)$ decreases steadily from $\frac{1}{2}\pi$ to 0 when $x$ increases from 0 to $\infty$.

We naturally call $AMN$ a *doubly asymptotic triangle* [Coxeter **3**, p. 188]. We have seen that such a "triangle" is determined by its one positive angle; in other words,

**16.36** *Two doubly asymptotic triangles are congruent if they have equal angles.*

Applying 16.34 to rays belonging to two parallel lines $LM$, $LN$, we obtain a third line parallel to both, forming a *trebly asymptotic triangle* $LMN$. In view of Bolyai's remark 15.24, we may regard such a triangle as a doubly asymptotic triangle whose angle is zero. Accordingly, we shall not be surprised to find that

**16.37** *Any two trebly asymptotic triangles are congruent.*

*Proof* (due to D. W. Crowe). Given any two trebly asymptotic triangles, dissect each into two right-angled doubly asymptotic triangles by drawing an *altitude* (perpendicular to one side and parallel to another, as in 16.35). By 16.36, all the four doubly asymptotic triangles are congruent. Therefore the two trebly asymptotic triangles must be congruent.

## EXERCISES

1. Draw figures for Theorems 16.33–16.35 in terms of the conformal and projective models.

2. If a quadrangle $ABED$ has right angles at $D$ and $E$ while $AD = BE$, then the angles at $A$ and $B$ are equal acute angles. (*Hint*: Draw $AM$ and $BM$ parallel to $D/E$; apply 16.31 to the asymptotic triangle $ABM$.)

3. The sum of the angles of any triangle is less than two right angles. (*Hint*: For a given triangle $ABC$, draw $AD$, $BE$, $CF$ perpendicular to the line joining the midpoints of $BC$ and $CA$.)

4. Given an asymptotic triangle $ABM$ with acute angles at both $A$ and $B$, draw $AD$ perpendicular to $BM$, and $BE$ perpendicular to $AM$, meeting in $H$. Draw $HF$ perpendicular to $AB$. Then $FH$ is parallel to $AM$ [Bonola **1**, p. 106]. What happens if we deal similarly with rays through $A$ and $B$ which are not parallel but ultraparallel?

5. If two trebly asymptotic triangles have a common side, by what isometry are

they related? (Of course, two trebly asymptotic triangles may have a common side without having a common altitude).

**6.** The inradius of a trebly asymptotic triangle is the distance whose angle of parallelism is 60°.

**7.** From any point on a side of a trebly asymptotic triangle, lines drawn perpendicular to the other two sides are themselves perpendicular [Bachmann **1**, p. 222].

## 16.4 THE FINITENESS OF TRIANGLES

*I could be bounded in a nutshell and count myself a king of infinite space.*

W. Shakespeare
(*Hamlet*, Act II, Scene 2)

One of the most elegant passages in the literature on hyperbolic geometry since the time of Lobachevsky is the proof by Liebmann [**1**, p. 43] that the area of a triangle remains finite when all its sides are infinite. C. L. Dodgson (*alias* Lewis Carroll) could not bring himself to accept this theorem; consequently he believed non-Euclidean geometry to be nonsense.

Instead of pursuing a philosophical discussion of the meaning of *area* [Carslaw **1**, pp. 84–90], let us be content to regard it as a numerical function, defined for every simple closed polygon, invariant under isometries, and additive when two polygons are juxtaposed.

