CHAPTER II

The Fundamental Group

§1. Introduction

For any topological space X and any point $x_0 \in X$, we will define a group, called *the fundamental group* of X, and denoted by $\pi(X, x_0)$. (Actually, the choice of the point x_0 is usually of minor importance, and hence it is often omitted from the notation.) We define this group by a very simple and intuitive procedure involving the use of closed paths in X. From the definition, it will be clear that the group is a topological invariant of X; i.e., if two spaces are homeomorphic, their fundamental groups are isomorphic. This gives us the possibility of proving that two spaces are not homeomorphic by proving that their fundamental groups are nonisomorphic. For example, this method suffices to distinguish between the various compact surfaces and in many other cases.

Now only does the fundamental group give information about spaces, but it also is often useful in studying continuous maps. As we shall see, any continuous map from a space X into a space Y induces a homomorphism of the fundamental group of X into that of Y. Certain topological properties of the continuous map will be reflected in the properties of this induced homomorphism. Thus, we can prove facts about certain continuous maps by studying the induced homomorphism of the fundamental groups.

We can summarize the above two paragraphs are follows: By using the fundamental group, topological problems about spaces and continuous maps can sometimes be reduced to purely algebraic problems about groups and homomorphisms. This is the basic strategy of the entire subject of algebraic topology: to find methods of reducing topological problems to questions of pure algebra, and then hope that algebraists can solve the latter.

This chapter will only give the basic definition and properties of the fundamental group and induced homomorphism, and determine its structure for a few very simple spaces. In later chapters we shall develop more general methods for determining the fundamental groups of some more interesting spaces.

§2. Basic Notation and Terminology

As usual, for any real numbers a and b such that $a < b$, $[a, b]$ denotes the closed interval of the real line with a and b as end points. For conciseness, we set $I = [0, 1]$. We note that, given any two closed intervals $[a, b]$ and $[c, d]$, there exist unique *linear* homeomorphisms

$$h_1, h_0 : [a, b] \to [c, d],$$

such that

$$h_0(a) = c, \qquad h_0(b) = d,$$

$$h_1(a) = d, \qquad h_1(b) = c.$$

We distinguish between these two by calling h_0 *orientation preserving* and h_1 *orientation reversing*.

A *path* or *arc* in a topological space X is a continuous map of some closed interval into X. The images of the end points of the interval are called the *end points* of the path or arc, an the path is said to *join* its end points. One of the end points is called the *initial* point, the other is called the *terminal* point (it is clear which is which).

A space X is called *arcwise connected* or *pathwise connected* if any two points of X can be joined by an arc. An arcwise-connected space is connected, but the converse statement is not true. The *arc components* of X are the maximal arcwise-connected subsets of X (by analogy with the ordinary components of X). Note that the arc components of X need not be closed sets. A space is *locally arcwise connected* if each point has a basic family of arcwise-connected neighborhoods (by analogy with ordinary local connectivity).

EXERCISE

2.1. Prove that a space which is connected and locally arcwise connected is arcwise connected.

Definition. Let $f_0, f_1 : [a, b] \to X$ be two paths in X such that $f_0(a) = f_1(a)$, $f_0(b) = f_1(b)$ (i.e., the two paths have the same initial and terminal points). We say that these two paths are *equivalent*, denoted by $f_0 \sim f_1$, if and only if there exists a continuous map

$$f : [a, b] \times I \to X,$$

such that

$$\left.\begin{array}{l} f(t,0)=f_0(t) \\ f(t,1)=f_1(t) \end{array}\right\} t\in[a,b],$$

$$\left.\begin{array}{l} f(a,s)=f_0(a)=f_1(a) \\ f(b,s)=f_0(b)=f_1(b) \end{array}\right\} s\in I.$$

Note that in the above definition we could replace I by any other closed interval if necessary. We leave it as an exercise to verify that this relation is reflexive, symmetric, and transitive.

Intuitively we say that two paths are equivalent if one can be continuously deformed into the other in the space X. During the defomation, the end points must remain fixed.

Our second basic definition is that of the *product* of two paths. The product of two paths is only defined if the terminal point of the first path is the initial point of the second path. If this condition holds, the product path is traversed by traversing the first path and then the second path, in the given order. To be precise, assume

$$f:[a,b]\to X,$$

$$g:[b,c]\to X$$

are paths such that $f(b)=g(b)$ (here $a<b<c$). Then the product $f\cdot g$ is defined by

$$(f\cdot g)t=\begin{cases} f(t), & t\in[a,b] \\ g(t), & t\in[b,c]. \end{cases} \tag{2.2.1}$$

It is a map $[a,c]\to X$. In the above definition, we had the rather cumbersome requirement that the domains of f and g had to be the intervals $[a,b]$ and $[b,c]$, respectively. We can remove this requirement by changing the domain of f or g by means of an orientation-preserving linear homeomorphism. Actually, in the future we shall only be interested in equivalence classes of paths rather than the paths themselves. By "equivalence class," we mean, with respect to the equivalence relation defined above and also with respect to the following obvious equivalence relation: If $f:[a,b]\to X$ and $g:[c,d]\to X$ are paths such that $g=fh$, where $h:[c,d]\to[a,b]$ is an *orientation-preserving* linear homeomorphism, then f and g are to be regarded as equivalent. Rather than considering paths whose domain is an arbitrary closed interval and allowing orientation-preserving linear homeomorphisms between any two such intervals, we find it technically simpler to demand that all paths be functions defined on one fixed interval, namely, the interval $I=[0,1]$. As a result of this simplification, the simple formula for the product of two paths, (2.2.1), has to be replaced by a more complicated formula. Also, it will not be immediately obvious that the multiplication of path classes is associative. However, the reader should keep in mind that there are various alternative ways of proceeding with this subject.

