CHAPTER 1

Transformations in the Euclidean Plane

In this chapter no attempt will be made to construct a geometry from its foundations; this will be left to later chapters. Our aim in the first chapter is the determination of the objectives in our study of geometry. We can best acquire such an orientation by having a close and well-planned look at a kind of geometry which is familiar to us. The euclidean plane serves this purpose because every student has studied it and knows a great deal about its properties. In this chapter we utilize this knowledge of the theorems in plane euclidean geometry as well as facts like the congruence theorems of triangles, Pythagoras' theorem, and others, without inquiring into their sources or their mutual relations. Another tool which will be freely used in this chapter is the analytic geometry of the plane. The student will have discovered that in many instances the methods provided by analytic geometry are easier to apply and less cumbersome than the "synthetic" treatment of geometrical problems which he had to apply in high school. Thus, again without attempting in this chapter to provide the theoretical foundations for analytic geometry, we will use it in a special form, in the plane of complex numbers. To each of the points in the plane a complex number will be assigned. Then the important transformations of the plane will be studied by observing their analytical meaning. By doing so we will be able to introduce and to become acquainted with concepts like isometry, similitude, rotation, and reflection. The algebraic notion of the group will help us in the organization of the details. In the last part of the chapter we will encounter the first example of a noneuclidean geometry.

1–1 THE PLANE OF COMPLEX NUMBERS

In this section we review some well-known properties of the complex numbers without, however, giving proofs for our statements. The proofs are readily available in books on elementary algebra.

An ordered couple of real numbers a and b is called a *complex number*, and the usual notation for it is $a + bi$. Two complex numbers $a + bi$ and $c + di$ are considered equal if and only if $a = c$ and $b = d$. Sometimes it is convenient to distinguish between the *real part* of $z = a + bi$, $\mathrm{Re}\,(z) = a$, and its *imaginary part*, $\mathrm{Im}\,(z) = b$. The sum and the product of complex numbers are

defined, respectively, by $(a + bi) + (c + di) = (a + c) + (b + d)i$ and $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$. This implies, for instance, $(0 + 1i)(0 + 1i) = -1 + 0i$. If we write $0 + bi = bi$ and identify the complex number $a + 0i$ with the real number a, then the last equation becomes $i^2 = -1$. The arithmetic of complex numbers follows the familiar laws. If Greek letters denote complex numbers, we have $\alpha + \beta = \beta + \alpha$, $\alpha\beta = \beta\alpha$, $(\alpha + \beta) + \delta = \alpha + (\beta + \delta)$, $(\alpha\beta)\delta = \alpha(\beta\delta)$, $\alpha(\beta + \delta) = \alpha\beta + \alpha\delta$, $1 \cdot \alpha = \alpha$, $0 + \alpha = \alpha$, $0 \cdot \alpha = 0$.

We introduce a plane cartesian coordinate system and represent the complex number $a + bi$ by a point with coordinates $x = a$, $y = b$. This representation is a one-to-one correspondence of all the complex numbers and all the points in the plane. In particular, the real numbers are represented by the points on the x-axis, and the imaginary numbers $0 + bi = bi$ by the points on the y-axis. Therefore we call this plane the *plane of complex numbers*, the x-axis the *real axis*, and the y-axis the *imaginary axis*. Because of the correspondence between the complex numbers and the points of the plane we will not have to distinguish between the concepts "point" and "number."

If the cartesian coordinate system is replaced by a polar coordinate system (Fig. 1–1) having the same origin and the real axis as its axis, then the complex number $a + bi$ will be represented by a point whose polar coordinates are ρ and ϕ. The cartesian and the polar coordinates are then connected by the relations $a = \rho \cos \phi$, $b = \rho \sin \phi$, and $\rho = \sqrt{a^2 + b^2}$. Therefore, $a + bi = \rho(\cos \phi + i \sin \phi)$. If we restrict the values of ρ and ϕ by $\rho > 0$ and $-\pi \leqq \phi < \pi$, then every nonzero complex number $a + bi$ determines a unique $\rho = \sqrt{a^2 + b^2}$ and a unique

$$\phi = \arccos \frac{a}{\sqrt{a^2 + b^2}} = \arcsin \frac{b}{\sqrt{a^2 + b^2}} .$$

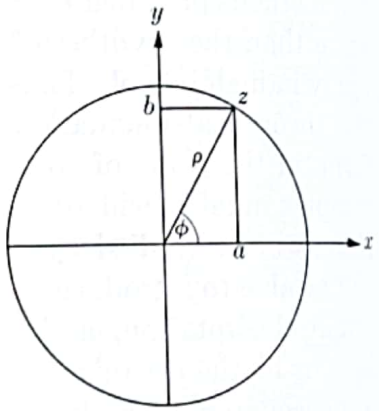

Figure 1-1

[The ambiguity of the first equality is removed by considering the second. For instance, for $a = -1$, $b = 1$, $\phi = \arccos(-1/\sqrt{2})$ could be $\frac{3}{4}\pi$ or $-\frac{3}{4}\pi$. However, $\phi = \arcsin(1/\sqrt{2})$ is $\pi/4$ or $\frac{3}{4}\pi$, which leaves only $\phi = \frac{3}{4}\pi$.]

The directed segment $(0, z)$ is called the *position vector* of the complex number z.