Let $ABM$ be any asymptotic triangle. Reflect it in the bisector $AF$ of the angle $A$ to obtain $AA_1N$, as in Figure 16.4*a*, $F$ being the point where the bisector meets the common parallel $MN$. Reflect the line $BM$ in the bisector $A_1F_1$ of $\angle NA_1M$ to obtain $A_2N$ (with $A_2$ on $AM$), and then reflect

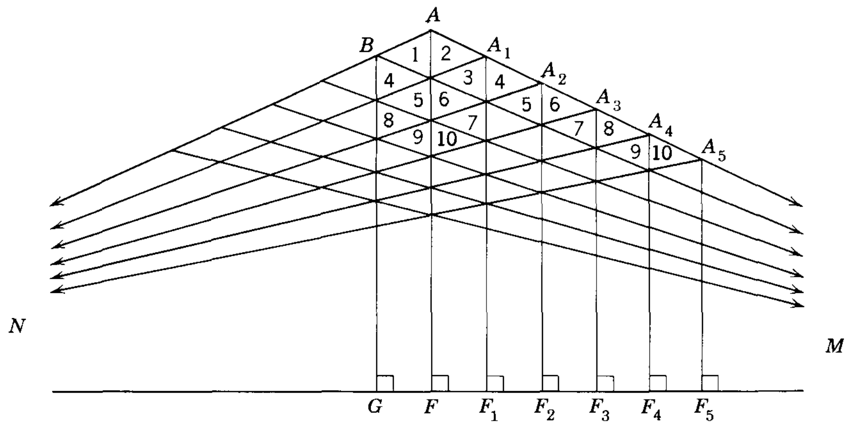



**Figure 16.4a**

this in $AF$. Continuing in this manner, we construct a network of triangles whose "vertical" sides bisect the angles at $B, A, A_1, A_2, A_3, \ldots$ and are perpendicular to $MN$ at $G, F, F_1, F_2, F_3, \ldots$. These points are evenly spaced along $MN$, since they are all derived from $F$ and $F_1$ by the group $D_\infty$ generated by reflections in $AF$ and $A_1F_1$; for instance, $G$ is the image of $F_1$ in the mirror $AF$. The numbered triangles which fit together to fill the asymptotic triangle $ABM$ are respectively congruent to those which fit together within the finite pentagon $ABGF_1A_1$; in fact, any two triangles that are numbered alike are related by some power of the translation from $G$ to $F_1$ (or from $F$ to $F_2$). Hence the area of the asymptotic triangle is less than or equal to the area of the pentagon:

**16.41** *Any asymptotic triangle has a finite area.*

Since any doubly asymptotic triangle (Figure 16.3*a*) can be dissected into two asymptotic triangles, it follows that

**16.42** *Any doubly asymptotic triangle has a finite area.*

By 16.36, the area of a doubly asymptotic triangle is a function of its angle. Comparing the triangles $AMN$ and $A'MN$ of Figure 16.3*d*, we see that this is a *decreasing* function: the larger triangle has the smaller angle.

Since any trebly asymptotic triangle can be dissected into two doubly asymptotic triangles (as in the proof of 16.37), 16.42 implies

**16.43** *Any trebly asymptotic triangle has a finite area.*

By 16.37, this area is a constant, depending only on our chosen unit of measurement.

## 16.5 AREA AND ANGULAR DEFECT

> *Gauss . . . did not recognize the existence of a logically sound non-Euclidean geometry by intuition or by a flash of genius: . . . on the contrary, he had spent upon this subject many laborious hours before he had overcome the inherited prejudice against it. [He] did not let any rumour of his opinions get abroad, being certain that he would be misunderstood. Only to a few trusted friends did he reveal something of his work.*
>
> R. Bonola [**1**, pp. 66-67]

János Bolyai, or Bolyai János (as it is written in Hungarian), announced his discovery of absolute geometry in an appendix to a book by his father, Bolyai Farkas, who was a friend of Gauss. When Gauss saw this book and read the appendix, he wrote a remarkable letter to his old friend, congratulating János and admitting that he himself had thought along the same lines without publishing the results. The original letter (of March 6, 1832) is lost, but the younger Bolyai's copy of it has been preserved, and it was eventually published in Gauss's collected works [Gauss **1**, vol. 8, pp. 220–225].