§3. Definition of the Fundamental Group of a Space

From now on, by a *path in X* we mean a continuous map $I \to X$. If f and g are paths in X such that the terminal point of f is the initial point of g, then the product $f \cdot g$ is defined by

$$(f \cdot g)t = \begin{cases} f(2t), & 0 \leqq t \leqq \frac{1}{2} \\ g(2t-1), & \frac{1}{2} \leqq t \leqq 1. \end{cases}$$

We say two paths, f_0 and f_1, are *equivalent* ($f_0 \sim f_1$) if the condition in §2 is satisfied.

Lemma 3.1. *The equivalence relation and the product we have defined are compatible in the following sense: If $f_0 \sim f_1$ and $g_0 \sim g_1$, then $f_0 \cdot g_0 \sim f_1 \cdot g_1$ (it is assumed, of course, that the terminal point of f_0 is the initial point of g_0).*

The proof may be left to the reader. In proving lemmas such as this, the following fact is often useful: *Let A and B be closed subsets of the topological space X such that $X = A \cup B$. If f is a function defined on X such that the restrictions $f|A$ and $f|B$ are both continuous, then f is continuous.* The proof, which is easy, is left to the reader. In the future, we will use this fact without comment.

As a result of Lemma 3.1, the multiplication of paths defines a multiplication of equivalence classes of paths (provided the terminal point of the first path and the initial point of the second path coincide). It is this multiplication of equivalence classes with which we are primarily concerned. Note that the multiplication of paths is not associative in general, i.e., $(f \cdot g) \cdot h \neq f \cdot (g \cdot h)$ (we assume both products are defined). However, we have

Lemma 3.2. *The multiplication of equivalence classes of paths is associative.*

Proof. It suffices to prove the following: Let f, g, and h be paths such that the terminal point of f = initial point of g, and the terminal point of g = initial point of h. Then

$$(f \cdot g) \cdot h \sim f \cdot (g \cdot h).$$

To prove this, consider the function $F : I \times I \to X$ defined by

$$F(t, s) = \begin{cases} f\left(\dfrac{4t}{1+s}\right), & 0 \leqq t \leqq \dfrac{s+1}{4} \\ g(4t-1-s), & \dfrac{s+1}{4} \leqq t \leqq \dfrac{s+2}{4} \\ h\left(1 - \dfrac{4(1-t)}{2-s}\right), & \dfrac{s+2}{4} \leqq t \leqq 1. \end{cases}$$

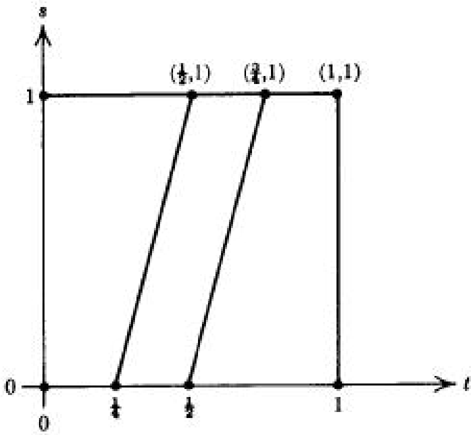

FIGURE 2.1. Proof of associativity.

Then, F is continuous, $F(t, 0) = [(f \cdot g) \cdot h]t$, and $F(t, 1) = [f \cdot (g \cdot h)]t$. The motivation for the definition of F is given in Figure 2.1. Q.E.D.

For any point $x \in X$, let us denote by $\mathscr{E}_x$ the equivalence class of the constant map of I into the point x of X. This path class has the following fundamental property:

Lemma 3.3. *Let α be an equivalence class of paths with initial point x and terminal point y. Then $\mathscr{E}_x \cdot \alpha = \alpha$ and $\alpha \cdot \mathscr{E}_y = \alpha$.*

PROOF. Let $e : I \to X$ be the constant map such that $e(I) = \{x\}$ and let $f : I \to X$ be a representative of the path class α. To prove the first relation, it suffices to prove that $e \cdot f \sim f$. Define $F : I \times I \to X$ by

$$F(t, s) = \begin{cases} x, & 0 \leqq t \leqq \frac{1}{2}s \\ f\left(\dfrac{2t - s}{2 - s}\right), & \frac{1}{2}s \leqq t \leqq 1. \end{cases}$$

Then $F(t, 0) = f(t)$ and $F(t, 1) = (e \cdot f)t$ as required. The motivation for the definition of F is shown in Figure 2.2. The proof that $\alpha \cdot \mathscr{E}_y = \alpha$ is similar and is left to the reader. Q.E.D.