The value $\rho > 0$ of a nonzero complex number $\rho(\cos \phi + i \sin \phi)$ is called its *absolute value*, and ϕ its *argument*. The absolute value of the number 0 is 0, while its argument is indeterminate. The notation for the absolute value of z is $|z|$, and for its argument is $\arg z$.

If $z = a + bi$ is a complex number, $a - bi$ is called its *conjugate* and is denoted by $\bar{z}$. The points z and $\bar{z}$ are symmetric with respect to the real axis. The product

$$z\bar{z} = (a + bi)(a - bi) = a^2 + b^2 = |z|^2$$

is the square of the absolute value. It is obvious that $|z| = |\bar{z}|$.

The multiplication of complex numbers is considerably facilitated by *De Moivre's theorem* (after the French, later British, mathematician Abraham de Moivre, 1667–1754). Let two complex numbers, z_1 and z_2, be given by their absolute values, ρ_1 and ρ_2, and their arguments, ϕ_1 and ϕ_2. Then

$$z_1 z_2 = \rho_1 \rho_2 [\cos(\phi_1 + \phi_2) + i \sin(\phi_1 + \phi_2)].$$

In words, the absolute value of the product is the product of the absolute values, while the argument of the product is the sum of the arguments. In particular, this implies the relation $|z_1 z_2| = |z_1|\,|z_2|$.

The circle with radius r about the origin has the equation $\rho = r$. In the special case of $r = 1$ this equation represents the so-called *unit circle.*

The length of the segment (z_1, z_2), that is, the distance d between the two points $z_1 = a_1 + b_1 i$ and $z_2 = a_2 + b_2 i$, can be determined by using Pythagoras' theorem,

$$d = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2},$$

and thus we obtain $d = |z_1 - z_2|$. In particular, the length of the position vector of z is $|z|$.

In calculus it is shown that $e^{i\phi} = \cos\phi + i \sin\phi$ for every real ϕ. Hence every complex number with absolute value ρ and argument ϕ can be written as $\rho e^{i\phi}$.

1–2 ISOMETRIES IN THE COMPLEX NUMBER PLANE

Transformations. Let S and T be two sets of points or numbers. We consider ordered pairs (s, t) such that $s \in S$ and $t \in T$. If in a set α of such ordered pairs each element of S appears exactly once as the leading term of a pair, then α is called a *mapping of S into T* or a *function* defined on S with values in T, and the notation is $\alpha\colon S \to T$. If $(s, t) \in \alpha$, we write $(s)\alpha = t$ or, if no ambiguity results, $s\alpha = t$. The element $s\alpha$ is called the *image* of s under α, and s is the *preimage* of $s\alpha$. S is the *domain* of α, and the set of all $s\alpha$ with $s \in S$ is called the *range* of α and is denoted by $S\alpha$. If none of the $s\alpha$ appear more than once in $S\alpha$, that is, if the equation $s_1\alpha = s_2\alpha$ implies that $s_1 = s_2$, then α is *one-to-one* or *injective.* If $S\alpha = T$, that is, if all elements of T are images under α, then α is called *surjective*, or we say that α maps S *onto* T. If a mapping is injective and surjective, it is said to be *bijective.*

One of the most important concepts in geometry is that of a *transformation.* Henceforth by this we will mean a bijective mapping of a set on itself (in contrast to other books in which more general transformations are treated). Therefore, if $\alpha\colon S \to T$ is a transformation, then $S = S\alpha = T$.

Neither of the requirements "injective" or "surjective" is redundant. For instance, the mapping which carries every integer x into the integer $2x$ is injective because for every x there exists one and only one $2x$. However, it is not a transformation in the sense used by us, because its domain is the set Z of all integers, while its range comprises the even integers only. It is "into," not

"onto." On the other hand, the mapping of Z that takes the even integers y into $y/2$ while the odd integers are left unchanged is surjective, since the domain and the range are both Z. However, this mapping is not injective, for instance $2 \to 1$ and $1 \to 1$; hence it is not a transformation either.

Transformations can, of course, be very general. The transformations of interest for geometry will always be restricted by the additional requirement that some property be *invariant*, that is, preserved. Thus we might, for instance, deal with transformations that preserve angles, distances, straightness of lines, areas, or volumes.

We shall write $\alpha\beta$ for the transformation which results from performing first α and then β, provided the domains and ranges of α and β all coincide. Thus, in contrast with other usages, we shall always be able to read products from left to right. The notation $(P)\alpha\beta$ makes this possible because $(P)\alpha\beta$ means $[(P)\alpha]\beta$. If α is a transformation, α^{-1} will be the *inverse transformation*, which is defined by $P\alpha^{-1} = Q$ if $Q\alpha = P$.

Isometries. The transformations of the plane which we intend to study now are distance-preserving. They are called congruences, congruent transformations, motions, or, as we will refer to them, *isometries*.

Definition. An *isometry* of the plane is a transformation α with the property that for any two points P and Q, the distance $|PQ|$ is equal to the distance $|P\alpha Q\alpha|$.

In the following, geometric transformations will be denoted principally by capital letters.

Theorem 1–2.1. Let a, b, c, d be complex numbers and $|a| = |c| = 1$. Then the mappings $M\colon z \to az + b$ and $K\colon z \to c\bar{z} + d$ are isometries of the plane.