This letter contains a wonderful proof that the area of a triangle $ABC$ is proportional to its *angular defect*

$$\pi - A - B - C:$$

the amount by which its angle sum falls short of two right angles. The following paraphrase fills up a few gaps in the argument, while retaining Gauss's systematic division into seven steps, numbered with Roman numerals.

I. *All trebly asymptotic triangles are congruent.* (This is our 16.37.)

II. *The area of a trebly asymptotic triangle has a finite value, say t.* (This is our 16.43.)

III. *The area of a doubly asymptotic triangle AMN is a function of its angle, NAM, say $f(\phi)$, where $\phi$ is the supplement of this angle.* Given the angle $\phi$, we can construct the triangle in a unique fashion (Figure 16.5*a*; cf. 16.3*c*). Gauss used the supplement, rather than the angle $NAM$ itself, to ensure that $f(\phi)$ is an *increasing* function of $\phi$. (See the remark after 16.42.)

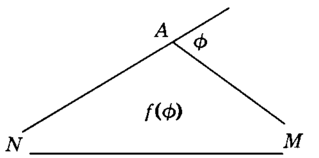



**Figure 16.5a**

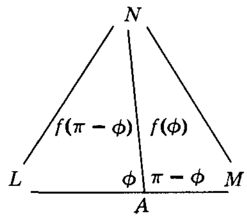



**Figure 16.5b**

IV. $f(\phi) + f(\pi - \phi) = t$.

This may be seen by fitting together two doubly asymptotic triangles $AMN$ and $ANL$ with supplementary angles, as in Figure 16.5*b*. Here it is understood that $0 < \phi < \pi$. But when $\phi$ approaches zero, the doubly asymptotic triangle collapses, and when $\phi$ approaches $\pi$ it tends to become trebly asymptotic. Hence

**16.51** $$f(0) = 0, \qquad f(\pi) = t,$$

and IV is valid for $0 \leqslant \phi \leqslant \pi$.

V. $f(\phi) + f(\psi) + f(\pi - \phi - \psi) = t$.

This, with $\phi > 0, \psi > 0, \phi + \psi < \pi$, may be seen by fitting together three doubly asymptotic triangles whose angles add up to $2\pi$, as in Figure 16.5*c*. It evidently remains valid when $\phi$ or $\psi$ is zero or $\phi + \psi = \pi$.

VI. $f(\phi) + f(\psi) = f(\phi + \psi)$.

This, with $\phi \geqslant 0, \psi \geqslant 0, \phi + \psi \leqslant \pi$, is obtained algebraically, by writing $\phi + \psi$ instead of $\phi$ in IV and then using V. It follows that $f(\phi)$ is simply a multiple of $\phi$, namely,

**16.52** $$f(\phi) = \mu\phi$$

where, by 16.51, $\mu = t/\pi$.

J. H. Lindsay has pointed out that this deduction can be made without assuming the function to be continuous. By VI, with $\phi = \psi$,

$$f(\phi) = \tfrac{1}{2} f(2\phi).$$

Thus 16.52 holds when $\phi = \frac{1}{2}\pi$, again when $\phi = \frac{1}{4}\pi$, and so on; that is, it holds when $\phi$ is $\pi$ divided by any power of 2. Appealing again to VI, we deduce that $f(\phi) = \mu\phi$ whenever $\phi = n\pi$, where $n$ is a number which terminates when expressed as a "decimal" in the scale of 2 [cf. Coxeter **3,** p. 102]. For brevity, let us call this a *binary* number.

Suppose, if possible, that, for some particular value of $\phi$, $f(\phi) \neq \mu\phi$. Choose a binary number $n$ between the two distinct real numbers $\phi/\pi$ and $f(\phi)/\mu\pi$. If $f(\phi) > \mu\phi$, so that

$$\phi < n\pi < \frac{f(\phi)}{\mu},$$

we have, since $f(\phi)$ is an increasing function,

$$f(\phi) < f(n\pi) = \mu n\pi < f(\phi),$$

which is absurd. If, on the other hand, $f(\phi) < \mu\phi$, we can argue the same way with all the inequalities reversed. Hence, in fact, $f(\phi) = \mu\phi$ for all the values of $\phi$ (from 0 to $\pi$).