For any path $f : I \to X$, let $\bar{f}$ denote the path defined by

$$\bar{f}(t) = f(1 - t), \qquad t \in I.$$

The path $\bar{f}$ is obtained by traversing the path f in the opposite direction.

Lemma 3.4 *Let α and $\bar{\alpha}$ denote the equivalence classes of the paths f and $\bar{f}$, respectively. Then,*

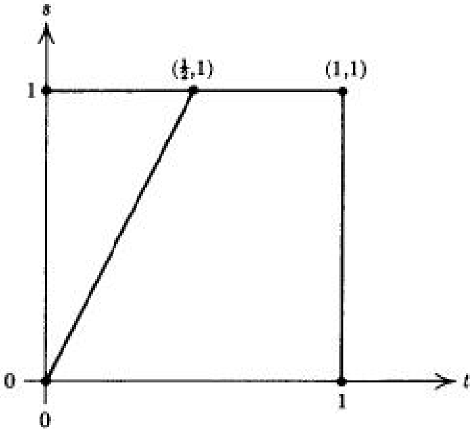

FIGURE 2.2. Proof of existence of units.

$$\alpha \cdot \bar{\alpha} = \mathscr{E}_x, \qquad \bar{\alpha} \cdot \alpha = \mathscr{E}_y,$$

where x and y are the initial and terminal points of the path f.

PROOF. To prove the first equation, it suffices to show that $f \cdot \bar{f} \sim e$, where e is the constant path at the point x. Therefore, we define $F : I \times I \to X$ by

$$F(t, s) = \begin{cases} f(2t), & 0 \leqq t \leqq \frac{1}{2}s \\ f(s), & \frac{1}{2}s \leqq t \leqq 1 - \frac{1}{2}s \\ f(2 - 2t), & 1 - \frac{1}{2}s \leqq t \leqq 1. \end{cases}$$

We then see that $F(t, 0) = x$, whereas $(f \cdot \bar{f})t = F(t, 1)$. Figure 2.3 explains the choice of the function F. We can also motivate the deformation of the path

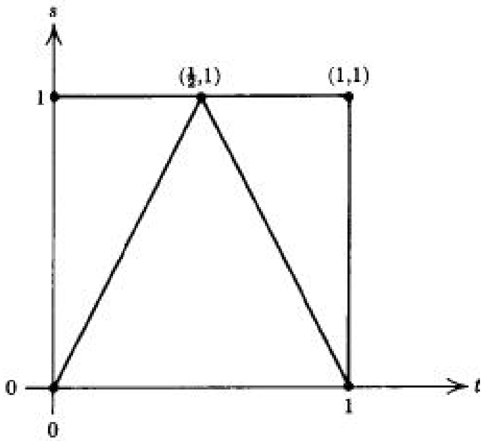

FIGURE 2.3. Proof of existence of inverses.

$f \cdot \bar{f}$ into the constant path e by a simple mechanical analogy. Consider the path f as an elastic "thread" in the space X from the point x to y; then $\bar{f}$ is another "thread" in the opposite direction, from y to x, and $f \cdot \bar{f}$ is represented by joining the two threads at the point y. We can now "pull in" the doubled thread to the point x because we do not need to keep it attached to the point y.

The proof that $\bar{\alpha} \cdot \alpha = \mathscr{E}_y$ is similar and is left to the reader. Q.E.D.

In view of these properties of the path class $\bar{\alpha}$, from now on we will denote it by α^{-1}. It is readily seen that the conditions of the lemma just proved characterize α^{-1} uniquely. Hence, if $f_0 \sim f_1$, then $\bar{f}_0 \sim \bar{f}_1$.

We can summarize the lemmas just proved by saying that the set of all path classes in X satisfies the axioms for a group, except that the product of two paths is not always defined.

Definition. A path, or path class, is called *closed*, or a *loop*, if the initial and terminal points are the same. The loop is said to be *based* at the common end point.

Let x be any point of X; it is readily seen that the set of all loops based at x is a group. This group is called the *fundamental group* or *Poincaré group* of X at the base point x and is denoted by $\pi(X, x)$.

Next, we will investigate the dependence of the group $\pi(X, x)$ on the base point x. Let x and y be two points in X, and let γ be a path class with initial point x and terminal point y (hence, x and y belong to the same arc component of X). Using the path γ, we define a mapping $u: \pi(X, x) \to \pi(X, y)$ by the formula $\alpha \to \gamma^{-1}\alpha\gamma$. We see immediately that this mapping is a homomorphism of $\pi(X, x)$ into $\pi(X, y)$. By using the path γ^{-1} instead of γ, we can define a homomorphism $v: \pi(X, y) \to \pi(X, x)$ in a similar manner. We immediately verify that the composed homomorphisms vu and uv are the identity maps of $\pi(X, x)$ and $\pi(X, y)$, respectively. Thus, u and v are isomorphisms, each of which is the inverse of the other. Thus, we have proved

Theorem. 3.5. *If X is arcwise connected, the groups $\pi(X, x)$ and $\pi(X, y)$ are isomorphic for any two points $x, y \in X$.*

The importance of this theorem is obvious; e.g., the question as to whether or not $\pi(X, x)$ has any given group theoretic property (e.g., it is abelian, finite, nilpotent, free, etc.) is independent of the point x, and thus depends only on the space X, provided X is arcwise connected.