Proof. It is obvious that K and M are single-valued mappings. In order to show that they are bijective, we start from $zM = az + b$ and obtain $z = (zM - b)/a$, which is always uniquely defined because $a \neq 0$ in view of $|a| = 1$. Thus, for each complex number zM there is exactly one z, namely $(zM - b)/a$, which is mapped on zM. Thus M is bijective. The mapping K also has a unique inverse by which $z \to (\bar{z} - \bar{d})/\bar{c}$. Hence K is also bijective. Now we have to show that distances remain unchanged. If z and w are two points, then

$$\begin{aligned} |zM - wM| &= |az + b - aw - b| = |a(z - w)| \\ &= |a|\,|z - w| = |z - w|, \end{aligned}$$

in view of $|a| = 1$. Also,

$$|zK - wK| = |c\bar{z} + d - c\bar{w} - d| = |c|\,|\bar{z} - \bar{w}| = \overline{|z - w|} = |z - w|.$$

Theorem 1–2.2. Given points z_0, z_1, w_0, and w_1, with $|z_1 - z_0| = |w_1 - w_0| \neq 0$, there exist exactly two transformations of the types $z \to az + b$ and $z \to c\bar{z} + d$, with $|a| = |c| = 1$, which map $z_0 \to w_0$ and $z_1 \to w_1$.

Proof. We have to determine a and b such that

$$az_0 + b = w_0, \qquad az_1 + b = w_1, \qquad |a| = 1.$$

Subtracting the two equations yields

$$w_1 - w_0 = a(z_1 - z_0), \qquad a = \frac{w_1 - w_0}{z_1 - z_0},$$

which is well defined, in view of $|z_1 - z_0| \neq 0$. Also,

$$|a| = \left|\frac{w_1 - w_0}{z_1 - z_0}\right| = \frac{|w_1 - w_0|}{|z_1 - z_0|} = 1.$$

For b we obtain by substitution

$$b = w_0 - az_0 = w_0 - z_0 \frac{w_1 - w_0}{z_1 - z_0}.$$

Similarly, $w_0 = c\bar{z}_0 + d$ and $w_1 = c\bar{z}_1 + d$ imply that $w_1 - w_0 = c(\bar{z}_1 - \bar{z}_0)$ and $c = (w_1 - w_0)/(\bar{z}_1 - \bar{z}_0)$, where

$$|c| = \left|\frac{w_1 - w_0}{\bar{z}_1 - \bar{z}_0}\right| = \frac{|w_1 - w_0|}{|z_1 - z_0|} = 1.$$

In order to be able to investigate the isometries of the plane we prove first that all isometries are collineations. A *collineation* is a transformation of the points of the plane which carries straight lines into straight lines.

Theorem 1–2.3. Every isometry of the number plane is a collineation.

Proof. If z, w, and t are three distinct points on a straight line, one of them, say w, lies between the two others. Then

$$|w - z| + |t - w| = |z - t|,$$

and if M is an isometry, then the images of z, w, and t behave in the same way, namely

$$|wM - zM| + |tM - wM| = |zM - tM|.$$

But this means that wM, zM, and tM are collinear, because otherwise they would form a triangle in which the sum of two sides is greater than the third side, that is,

$$|wM - zM| + |tM - wM| > |zM - tM|.$$

Corollary. Isometries map parallel lines on parallel lines.

Theorem 1–2.4. An isometry is uniquely determined by a triangle and its congruent image.

Proof (Fig. 1–2). Let z_0, z_1, and z_2 be the vertices of the given triangle, and let w_0, w_1, and w_2 be those of its image, with

$$|z_k - z_l| = |w_k - w_l|, \qquad k, l = 0, 1, 2.$$

Let z be a point distinct from the z_k's. The parallel to the line (z_0, z_1) through z meets the line (z_0, z_2) in a point z'', and the parallel to (z_0, z_2) through z meets (z_0, z_1) in a point z'. If w is the image of z under an isometry which also carries the z_k's into the w_k's, then, according to Theorem 1–2.3 and the Corollary, the lines (z, z') and (z, z'') will be mapped on corresponding lines (w, w') and (w, w''), with w' on (w_0, w_1) and w'' on (w_0, w_2). Then

$$|z_0 - z'| = |w_0 - w'|, \qquad |z_1 - z'| = |w_1 - w'|, \qquad |z_0 - z''| = |w_0 - w''|$$

and $|z_2 - z''| = |w_2 - w''|$, which determines the position of w' and w'' uniquely. But then also the position of w is uniquely defined. Since z was arbitrarily chosen, this proves the theorem.

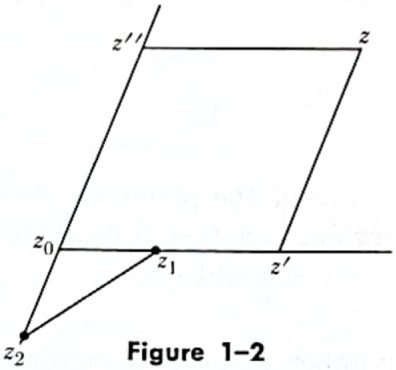

Figure 1–2

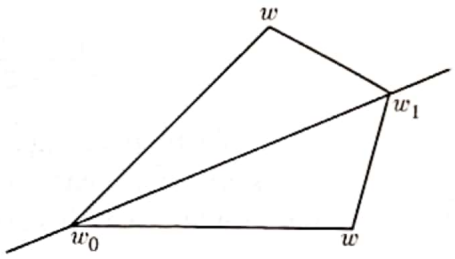

Figure 1–3

Theorem 1–2.5. There are exactly two isometries mapping two given points z_0 and z_1 onto two given points w_0 and w_1, provided