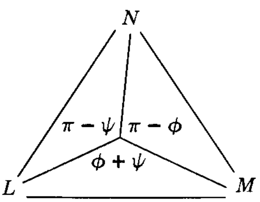



**Figure 16.5c**

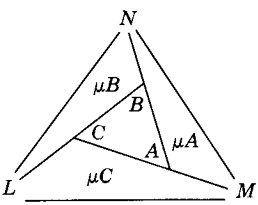



**Figure 16.5d**

VII. *The area* $\Delta$ *of any triangle ABC (with finite sides) is a constant multiple of its angular defect:*

$$\Delta = \mu(\pi - A - B - C).$$

For this final step, Gauss exhibited $ABC$ as part of a trebly asymptotic triangle by extending its sides in cyclic order, as in Figure 16.5*d*. The remaining parts are doubly asymptotic triangles whose areas are $\mu A$, $\mu B$, $\mu C$. Hence

$$\Delta + \mu A + \mu B + \mu C = t = \mu\pi,$$

and the desired formula follows at once.

If we wish, we can follow Lobachevsky in using such a unit of measurement* that the area of a trebly asymptotic triangle is $\pi$. Then $\mu = 1$, and the formula is simply

**16.53** $$\Delta = \pi - A - B - C.$$

This is strikingly reminiscent of the formula 6.92, which tells us that the area of a spherical triangle drawn on a sphere of radius $R$ is

$$(A + B + C - \pi)R^2.$$

In fact, setting $R^2 = -1$, we find that Gauss's result agrees formally with the area of a triangle drawn on a sphere of radius $i$. Long before the time of Gauss, it was suggested by J. H. Lambert (1728–1777) that, if a non-Euclidean plane exists, it should resemble a sphere of radius $i$. This analogy enabled him to derive the formulas of hyperbolic trigonometry (which were later developed rigorously by Lobachevsky) from the classical formulas of spherical trigonometry. Its full significance did not appear till Minkowski (1864–1909) discovered the geometry of space-time, which provided a geometrical basis for Einstein's special theory of relativity. We know now that, in a (2 + 1)-dimensional space-time, the hyperbolic plane can be represented without distortion on either sheet of a *sphere of time-like radius.* In the underlying affine space, this kind of sphere is a hyperboloid of two sheets.†

## EXERCISES

**1.** Gauss's formula 16.53 remains valid when the triangle has one or more zero angles.

**2.** The area of any simple $p$-gon is equal to its angular defect: the amount by which its angle sum falls short of that of a $p$-gon in the Euclidean plane. (*Hint:* Dissect the polygon into triangles. Of course, we are now assuming $\mu = 1$.) In Figure 16.4*a*, the area of $ABM$ is equal to that of $ABGF_1A_1$.

**3.** The product of three translations along the directed sides of a triangle (through the lengths of these sides themselves) is a rotation through the angular defect of the triangle. (These translations are half as long as those in Donkin's theorem, 15.31.) [Lamb **1**, p. 7.]

**4.** The product of half-turns about the midpoints of the sides of a simple quadrangle (in their natural order) is a rotation through the angular defect of the quadrangle.

**5.** Any polygon whose angle sum is a submultiple of $2\pi$ can be repeated, by half-

* Coxeter, Hyperbolic triangles, *Scripta Mathematica,* **22** (1956), p. 9.

† Coxeter, A geometrical background for de Sitter's world, *American Mathematical Monthly,* **50** (1943), p. 220.

turns about the midpoints of its sides, so as to cover the whole plane without interstices [cf. Somerville **1**, p. 86, Ex. 15]. (*Hint:* See Figures 4.2*b* and *c*.)