On the other hand, we must keep in mind that there is no *canonical* or *natural* isomorphism between $\pi(X, x)$ and $\pi(X, y)$; corresponding to each choice of a path class from x to y there will be an isomorphism, from $\pi(X, x)$ to $\pi(X, y)$, and, in general, different path classes will give rise to different isomorphisms.

EXERCISES

3.1. Under what conditions will two path classes, γ and γ', from x to y give rise to the same isomorphism of $\pi(X, x)$ onto $\pi(X, y)$?

3.2. Let X be an arcwise-connected space. Under what conditions is the following statement true: For any two points $x, y \in X$, all path classes from x to y give rise to the same isomorphism of $\pi(X, x)$ onto $\pi(X, y)$?

3.3. Let $f, g : I \to X$ be two paths with initial point x_0 and terminal point x_1. Prove that $f \sim g$ if and only if $f \cdot \bar{g}$ is equivalent to the constant path at x_0 ($\bar{g}$ is defined as in Lemma 3.4).

We will actually determine the structure of the fundamental group of various spaces later in this chapter and in Chapter IV.

§4. The Effect of a Continuous Mapping on the Fundamental Group

Let $\varphi : X \to Y$ be a continuous mapping, and let $f_0, f_1 : I \to X$ be paths in X. It is readily seen that if f_0 and f_1 are equivalent, then so are the paths φf_0 and φf_1 represented by the composed functions. Thus, if α denotes the path class that contains f_0 and f_1, it makes sense to denote by $\varphi_*(\alpha)$ the path class that contains the paths φf_0 and φf_1. $\varphi_*(\alpha)$ is the image of th path class α in the space Y, and it is readily verified that the mapping φ_* which sends α into $\varphi_*(\alpha)$ has the following properties:

(a) If α and β are path classes in X such that $\alpha \cdot \beta$ is defined, then $\varphi_*(\alpha \cdot \beta) = (\varphi_* \alpha) \cdot (\varphi_* \beta)$.
(b) For any point $x \in X$, $\varphi_*(\mathscr{E}_x) = \mathscr{E}_{\varphi(x)}$.
(c) $\varphi_*(\alpha^{-1}) = (\varphi_* \alpha)^{-1}$.

For these reasons, we shall call φ_* a "homomorphism," or, the "homomorphism induced by φ."

If $\psi : Y \to Z$ is also a continuous map, then we can verify the following property easily:

(d) $(\psi\varphi)_* = \psi_* \varphi_*$.

Finally, if $\varphi : X \to X$ is the identity map, then

(e) $\varphi_*(\alpha) = \alpha$ for any path class α in X; i.e., φ_* is the identity homomorphism.

Note that, in view of these properties, a continuous map $\varphi : X \to Y$ induces a homomorphism $\varphi_* : \pi(X, x) \to \pi(Y, \varphi(x))$; and, if φ is a homomorphism, then φ_* is an isomorphism. This induced homeomorphism will be extremely important in studying the fundamental group.

Caution: If φ is a one-to-one map, it does *not* follow that φ^* is one-to-one; similarly, if φ is onto, it does not follow that φ_* is onto. We shall see examples to illustrate this point later.

EXERCISE

4.1. Let $\varphi : X \to Y$ be a continuous map and let γ be a class of paths in X from x_0 to x_1. Prove that the following diagram is commutative:

$$\begin{array}{ccc} \pi(X, x_0) & \xrightarrow{\varphi_*} & \pi(Y, \varphi(x_0)) \\ \Big\downarrow u & & \Big\downarrow v \\ \pi(X, x_1) & \xrightarrow{\varphi_*} & \pi(Y, \varphi(x_1)). \end{array}$$

Here the isomorphism u is defined by $u(\alpha) = \gamma^{-1}\alpha\gamma$, and v is defined similarly using $\varphi_*(\gamma)$ in place of γ. [NOTE: An important special case occurs if $\varphi(x_0) = \varphi(x_1)$. Then, $\varphi_*(\gamma)$ is an element of the group $\pi(Y, \varphi(x_0))$.]

To make further progress in the study of the induced homomorphism φ_*, we must introduce the important notion of *homotopy* of continuous maps.

Definition. Two continuous maps $\varphi_0, \varphi_1 : X \to Y$ are *homotopic* if and only if there exists a continuous map $\varphi : X \times I \to Y$ such that, for $x \in X$,

$$\varphi(x, 0) = \varphi_0(x),$$
$$\varphi(x, 1) = \varphi_1(x).$$

If two maps φ_0 and φ_1 are homotopic, we shall denote this by $\varphi_0 \simeq \varphi_1$. We leave it to the reader to verify that this is an equivalence relation on the set of all continuous maps $X \to Y$. The equivalence classes are called *homotopy classes* of maps.