$$|z_0 - z_1| = |w_0 - w_1| \neq 0. \tag{1}$$

Proof (Fig. 1–3). In view of Theorems 1–2.1 and 1–2.2, there are at least two such isometries. We have only to show that no more such isometries exist. Now, any isometry mapping z_0 on w_0 and z_1 on w_1 is a collineation, and therefore maps the straight line through z_0 and z_1 onto the straight line through w_0 and w_1. Every point z is mapped on a point w so that

$$|z - z_0| = |w - w_0|, \qquad |z - z_1| = |w - w_1|. \tag{2}$$

If z, z_0, and z_1 are collinear, so are w, w_0, and w_1, in view of Theorem 1–2.3. Then Eqs. (1) and (2) determine uniquely the position of w on the line through

w_0 and w_1. If z is not collinear with z_0 and z_1, then z_0, z_1, and z form a triangle. The triangle with vertices w_0, w_1, and w is congruent to this triangle. The vertices w_0 and w_1 being fixed, there are two possible positions for the triangle, with w lying symmetrically on either side of the line through w_0 and w_1. Thus every point z has one or two images w, according to whether z is collinear or noncollinear with z_0 and z_1. But, by Theorem 1–2.4, each of these triples, w_0, w_1, and w, determines an isometry, and our theorem is proved.

We may summarize the results of this section as follows.

Theorem 1–2.6. The set $\mathfrak{M}$ of all isometries of the number plane is composed of two classes $\mathfrak{M}_+$ and $\mathfrak{M}_-$. The class $\mathfrak{M}_+$ consists of all isometries of the form $z \to az + b$ ($|a| = 1$), and the class $\mathfrak{M}_-$ of all isometries of the form $z \to c\bar{z} + d$ ($|c| = 1$).

The isometries of $\mathfrak{M}_+$ are called *direct*, those of $\mathfrak{M}_-$ *opposite*.

Coordinate transformations. Another concept closely related to the isometries is that of the *coordinate transformations*. By this we mean the introduction of another cartesian coordinate system for assigning values to the points of the plane.

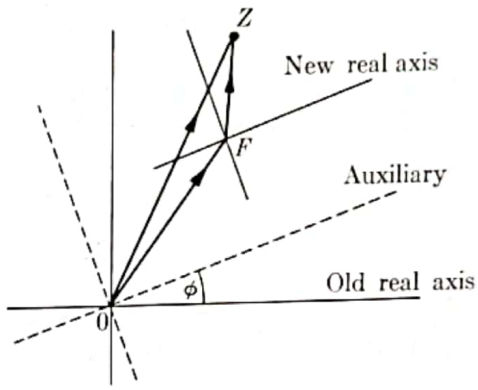

Figure 1–4

Let the origin of the new system be at the point F (Fig. 1–4) whose complex value is f, and let its positive real axis be obtained from the old positive real axis by spinning it through an angle ϕ. Assume also that the positive imaginary axis transforms into the new positive imaginary axis by a rotation through the same angle ϕ. First consider an auxiliary system with origin 0, whose axes are parallel to those of the new system.

If a subscript 1 designates values in the auxiliary system, a subscript 2 those in the new system, and a subscript 0 the old value, let Z be an arbitrary point with $Z_0 = z$. In the auxiliary system all arguments diminish by ϕ, and thus, according to De Moivre's theorem,

$$Z_1 = z[\cos(-\phi) + i \sin(-\phi)] = z(\cos\phi - i\sin\phi),$$

and $F_1 = f(\cos\phi - i\sin\phi)$. The value Z_2 can be obtained by subtraction of F_1 from Z_1 such that

$$Z_2 = Z_1 - F_1 = z(\cos\phi - i\sin\phi) - f(\cos\phi - i\sin\phi) = az + b,$$

say, where a and b are uniquely determined by the choice of f and ϕ, and $|a| = 1$. Conversely, it can be shown (Exercise 9) that for any given a and b with $|a| = 1$

there exists a unique coordinate transformation such that $Z_2 = az + b$ or, for short, $z \to az + b$.

Another case arises if ϕ is again the angle that spins the positive real axis into its new position, but if the positive imaginary axis is turned by this angle into the new *negative* imaginary axis. Then it can be shown (Exercise 10) that the formula becomes $z \to a\bar{z} + b$, $|a| = 1$.

Thus the same formulas $z \to az + b$ and $z \to a\bar{z} + b$ applied to two different operations: to isometries where the points were moved and the coordinate system remained unchanged, and to coordinate transformations where the same points were observed from a new system.

EXERCISES

1. Are the following mappings transformations of the complex number plane?

   (a) $z \to z^3$ (b) $x + yi \to x^3 + y^3 i$

   (c) $x + yi \to x^2 + y^2 i$

2. Are the following mappings transformations of the number plane? Are they collineations? Find the image and the preimage of i in each of them.

   (a) $z \to |z|$ (b) $z \to \bar{z}$

   (c) $z \to z\bar{z}$ (d) $z \to \mathrm{Re}\,(z)$

3. Show that $z \to z^2$ is not a transformation of the whole number plane. Is this mapping distance-preserving? Are straight lines carried into straight lines?