## 16.6 CIRCLES, HOROCYCLES, AND EQUIDISTANT CURVES

*A circle is the orthogonal trajectory of a pencil of lines with a real vertex . . . A horacycle is the orthogonal trajectory of a pencil of parallel lines. . . . The orthogonal trajectory of a pencil of lines with an ideal vertex . . . is called an equidistant-curve.*

D. M. Y. Sommerville (1879–1934)
[Sommerville **1**, pp. 51–52]

By 15.26, any two distinct lines are either intersecting, parallel, or ultraparallel. In other words, they belong to a *pencil* of lines of one of three kinds: an ordinary pencil, consisting of all the lines through one point, a pencil of parallels, consisting of all the lines parallel to a given ray, or a *pencil of ultraparallels,* consisting of all the lines perpendicular to a given line. By 15.32, the product of reflections in the two lines is a rotation, a parallel displacement, or a translation, respectively. By fixing one of the two lines and allowing the other to vary in the pencil, we see that each of these three kinds of direct isometry can be applied as a continuous motion.

A *circle* with center $O$ may be defined either as in § 15.1 or to be the locus of a point $P$ which is derived from a fixed point $Q$ (distinct from $O$) by continuous rotation about $O$, or to be the locus of the image of $Q$ by reflection in all the lines through $O$. When the radius $OQ$ becomes infinite, we have a *horocycle* with center $M$ (at infinity): the locus of a point which is derived from a fixed point $Q$ by a continuous parallel displacement, or the locus of the image of $Q$ by reflection in all the lines parallel to the ray $QM$ [Coxeter **3**, p. 213]. The rays parallel to $QM$ are called the *diameters* of the horocycle. The first "o" in the word "horocycle" is short, as in "horror."

The locus of a point at a constant distance from a fixed line $o$ is not a pair of parallel lines, as it would be in the Euclidean plane, but an *equidistant curve* (or "hypercycle"), having two branches, one on each side of its *axis* $o$. Either branch may be described as the locus of a point which is derived from a fixed point $Q$ (not on $o$) by continuous translation along $o$, or as the locus of the image of $Q$ by reflection in all the lines perpendicular to $o$.

Orthogonal to the pencil of lines through $O$ we have a pencil of concentric circles. A rotation about $O$ permutes the lines and slides each circle along itself. Orthogonal to the pencil of parallels with a common end $M$ we have a pencil of *concentric horocycles.* A parallel displacement with center $M$ permutes the parallel lines and slides each horocycle along itself. Orthogonal to the pencil of ultraparallels perpendicular to $o$ we have a pencil of *coaxal equidistant curves.* A translation along $o$ permutes the ultraparallel lines and slides each equidistant curve along itself.

We are now ready to fulfill the promise, made after 15.31, to show that "the product of two translations with nonintersecting axes may be a rotation." Referring to Figure 16.3*d*, we see that the line through *C* perpendicular to *AB* is ultraparallel to *AM* and *AN*. Therefore, it has a common perpendicular *GH* with *AM*, and a common perpendicular *FE* with *AN*, forming a pentagon *AEFGH* with right angles at *E*, *F*, *G*, *H* as in Figure 16.6*a*. The remaining angle (at *A*) may be as small as we please; if it is zero, the pentagon is "asymptotic." The product of reflections in *AE* and *FG* is a translation along *EF* (through 2*EF*). The product of reflections in *FG* and *AH* is a translation along *GH* (through 2*GH*). Hence the product of these two translations is the same as the product of reflections in *AE* and *AH*, which is a rotation or, if *A* is an "end," a parallel displacement. Since the axes of the two translations are both perpendicular to *FG*, we have thus proved that the product of translations along two ultraparallel lines may be either a rotation or a parallel displacement. (Of course, it may just as easily be another translation.)