To better visualize the geometric content of the definition, let us write $\varphi_t(x) = \varphi(x, t)$ for any $(x, t) \in X \times I$. Then, for any $t \in I$,

$$\varphi_t : X \to Y$$

is a continuous map. Think of the parameter t as representing time. Then, at time $t = 0$, we have the map φ_0, and, as t varies, the map φ_t varies *continuously* so that at time $t = 1$ we have the map φ_1. For this reason, a homotopy is often spoken of as a continuous deformation of a map.[1]

[1] The student who is familiar with the compact-open topology for function spaces will recognize that two maps $\varphi_0, \varphi_1 : X \to Y$ are homotopic if and only if they can be joined by an arc in the space of all continuous functions $X \to Y$ (provided X and Y satisfy certain hypotheses). Indeed, the map $t \to \varphi_t$ in the above notation is a path from φ_0 to φ_1.

Definition. Two maps $\varphi_0, \varphi_1 : X \to Y$ are *homotopic relative to the subset A of X* if and only if there exists a continuous map $\varphi : X \times I \to Y$ such that

$$\begin{aligned} \varphi(x, 0) &= \varphi_0(x), && x \in X, \\ \varphi(x, 1) &= \varphi_1(x), && x \in X, \\ \varphi(a, t) &= \varphi_0(a) = \varphi_1(a), && a \in A, t \in I. \end{aligned}$$

Note that this condition implies $\varphi_0 | A = \varphi_1 | A$.

Theorem 4.1. *Let $\varphi_0, \varphi_1 : X \to Y$ be maps that are homotopic relative to the subset $\{x\}$. Then*

$$\varphi_{0*} = \varphi_{1*} : \pi(X, x) \to \pi(Y, \varphi_0(x)),$$

i.e., the induced homomorphisms are the same.

PROOF. The proof is immediate.

Unfortunately, the condition that the homotopy should be relative to the base point x is too restrictive for many purposes. This condition can be omitted, but we then complicate the statement of the theorem. We shall, however, do this in §8.

We shall now apply some of these results.

Definition. A subset A of a topological space X is called a *retract* of X if there exists a continuous map $r : X \to A$ (called a *retraction*) such that $r(a) = a$ for any $a \in A$.

As we shall see shortly, it is a rather strong condition to require that a subset A be a retract of X. A simple example of a retract of a space is the "center circle" of a Möbius strip. (What is the retraction in this case?)

Now let $r : X \to A$ be a retraction, as in the above definition, and $i : A \to X$ the inclusion map. For any point $a \in A$, consider the induced homomorphisms

$$\begin{aligned} i_* &: \pi(A, a) \to \pi(X, a), \\ r_* &: \pi(X, a) \to \pi(A, a). \end{aligned}$$

Because ri = identity map, we conclude that $r_* i_*$ = identity homomorphism of the group $\pi(A, a)$, by properties (d) and (e) given previously. From this we conclude that *i_* is a monomorphism* and *r_* is an epimorphism.* Moreover, the condition that $r_* i_*$ = identity imposes strong restrictions on the subgroup $i_* \pi(A, a)$ of $\pi(X, a)$.

We shall actually use this result later to prove that certain subspaces are not retracts.

EXERCISES

4.2. Show that a retract of a Hausdorff space must be a closed subset.

4.3. Prove that if A is a retract of X, $r: X \to A$ is a retraction, $i: A \to X$ is the inclusion, and $i_*\pi(A)$ is a normal subgroup of $\pi(X)$, then $\pi(X)$ is the direct product of the subgroups image i_* and kernel r_* (see §2 of Chapter III for the definition of direct product of groups).

4.4. Let A be a subspace of X, and let Y be a nonempty topological space. Prove that $A \times Y$ is a retract of $X \times Y$ if and only if A is a retract of X.

4.5. Prove that the relation "is a retract of" is transitive, i.e., if A is a retract of B and B is a retract of C, then A is a retract of C.

We now introduce the notion of *deformation retract*. The subspace A is a deformation retract of X if there exists a retraction $r: X \to A$ homotopic to the identity map $X \to X$. The precise definition is as follows:

Definition. A subset A of X is a *deformation retract*[2] of X if there exists a retraction $r: X \to A$ and a homotopy $f: X \times I \to X$ such that

$$\left.\begin{aligned} f(x, 0) &= x \\ f(x, 1) &= r(x) \end{aligned}\right\} \quad x \in X,$$

$$f(a, t) = a, \quad a \in A, t \in I.$$

Theorem 4.2. *If A is a deformation retract of X, then the inclusion map $i: A \to X$ induces an isomorphism of $\pi(A, a)$ onto $\pi(X, a)$ for any $a \in A$.*

Proof. As above, r_*i_* is the identity map of $\pi(A, a)$. We will complete the proof by showing that i_*r_* is the identity map of $\pi(X, a)$. This follows because ir is homotopic to the identity map $X \to X$ (relative to $\{a\}$); hence, Theorem 4.1 is applicable. Q.E.D.