4. Show that T: $z \to 1/z$ is a transformation of the number plane with the point 0 removed. Find the image under T of

   (a) the real axis,
   (b) the imaginary axis,
   (c) the line containing all points with real part 1,
   (d) the circle about 0 with radius r,
   (e) the angle bisector of the first quadrant.

   What are the preimages of all these?

5. Are the transformations $z \to 2iz$ and $z \to 2i\bar{z}$ isometries? Are they angle-preserving (conformal)? Note that an angle α and an angle $-\alpha$ are not to be considered as equal.

6. Find the isometries of the number plane taking

   (a) i into -1, 2 into $2i$, (b) 1 into 2, $1 + i$ into 1.

   Why are the two isometries for (b) related to each other in such a simple way?

7. Are the following isometries direct or opposite?

   (a) $M\colon 0 \to 0,\ 1 \to i$, and $(i)M > 0$ (b) M^{-1}
   (c) $N\colon 1 \to 1 + i,\ i \to 2i,\ -1 \to 0$

8. In the proof of Theorem 1–2.5, a geometric argument was used to show that there are two possible positions for the point w, if z_1, z_0, and z are not collinear. Replace this argument by an algebraic reasoning.

9. Prove that for every a and b with $|a| = 1$ there exist unique coordinate transformations $z \to az + b$ and $z \to a\bar{z} + b$.

10. Show that the second type of coordinate transformation described on p. 8 indeed yields a formula $z \to a\bar{z} + b$, $|a| = 1$.

11. Find the transformation if a new system has its origin at $\sqrt{3} - i$, if the new positive real axis passes through 0, and if the new positive imaginary axis intersects the old positive real axis.

12. Describe the coordinate transformations

   (a) $z \to \bar{z}$ (b) $z \to i\bar{z}$ (c) $z \to -iz + 1$.

1–3 GROUPS

Basic concepts. In order to continue our work on isometries we shall need the algebraic concept of a group.

A set S is said to be *closed under a binary operation* $(*)$ if every ordered pair of elements a and b in S determines a unique element of S, which will be called $a * b$. Not every set is closed under every binary operation. For instance, let T be the set of all integers <10 and let the operation $(*)$ be addition. The sum $c = a + b$ of two elements of T might exceed 10 and therefore does not necessarily belong to T. We say then that T is not closed under addition. On the other hand, the set of all integers is closed under addition, and it is also closed under multiplication. Another example is that the set of all points in the plane is closed under the operation of "finding the midpoint."

A set S provided with a binary operation $(*)$ is called a *group* $(S, *)$ if it satisfies the following four postulates.

Gp 1. S is closed under $(*)$.
Gp 2. For all elements a, b, c in S, $(a * b) * c = a * (b * c)$.
This property is called *associativity*.
Gp 3. S contains an element e which satisfies $e * a = a * e = a$, for all a in S.
We call e the *identity* element of the group.
Gp 4. For each element a of S there exists a unique element a' of S such that $a * a' = a' * a = e$.
We call a' the *inverse* of a.

If $a * b$ is replaced by $a + b$ or ab, then the group will be called *additive* or *multiplicative*, respectively, and the identity element e will then be replaced by the zero element 0 or the unity element 1, respectively. The inverse a' will become $-a$ in the additive case and a^{-1} in the multiplicative case. In order to make our notation as short as possible, we will in general write our groups multiplicatively.

If there is no ambiguity as to the operation used, we will write "group S" instead of "the group $(S, *)$."

Examples of groups will be found in the exercises. Here we mention only that the complex numbers form a group under addition, while the nonzero complex numbers form a multiplicative group.

The set S over which the group is defined cannot be the null set because of Gp 3. If S is finite, the number of its elements is called the *order* of the group. It is important to note that ab is not necessarily the same element as ba. If, for all elements of S, $ab = ba$, we call the group *commutative* or *abelian* (in honor of the Norwegian mathematician N. H. Abel, 1802–1829). The property characterizing such a group is referred to as *commutativity*.

The postulate Gp 2 makes the parentheses in $(ab)c$ and $a(bc)$ unnecessary; we may write abc without any ambiguity. It can be proved by induction that in products of more than three elements the parentheses can also be ignored. Consequently, powers of elements can be defined, such as $a^2 = aa$, $a^3 = aaa$, $a^{-2} = a^{-1}a^{-1}$, and so forth.

A few rules for groups follow, with a, b, and c in each case standing for arbitrary group elements.

Theorem 1-3.1. The inverse of a^{-1} is a.

Proof. $a^{-1}a = aa^{-1} = 1$, by Gp 4.

Theorem 1-3.2. $(ab)^{-1} = b^{-1}a^{-1}$.

Proof. $abb^{-1}a^{-1} = 1$.

Corollary. $(a_1a_2 \ldots a_n)^{-1} = a_n^{-1} \ldots a_2^{-1}a_1^{-1}$. Proof by induction.

Theorem 1-3.3. Each of $ac = bc$ and $ca = cb$ implies $a = b$.

Proof. From $ac = bc$ we obtain by right multiplication with c^{-1}, $acc^{-1} = bcc^{-1}$, and hence $a = b$. The other result follows by left multiplication with c^{-1}.

Subgroups and cosets. We define a *subgroup* of a group G as a group whose elements are elements of G and whose operation is the same as that of G. Evidently, a subgroup of a subgroup of G is again a subgroup of G.