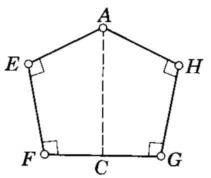



**Figure 16.6a**

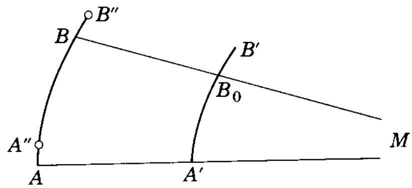



**Figure 16.6b**

The product of translations along two parallel lines, *AM* and *BM*, leaves invariant the common end *M*; therefore it cannot be a rotation, but must be either a translation along another line through *M* or a parallel displacement with center *M*. We shall soon see that the latter possibility arises when the two given translations are of equal length, one towards *M* and the other away from *M*. In fact, the translation along *AM* from *A* to *A'* (Figure 16.6*b*) transforms the arc *AB* of a horocycle through *A* into an equal arc *A'B'* of the concentric horocycle through *A'*. Let $B_0$ denote the point in which the latter arc is cut by the diameter through *B*. The translation along this diameter from $B_0$ to *B* transforms the arc $B_0A'$ of the second horocycle into the equal arc *BA''* of the first. Thus the product of the two translations is the parallel displacement that transforms the arc *AB* into *A''B''*; it slides this horocycle (and every concentric horocycle) along itself.

**EXERCISES**

**1.** The three vertices of a (finite) triangle all lie on each of three equidistant curves, whose axes join midpoints of pairs of sides, and on a fourth "cycle," which may be either a circle or a horocycle or another equidistant curve (with all three vertices on one branch). [Sommerville **1**, pp. 54, 189.]

**2.** The three sides of a (finite) triangle all touch a circle (the incircle) and three other "cycles," each of which may be of any one of the three kinds.

**3.** In Figure 15.2*a*, the horocycle through $J$ with diameter $p_1$ passes also through $L$.

**4.** How many horocycles will pass through two given points?

**5.** An equidistant curve may have as many as four intersections with a circle or a horocycle or another equidistant curve.

**6.** Develop the analogy between conics in the affine plane and generalized circles in the hyperbolic plane. A horocycle, like a parabola, goes to infinity in one direction: if the points $P$ and $Q$ on it are variable and fixed, respectively, the limiting position of the line $QP$ is the diameter through $Q$. An equidistant curve, like a hyperbola, has two branches.

**7.** Unlike the conjugate axis of a hyperbola, the axis of an equidistant curve is on the *concave* side of each branch.

## 16.7 POINCARÉ'S "HALF-PLANE" MODEL

*There is a gain in simplicity when the fundamental circle is taken as a straight line, say the axis of x. . . . We may avoid dealing with pairs of points by considering only those points above the x-axis. A proper circle is represented by a circle lying entirely above the x-axis; a horocycle by a circle touching the x-axis; an equidistant-curve by the upper part of a circle cutting the x-axis together with the reflexion of the part which lies below the axis.*

D. M. Y. Sommerville [**1**, pp. 188-189]

From the conformal model (Figure 16.2*a*) in which the lines are represented by circles (and lines) orthogonal to a fixed circle $\Omega$, Poincaré derived another conformal model by inversion in a circle whose center lies on $\Omega$. The inverse of $\Omega$ is a line, say a "horizontal" line, which we shall again denote by $\Omega$. The points of the hyperbolic plane are represented by pairs of points which are images of each other by reflection in $\Omega$, and the lines are represented by circles and lines orthogonal to $\Omega$, that is, circles whose centers lie on $\Omega$, and vertical lines [Burnside **1**, p. 387].

Through a pair of points which are images in $\Omega$, we can draw an intersecting pencil of coaxal circles (like Figure 6.5*a* turned through a right angle) representing an ordinary pencil of lines. The orthogonal nonintersecting pencil, having $\Omega$ for its radical axis, represents a pencil of concentric circles. The limiting points of the nonintersecting pencil represent the common center of the concentric circles.