We shall use this theorem in two different ways. On the one hand, we shall use it throughout the rest of this book to prove that two spaces have isomorphic fundamental groups. On the other hand, we can use it to prove that a subspace is not a deformation retract by proving the fundamental groups are not isomorphic. In particular, we shall be able to prove that certain retracts are not deformation retracts.

Definition. A topological space X is *contractible to a point* if there exists a point $x_0 \in X$ such that $\{x_0\}$ is a deformation retract of X.

Definition. A topological space X is *simply connected* if it is arcwise connected and $\pi(X, x) = \{1\}$ for some (and hence any) $x \in X$.

Corollary 4.3. *If X is contractible to a point, then X is simply connected.*

[2] Some authors define this term in a slightly weaker fashion.

Examples

4.1. A subset X of the plane or, more generally, of Euclidean n-space $\mathbf{R}^n$ is called *convex* if the line segment joining any two points of X lies entirely in X. We assert that *any convex subset X of $\mathbf{R}^n$ is contractible to a point.* To prove this, choose an arbitrary point $x_0 \in X$, and then define $f : X \times I \to X$ by the formula

$$f(x, t) = (1 - t)x + ix_0$$

for any $(x, t) \in X \times I$ [i.e., $f(x, t)$ is the point on the line segment joining x and x_0 which divides it in the ratio $(1 - t) : t$]. Then f is continuous, $f(x, 0) = x$, and $f(x, 1) = x_0$, as required. More generally, we may define a subset X of $\mathbf{R}^n$ to be *starlike with respect to the point* $x_0 \in X$ provided the line segment joining x and x_0 lies entirely in X for any $x \in X$. Then, the same proof suffices to show that if X is starlike with respect to x_0, it is contractible to the point x_0.

4.2. We assert that the unit $(n - 1)$-sphere S^{n-1} is a deformation retract of $E^n - \{0\}$, the closed unit n-dimensional disc minus the origin. To prove this, define a map $f : X \times I \to X$, where

$$X = E^n - \{0\} = \{x \in \mathbf{R}^n : 0 < |x| \leqq 1\},$$

by the formula

$$f(x, t) = (1 - t)x + t \cdot \frac{x}{|x|}.$$

(The reader should draw a picture to show what happens here when $n = 2$ or $n = 3$.) Then f is continuous, $f(x, 0) = x$, $f(x, 1) = x/|x| \in S^{n-1}$, and, if $x \in S^{n-1}$, then $f(x, t) = x$ for all $t \in I$. In particular, for $n = 2$, we see that the boundary circle is a deformation retract of a punctured disc.

EXERCISES

4.6. Let x_0 be any point in the plane $\mathbf{R}^2$. Find a circle C in $\mathbf{R}^2$ which is a deformation retract of $\mathbf{R}^2 - \{x_0\}$. What is the n-dimensional analog of this fact?

4.7. Find a circle C which is a deformation retract of the Möbius strip.

4.8. Let T be a torus and let X be the complement of a point in T. Find a subset of X which is homeomorphic to a figure "8" curve (i.e., the union of two circles with a single point in common) and which is a deformation retract of X.

4.9. Generalize Exercise 4.8 to arbitrary compact surfaces, i.e., let S be a compact surface and let X be the complement of a point in S. Find a subset A of X such that (a) A is homeomorphic to the union of a finite number of circles and (b) A is a deformation retract of X. (HINT: Consider the representation of S as the space obtained by identifying in pairs the edges of a certain polygon.)

4.10. Let x and y be distinct points of a simply connected space X. Prove that there is a *unique* path class in X with initial point x and terminal point y.

4.11. Let X be a topological space, and for each positive integer n let X_n be an arcwise-connected subspace containing the base point $x_0 \in X$. Assume that the subspaces X_n are nested, i.e., $X_n \subset X_{n+1}$ for all n, that

$$X = \bigcup_{n=1}^{\infty} X_n,$$

and that for any compact subset A of X there exists an integer n such that $A \subset X_n$. (EXAMPLE: Each X_n is open.) Let $i_n : \pi(X_n) \to \pi(X)$ and $j_{mn} : \pi(X_m) \to \pi(X_n)$, $m < n$, denote homomorphisms induced by inclusion maps. Prove the following two statements: (a) For any $\alpha \in \pi(X)$, there exists an integer n and an element $\alpha' \in \pi(X_n)$ such that $i_n(\alpha') = \alpha$. (b) If $\beta \in \pi(X_m)$ and $i_m(\beta) = 1$, then there exists an integer $n \geqq m$ such that $j_{mn}(\beta) = 1$. [REMARK: These two statements imply that $\pi(X)$ is the *direct limit* of the sequence of groups $\pi(X_n)$ and homomorphisms j_{mn}. We shall see examples later on where the hypotheses of this exercise are valid.] If the homomorphisms $j_{n,n+1}$ are monomorphisms for all n, prove that each i_n is also a monomorphism and that $\pi(X)$ is the union of the subgroups $i_n\pi(X_n)$.