Theorem 1-3.4. A nonempty subset H of a group G is a subgroup of G if and only if for every two elements a and b of H, (i) b^{-1} and ab are in H, or (ii) ab^{-1} is in H.

Proof. Suppose that H is a subgroup, containing a and b. Then, by Gp 4, b^{-1} is in H, and, by Gp 1, ab and ab^{-1} are as well. For the converse, in the case (ii), suppose that $ab^{-1} \in H$ whenever a and $b \in H$. From $a \in H$ we obtain $aa^{-1} = 1 \in H$. Now, 1 and $b \in H$ imply $1b^{-1} = b^{-1} \in H$, and a and $b^{-1} \in H$ imply $a(b^{-1})^{-1} = ab \in H$. Thus Gp 3, Gp 4, and Gp 1 are satisfied for H. Since the operation is the same as in G, Gp 2 is fulfilled. Hence H is a group. When the conditions (i) hold, then $ab^{-1} \in H$, and H is a subgroup.

Definition. In a group G with a subgroup H the set C of elements ha (h arbitrary in H, a fixed in G) is denoted Ha and called a *right coset* with respect to H. Correspondingly, aH is a *left coset.*

Every element of G belongs to a coset with respect to H because obviously $b \in Hb$ and $b \in bH$. A further consequence of this fact is the following theorem.

Theorem 1–3.5. If the cosets Ha and Hb have an element in common, they coincide.

Proof. Let c be in Ha and Hb. Then there are elements h and k in H such that $c = ha$ and $c = kb$. This implies $ha = kb$, $a = h^{-1}kb$. Now $h^{-1}k \in H$, and therefore $a \in Hb$ and, with l an arbitrary element of H, $la \in Hb$. This implies $Ha \subseteq Hb$. On the other hand, $b = k^{-1}ha$, which results in $Hb \subseteq Ha$. Hence $Ha = Hb$.

Theorem 1–3.6. If the number of left or right cosets with respect to a subgroup is finite, then there are as many left as right cosets.

Proof. There is a one-to-one correspondence between left and right cosets because $(aH)^{-1} = H^{-1}a^{-1} = Ha^{-1}$; that is, the inverses of a left coset form a right coset.

Definition. The number of cosets with respect to H is the *index* of H in G.

It should be remarked that Theorem 1–3.6 in its present formulation does not hold for infinitely many cosets. However, if "cardinality" is substituted for "number," the theorem becomes valid also for this case, and the proof remains correct without any change.

Corollary. The cosets of a group G with respect to a subgroup H are disjoint and exhaust G.

Theorem 1–3.7. Two elements a and b of G belong to the same coset with respect to H if and only if $ab^{-1} \in H$.

Proof. $ab^{-1} \in H$ implies $a \in Hb$ and therefore $Ha = Hb$. Conversely, if $Ha = Hb$, then $a \in Hb$ and $ab^{-1} \in H$.

As an example let us mention that $\mathfrak{M}_-$, the set of all opposite isometries, is a coset of $\mathfrak{M}$ with respect to $\mathfrak{M}_+$. A detailed proof of this statement will be given later.

Equivalence. At this point, a very useful abstract concept will be introduced, that of an *equivalence relation*.

In a set S, let a relation $(\sim)$ between elements be given such that for all x, y, and z in S,

Ev 1. $x \sim y$ implies $y \sim x$ (*symmetry*).
Ev 2. $x \sim x$ (*reflexivity*).
Ev 3. If $x \sim y$ and $y \sim z$, then $x \sim z$ (*transitivity*).

Then this relation is an equivalence relation.

It can be shown that the existence of an equivalence relation in a set brings about the partition of the set into *equivalence classes* such that each set element belongs to one and only one class to each of whose elements it is related, whereas it is related to none of the elements of other classes.

If, in a group G with subgroup H, $x \sim y$ means $xy^{-1} \in H$, then the relation $(\sim)$ can be shown to satisfy the postulates Ev 1 through 3. The equivalence classes of G are then exactly the cosets with respect to H. The proof will be left for the exercises.

Conjugate and normal subgroups. If again H is a subgroup of a group G, the elements $g^{-1}hg$, with g in G and h running through H, form a subset of G that may be written $g^{-1}Hg$. Moreover the following result holds.

Theorem 1–3.8. For a given subgroup H of G and a given element g in G, $K = g^{-1}Hg$ is a subgroup.

Proof. Let h and k be in H. Then $g^{-1}kg$ and $g^{-1}hg$ are in K, and

$$g^{-1}kg(g^{-1}hg)^{-1} = g^{-1}kgg^{-1}h^{-1}g = g^{-1}kh^{-1}g \in K.$$

By Theorem 1–3.4, K is a subgroup.

We call $g^{-1}ag$ a *conjugate* of a and $g^{-1}Hg$ a *conjugate subgroup* of H. Important in particular are those subgroups which coincide with all their conjugates. Such subgroups are called self-conjugate, invariant, or, as we shall say, *normal subgroups*. If H is a normal subgroup of G, we write $H \lhd G$.

For a normal subgroup H we have, therefore, $H = g^{-1}Hg$ for all elements g, or $gH = Hg$. This means that the right and left cosets with respect to a normal subgroup H coincide. It does not mean, however, that for every element h of H, $gh = hg$ for every element g. On the contrary, in general we will have $gh = h'g$, where h' is an element of H other than h.