Another pencil of circles (situated as in Figure 6.5*a* itself) can be drawn through two points on $\Omega$. One member of this pencil, having its center on $\Omega$, represents a line $o$. The remaining circles (or strictly, pairs of them related by reflection in Ω) represent coaxal equidistant curves with axis $o$. For, the orthogonal nonintersecting pencil represents the pencil of ultraparallel lines perpendicular to $o$.

A tangent pencil of circles whose centers lie on Ω (Figure 6.5$b$) represents a pencil of parallels, whereas the orthogonal tangent pencil (touching Ω) represents a pencil of concentric horocycles. One particular pencil of parallels (special in the model but, of course, not special in the hyperbolic geometry itself) is represented by all the vertical lines (which pass, like Ω itself, through the point at infinity of the inversive plane). The horocycles having these lines for diameters are represented by all the horizontal lines except Ω (or strictly, pairs of such lines related by reflection in Ω). Since reflections in the vertical lines represent reflections in the parallel lines, horizontal translations represent parallel displacements. Hence the horizontal lines (other than Ω itself) represent the horocycles *isometrically:* equal segments represent equal arcs.

**EXERCISES**

1. What figure is represented by two lines forn :ng an angle that is bisected by Ω?

2. When two ultraparallel lines are represented by nonintersecting circles (in either of Poincaré's conformal models), the distance between the lines, measured along their common perpendicular, appears as the *inversive distance* between the circles (see Exercise 5 of § 6.6).

3. The angle of parallelism (Figure 16.3$d$ on page 293) is

$$\Pi(x) = 2 \arctan e^{-x}.$$

## 16.8 THE HOROSPHERE AND THE EUCLIDEAN PLANE

*F. L. Wachter (1792-1817) . . . in a letter to Gauss (Dec., 1816) . . . speaks of the surface to which a sphere tends as its radius approaches infinity. . . . He affirms that even in the case of the Fifth Postulate being false, there would be a geometry on this surface identical with that of the ordinary plane.*

R. Bonola [**1**, *pp.* 62-63]

The ideas in §§ 16.6 and 16.7 extend in an obvious manner from two to three dimensions. The locus of images of a point $Q$ by reflection in all the planes through a point $O$ is a *sphere* with radius $OQ$. As a limiting case we have a *horosphere* with center $M$ (at infinity): the locus of images of a point $Q$ by reflection in all the planes parallel to the ray $QM$ [Coxeter **3**, p. 218]. The locus of images of a point $Q$ by reflection in all the planes perpendicular to a fixed plane $\omega$ is one sheet of an *equidistant surface,* which consists of points at a constant distance from $\omega$ on either side.

There is a conformal model in inversive space in which the points of hyperbolic space are represented by pairs of points related by reflection in a fixed "horizontal" plane $\Omega$, and the planes are represented by spheres and planes orthogonal to $\Omega$, that is, spheres whose centers lie on $\Omega$, and vertical planes. The representation of lines (which are intersections of planes) follows immediately. Of particular interest is the bundle of vertical lines, which represents the bundle of lines parallel to a given ray $QM$ (special in the model, though not in the hyperbolic geometry itself). The horospheres that have these lines for diameters are represented by all the horizontal planes except $\Omega$. Since every vertical plane provides a model (of the kind described in § 16.7) for a plane in the hyperbolic space, each horizontal plane (except $\Omega$) represents a horosphere, and every line in the plane represents a horocycle on the horosphere. Since distances along such lines agree with distances along the corresponding horocycles, the representation of the horosphere by the Euclidean plane is *isometric:* for any figure in the Euclidean plane there is a congruent figure on the horosphere (with lines replaced by horocycles).

This astonishing theorem was discovered independently by Bolyai and Lobachevsky. For two different proofs, see Coxeter [**3,** pp. 197, 251]. It means that, along with ordinary spherical geometry, the inhabitants of a hyperbolic world would also study horospherical geometry, which is the same as Euclidean geometry!