§5. The Fundamental Group of a Circle is Infinite Cyclic

Let S^1 denote the unit circle in the Euclidean plane $\mathbf{R}^2$, $S^1 = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$ (or, equivalently, in the complex plane $\mathbf{C}$). Let $f : I \to S^1$ denote the closed path that goes around the circle exactly once, defined by

$$f(t) = (\cos 2\pi t, \sin 2\pi t), \quad 0 \leqq t \leqq 1,$$

and denote the equivalence class of f by the symbol α.

Theorem 5.1. *The fundamental group $\pi(S^1, (1, 0))$ is an infinite cyclic group generated by the path class α.*

PROOF. Let $g : I \to S^1$, $g(0) = g(1) = (1, 0)$ be a closed path in S^1. We shall prove first that g belongs to the equivalence class α^m for some integer m (m may be positive, negative, or zero). Let

$$U_1 = \{(x, y) \in S^1 : y > -\tfrac{1}{10}\},$$
$$U_2 = \{(x, y) \in S^1 : y < +\tfrac{1}{10}\}.$$

Then, U_1 and U_2 are connected open subsets of S^1, each of which is slightly larger than a semicircle, and $U_1 \cup U_2 = S^1$. Obviously U_1 and U_2 are each homeomorphic to an open interval of the real line, hence, each is contractible. In the case where $g(I) \subset U_1$ or $g(I) \subset U_2$, it is then clear that g is equivalent to the constant path, and hence belongs to the equivalence class of α^0. We put this case aside and assume from now on that $g(I) \not\subset U_1$ and $g(I) \not\subset U_2$.

We next assert that it is possible to divide the unit interval into subintervals $[0, t_1]$, $[t_1, t_2]$, $\ldots$, $[t_{n-1}, 1]$, where $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$, such that the following conditions hold:

(a) $$g([t_i, t_{i+1}]) \subset U_1 \quad \text{or}$$
$$g([t_i, t_{i+1}]) \subset U_2 \quad \text{for} \quad 0 \leqq i < n.$$

(b) $$g([t_{i-1}, t_i]) \quad \text{and} \quad g([t_i, t_{i+1}])$$

are not both contained in the *same* open set U_j, $j = 1$ or 2.

This assertion may be proved as follows. $\{g^{-1}(U_1), g^{-1}(U_2)\}$ is an open covering of the compact metric space I; let $\mathscr{E}$ be a Lebesgue number[3] of this covering.

Divide the unit interval in any way whatsoever into subintervals of length $< \varepsilon$. With this subdivision, condition (a) will hold; however, condition (b) may not hold. If two consecutive subintervals are mapped by g into the same set U_j, then amalgamate these two subintervals into a single subinterval by omitting the common end point. Continue this process of amalgamation until condition (b) holds.

Let β denote the equivalence class of the path g, and let β_i denote the equivalence class of $g|[t_{i-1}, t_i]$ for $1 \leqq i \leqq n$. Then, obviously, β is a product,

$$\beta = \beta_1 \cdot \beta_2 \cdot \ldots \cdot \beta_n.$$

Each β_i is a path in U_1 or U_2. Because of condition (b), it is clear that $g(t_i) \in U_1 \cap U_2$. $U_1 \cap U_2$ has two components, one of which contains the point $(1, 0)$, and the other of which contains the point $(-1, 0)$. For each index i, $0 < i < n$, choose a path class γ_i in $U_1 \cap U_2$ with initial point $g(t_i)$ and terminal point $(1, 0)$ or $(-1, 0)$, depending on which component of $U_1 \cap U_2$ contains $g(t_i)$. Let

$$\delta_1 = \beta_1 \gamma_1,$$
$$\delta_i = \gamma_{i-1}^{-1} \beta_i \gamma_i \quad \text{for} \quad 1 < i < n,$$
$$\delta_n = \gamma_{n-1}^{-1} \beta_n.$$

Then, it is clear that

$$\beta = \delta_1 \delta_2 \cdots \delta_n. \tag{2.51}$$

where each δ_i is a path class in U_1 or U_2 having its initial and terminal points in the set $\{(1, 0), (-1, 0)\}$. For any index i, if δ_i is a closed path class, then $\delta_i = 1$, because U_1 and U_2 are simply connected. We may therefore assume that any such δ_i has been dropped from formula (2.5-1), and, changing notation if necessary, that $\delta_i, \delta_2, \ldots,$ and δ_n are not closed paths.

Becuase U_1 is simply connected, there is a unique path class η_1 in U_1 with

[3] We say ε is a *Lebesgue number* of a covering of a metric space X if the following condition holds: Any subset of X of diameter $< \varepsilon$ is contained in some set of the covering. It is a theorem that any open covering of a compact metric space has a Lebesgue number. The reader may either prove this as an exercise or look up the proof in a textbook on general topology.

initial point (1, 0) and terminal point $(-1, 0)$ (see Exercise 4.10). Also, η_1^{-1} is the unique path class in U_1 with initial point $(-1, 0)$ and terminal point (1, 0). Analogously, we denote by η_2 the unique path class in U_2 with initial point $(-1, 0)$ and terminal point (1, 0). Note that $\eta_1\eta_2 = \alpha$.