Every group G has two trivial normal subgroups, namely G itself and the group consisting of 1 alone. Indeed, $gG = G = Gg$ and $g \cdot 1 = g = 1 \cdot g$. In an abelian group all multiplication is commutative, and thus every subgroup is normal.

Another definition of a normal subgroup is equivalent to our definition although it seems weaker. By this definition H is a normal subgroup of G if, whenever $h \in H$, $g^{-1}hg$ is also in H for all $g \in G$. Now it follows that $g^{-1}Hg \subseteq H$ for all g in G, that is, also for g^{-1}. Hence $(g^{-1})^{-1}Hg^{-1} = gHg^{-1} \subseteq H$, and therefore $H \subseteq g^{-1}Hg$. But $H \subseteq g^{-1}Hg$ and $g^{-1}Hg \subseteq H$ imply the equality of H and $g^{-1}Hg$, and thus we have come back to our original definition. Conversely, it is obvious that our first definition had the weaker definition as a consequence.

Theorem 1–3.9. A subgroup H of index 2 in G is always normal.

Proof. There is only one left coset, gH, distinct from H, and only one such right coset. Since g is not in H, g is in the right coset which is, therefore, Hg. Thus $Hg = gH$, and H is normal in G.

Isomorphism. One further concept that is of great importance is that of *isomorphism*. Two groups G and $G\alpha$ are *isomorphic*, in symbols $G \cong G\alpha$, if there exists a bijective mapping α of $(G, \circ)$ on $(G, *)$ taking the elements 1, $a, b, \ldots$ of G into the elements $1\alpha, a\alpha, b\alpha, \ldots$ of $G\alpha$ such that

$$(a \circ b)\alpha = a\alpha * b\alpha,$$

for all elements a and b of G. The significance of this requirement is that multiplication is preserved under the isomorphism. It is easily verified that isomorphism is an equivalence relation, namely that (a) $G \cong G$, (b) $G \cong H$ implies $H \cong G$, (c) $G \cong H$ and $H \cong K$ imply $G \cong K$. At times it will seem that establishing an isomorphism between two groups is just another way of saying that they are "essentially the same group," or that they are "abstractly identical." Indeed, the isomorphism notion only provides a well-defined term to describe the situation. Examples of isomorphic groups will be mentioned in the exercises.

A special kind of isomorphism is the *automorphism*. An automorphism is an isomorphism of a group with itself. In the next theorem we will meet special automorphisms.

Theorem 1–3.10. The mappings α_g: $G \to g^{-1}Gg$, where g is a fixed element of the group G, are automorphisms of G. If H is a subgroup of G, then the mappings α_g, for all elements g in H, form a group.

The mappings α_g are called *inner automorphisms*.

Proof. The mapping α_g has as inverse α_g^{-1}: $G \to gGg^{-1}$ which is single-valued. Thus α_g is a bijective mapping of G on itself. For establishing an automorphism we have to prove, for two arbitrary elements a and b of G, that $ab \to (g^{-1}ag)(g^{-1}bg)$. But indeed $ab \to g^{-1}(ab)g = g^{-1}agg^{-1}bg$. In order to prove that the inner automorphisms form a group, we observe first that the identity mapping α_1 is an automorphism, and that $\alpha_{g^{-1}}$ is the inverse of α_g.

The closure is obvious. Namely, let g and h be in H. Then

$$\alpha_g\alpha_h\colon G \to h^{-1}(g^{-1}Gg)h = (gh)^{-1}G(gh),$$

and $\alpha_g\alpha_h = \alpha_{gh}$. Since associativity of one-to-one mappings of a set onto itself can be taken for granted, the theorem is proved.

If $b = g^{-1}cg$, we say that c has been *transformed by the element* g to yield b. The inner automorphism α_g, therefore, transforms the group G by the element g.

Transformation groups. We are now ready to introduce *transformation groups*. We mentioned already that the result of performing a transformation α on a set S, followed by a second transformation β, is a third transformation which we call $\alpha\beta$. We always assume that S is the domain and range of α, β, and $\alpha\beta$. We may consider this composition of transformations as an operation and inquire whether the transformations form a group under this operation. Since $\alpha\beta$ is uniquely determined, Gp 1 is fulfilled. In order to verify Gp 2 ,we observe that

$$(S\alpha\beta)\delta = [(S\alpha)\beta]\delta = (S\alpha)\beta\delta.$$

The transformation which leaves all elements of S unchanged is called the *identity transformation* I and has the property required in Gp 3. To each transformation α, as a bijective mapping of S on itself, there exists a unique α^{-1} such that $\alpha\alpha^{-1} = \alpha^{-1}\alpha = I$. This *inverse transformation* of α maps each element of S on its preimage under α.

We can summarize these results as:

Theorem 1-3.11. The transformations of a set S form a group $\mathcal{G}(S)$ under composition.

$\mathcal{G}(S)$ and its subgroups are called *transformation groups*.

In the special case in which S is finite, every transformation becomes a permutation of the elements of S. The group $\mathcal{G}(S)$ is then a so-called *symmetric group* of permutations. If S has n elements, $\mathcal{G}(S)$ has order $n!$.

EXERCISES

1. Which of the following sets are groups?
   (a) All rationals under addition
   (b) All rationals under multiplication
   (c) The complex numbers z with $|z| = 1$, under multiplication
   (d) All integers under subtraction
   (e) All even integers under addition
2. Let $n > 1$ be an integer. Do the n roots of the equation $x^n = 1$ form a group under addition? under multiplication?

3. Over the set of all rationals distinct from -1 let an operation be defined by $a * b = a + b + ab$ (usual addition and multiplication).
   (a) Show that the result is a group.
   (b) Is the group abelian?
   (c) Do the integers $\neq -1$ form a subgroup?
4. Is the multiplicative group of nonzero reals isomorphic with the additive group of all reals?
5. Prove:
   (a) Under an isomorphism between two groups the identity elements correspond, and all the inverses of corresponding elements correspond themselves.
   (b) A group isomorphic to an abelian group is itself abelian.
6. (a) Let n be an integer >0. Two integers are to be considered as identical when their difference is divisible by n. Then prove that the n integers $0, 1, 2, \ldots, n-1$ form an additive group G.

   Prove:
   (b) The integers $1, 2, \ldots, n-1$ form a multiplicative group H, provided n is prime.
   (c) G, with $n = 4$, and H, with $n = 5$, are isomorphic.
7. Prove that the set of all even integers forms a group under addition. What kind of a set is formed by the odd integers?
8. Prove: The group of all even integers is isomorphic to the group of all integers under addition.
9. Prove: If a subgroup H is finite, each coset Ha of H has exactly as many elements as H has.
10. Prove: A nonempty subset C of a group is a coset if and only if the following holds true. If a, b, and c are elements of C, then $ab^{-1}c$ is also in C.
11. If a and b are elements of a group, are ab and ba conjugate?
12. G is the group of all transformations of the set of reals that can be expressed by $x \to ax + b$, where $a \neq 0$ and b are real.
    (a) Prove: All the elements of G with $a = 1$ form a normal subgroup H of G.
    (b) Describe the right and left cosets of G with respect to H.
    (c) Does the subset of all elements of G with $b = 0$ form a normal subgroup of G?
    (d) Is G abelian? is H abelian?
13. Prove: If H and K are normal subgroups of a group G, so are $H \cap K$ and HK (the set of all products hk with $h \in H$ and $k \in K$).
14. (a) Do $K \lhd H$ and $H \lhd G$ imply $K \lhd G$?
    (b) Do $H \lhd G$ and $K \lhd G$ and $K \subset H$ imply $K \lhd H$?
15. Let Z (the "center") be the set of all elements z of a group G with the property that $zg = gz$ for all elements g in G. Is Z a normal subgroup of G?

16. Prove: The transformation of all group elements by an element g brings about the identity automorphism if and only if g is in the center (Exercise 15).
17. Are " is father of" and "is brother of" equivalence relations for the set of all human males?
18. If parallel lines in the plane are defined as lines with equal slope, show that parallelism is an equivalence relation for the set of all lines in the plane. What are the equivalence classes?
19. Prove the existence of the equivalence classes for an equivalence relation, as described in the text.
20. (a) Prove that $(\sim)$ is an equivalence relation if $x \sim y$ in a group G with subgroup H means $xy^{-1} \in H$.
    (b) Prove that the equivalence classes are the cosets of G with respect to H.

1–4 COLLINEATION GROUPS OF THE COMPLEX NUMBER PLANE

Translations. The transformations which we met in Section 1–2 were of the types $z \to az + b$ and $z \to a\bar{z} + b$, with the restriction $|a| = 1$. In this section we will, temporarily, remove the restriction $|a| = 1$ and study the resulting transformations, with special attention to the various groups that are formed by them. The requirement $|a| = 1$ was necessary and sufficient for the transformations to be distance-preserving. Thus the transformations $z \to az + b$ and $z \to a\bar{z} + b$ will be of a more general type, and the properties preserved by them, namely, similarity of triangles and of figures in general will, therefore, turn out to be weaker than the equality of distance. It is well known from elementary geometry that congruent figures are also similar. Thus isometries are special similarity-preserving transformations, and in this section we will see how the isometries fit into the more general picture.

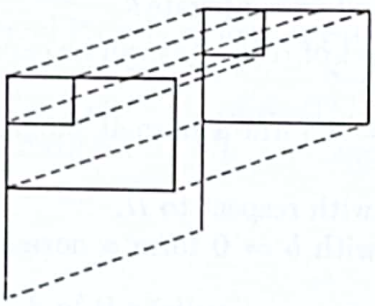

Figure 1–5

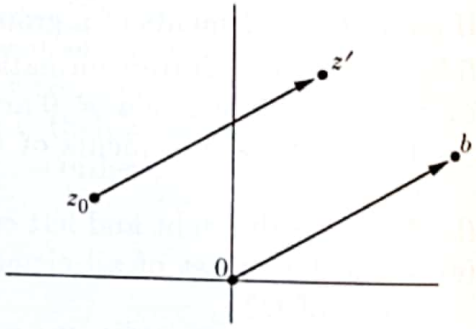

Figure 1–6

We start from one of the simplest types of transformation, the so-called *translations*, represented by $z \to z + b$. By Theorem 1–2.1, translations are isometries. Geometrically it is easy to see (Fig. 1–5) that a translation takes every point z into a point z' such that the directed segment from z to z' is parallel and equal in length to the position vector of b and equally directed (Fig. 1–6).