Thus, we see that, for each index i,

$$\delta_i = \eta_1^{\pm 1} \quad \text{or} \quad \delta_i = \eta_2^{\pm 1}.$$

In view of condition (b) above, if $\delta_i = \eta_1^{\pm 1}$, then $\delta_{i+1} = \eta_2^{\pm 1}$, while if $\delta_i = \eta_2^{\pm 1}$, then $\delta_{i+1} = \eta_1^{\pm 1}$. Therefore only the following possibilities remain:

$$\beta = 1,$$

$$\beta = \eta_1\eta_2\eta_1\eta_2\cdots\eta_1\eta_2,$$

or

$$\beta = \eta_2^{-1}\eta_1^{-1}\eta_2^{-1}\eta_1^{-1}\cdots\eta_2^{-1}\eta_1^{-1}.$$

In the second case $\beta = \alpha^m$ for some $m > 0$, whereas in the third case $\beta = \alpha^m$ for some integer $m < 0$. Thus, we have $\beta = \alpha^m$ in all cases.

From this it follows that $\pi(S^1)$ is a cyclic group. However, this argument gives no hint as to the order of $\pi(S^1)$. In §3 of Chapter V we will complete the proof by showing that $\pi(S^1)$ is an infinite group, using the theory of covering spaces; another proof is given in the discussion of Example 7.1 of Chapter V. When we introduce homology theory later on, it will be easy to give still other proofs.

It would be possible to give a direct, ad hoc proof now that $\pi(S^1)$ is infinite; see Massey ([2], Chapter II) or Ahlfors and Sario ([1], Chapter I, Section 10). It is also possible to give a proof using the concept of the *winding number* or *index* of a closed path in the plane with respect to a point; this is explained in most textbooks on complex function theory. The theory of the winding number or index can also be developed in the context of real function theory.

Given the fundamental importance of Theorem 5.1 and its basic intuitive appeal, it is not surprising that there should be so many different proofs available. Q.E.D.

As a corollary of Theorem 5.1, we see that the fundamental group of any space with a circle as deformation retract is infinite cyclic. Examples of such spaces are the Möbius strip, a punctured disc, the punctured plane, a region in the plane bounded by two concentric circles, etc. (see the exercises in the preceding section).

Exercises

5.1. Let $\{U_i\}$ be an open covering of the space X having the following properties: (a) There exists a point x_0 such that $x_0 \in U_i$ for all i. (b) Each U_i is simply connected.

(c) If $i \neq j$, then $U_i \cap U_j$ is arcwise connected. Prove that X is simply connected. [HINT: To prove any loop $f: I \to X$ based at x_0 is trivial, first consider the open covering $\{f^{-1}(U_i)\}$ of the compact metric space I and make use of the Lebesgue number of this covering.]
Remark. The two most important cases of this exercise are the following: (1) A covering by two open sets and (2) the sets U_i are linearly ordered by inclusion. The student should restate the exercise for these two special cases.

5.2. Use the result of Exercise 5.2, remark (1), to prove that the unit 2-sphere S^2 or, more generally, the n-sphere S^n, $n \geq 2$, is simply connected.

5.3. Prove that $\mathbf{R}^2$ and $\mathbf{R}^n$ are not homeomorphic if $n \neq 2$. (HINT: Consider the complement of a point in $\mathbf{R}^2$ or $\mathbf{R}^n$.)

5.4. Prove that any homeomorphism of the closed disc E^2 onto itself maps S^1 onto S^1 and U^2 onto U^2.

§6. Application: The Brouwer Fixed-Point Theorem in Dimension 2

One of the best known theorems of topology is the following fixed-point theorem of L.E.J. Brouwer. Let E^n denote the closed unit ball in Euclidean n-space $\mathbf{R}^n$:

$$E^n = \{x \in \mathbf{R}^n : |x| \leqq 1\}.$$

Theorem 6.1. *Any continuous map f of E^n into itself has at least one fixed point, i.e., a point x such that $f(x) = x$.*

We shall only prove this theorem for $n \leqq 2$. Before going into the proof, it seems worthwhile to indicate why there should be interest in fixed-point theorems such as this one.

Suppose we have a system of n equations in n unknowns:

$$\begin{aligned} g_1(x_1, \ldots, x_n) &= 0, \\ g_2(x_1, \ldots, x_n) &= 0, \\ &\vdots \\ g_n(x_1, \ldots, x_n) &= 0. \end{aligned} \tag{2.6.1}$$

Here the g_i's are assumed to be continuous real-valued functions of the real variables $x_1, \ldots, x_n$. It is often an important problem to be able to decide whether or not such a system of equations has a solution. We can transform this problem into a fixed-point problem as follows. Let

$$h_i(x_i, \ldots, x_n) = g_i(x_1, \ldots, x_n) + x_i$$