Module MA3427: Algebraic Topology I
Michaelmas Term 2010

D. R. Wilkins

Contents

1 Topological Spaces

1.1 Notions of Continuity

The concept of continuity is fundamental in large parts of contemporary mathematics. In the nineteenth century, precise definitions of continuity were formulated for functions of a real or complex variable, enabling mathematicians to produce rigorous proofs of fundamental theorems of real and complex analysis, such as the Intermediate Value Theorem, Taylor's Theorem, the Fundamental Theorem of Calculus, and Cauchy's Theorem.

In the early years of the Twentieth Century, the concept of continuity was generalized so as to be applicable to functions between metric spaces, and subsequently to functions between topological spaces.

1.2 Topological Spaces

Definition A *topological space* X consists of a set X together with a collection of subsets, referred to as *open sets*, such that the following conditions are satisfied:—

(i) the empty set $\emptyset$ and the whole set X are open sets,

(ii) the union of any collection of open sets is itself an open set,

(iii) the intersection of any *finite* collection of open sets is itself an open set.

The collection consisting of all the open sets in a topological space X is referred to as a *topology* on the set X.

Remark If it is necessary to specify explicitly the topology on a topological space then one denotes by (X, τ) the topological space whose underlying set is X and whose topology is τ. However if no confusion will arise then it is customary to denote this topological space simply by X.

1.3 Subsets of Euclidean Space

Let X be a subset of n-dimensional Euclidean space $\mathbb{R}^n$. The *Euclidean distance* $|\mathbf{x} - \mathbf{y}|$ between two points $\mathbf{x}$ and $\mathbf{y}$ of X is defined as follows:

$$|\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2},$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. The Euclidean distances between any three points $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ of X satisfy the *Triangle Inequality*:

$$|\mathbf{x} - \mathbf{z}| \leq |\mathbf{x} - \mathbf{y}| + |\mathbf{y} - \mathbf{z}|.$$

A subset V of X is said to be *open* in X if, given any point $\mathbf{v}$ of V, there exists some $\delta > 0$ such that

$$\{\mathbf{x} \in X : |\mathbf{x} - \mathbf{v}| < \delta\} \subset V.$$

The empty set is also considered to be open in X.

Both $\emptyset$ and X are open sets in X. Also it is not difficult to show that any union of open sets in X is open in X, and that any finite intersection of open sets in X is open in X. (This will be proved in more generality for open sets in metric spaces.) Thus the collection of open sets in a subset X of a Euclidean space $\mathbb{R}^n$ satisfies the topological space axioms. Thus every subset X of $\mathbb{R}^n$ is a topological space with these open sets. This topology on a subset X of $\mathbb{R}^n$ is referred to as the *usual topology* on X, generated by the Euclidean distance function.

In particular $\mathbb{R}^n$ is itself a topological space.

1.4 Open Sets in Metric Spaces

Definition A *metric space* (X, d) consists of a set X together with a *distance function* $d: X \times X \to [0, +\infty)$ on X satisfying the following axioms:

(i) $d(x, y) \geq 0$ for all $x, y \in X$,

(ii) $d(x, y) = d(y, x)$ for all $x, y \in X$,

(iii) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$,

(iv) $d(x, y) = 0$ if and only if $x = y$.

The quantity $d(x, y)$ should be thought of as measuring the *distance* between the points x and y. The inequality $d(x, z) \leq d(x, y) + d(y, z)$ is referred to as the *Triangle Inequality*. The elements of a metric space are usually referred to as *points* of that metric space.

An n-dimensional Euclidean space $\mathbb{R}^n$ is a metric space with with respect to the *Euclidean distance function* d, defined by

$$d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Any subset X of $\mathbb{R}^n$ may be regarded as a metric space whose distance function is the restriction to X of the Euclidean distance function on $\mathbb{R}^n$ defined above.

Definition Let (X, d) be a metric space. Given a point x of X and $r \geq 0$, the *open ball* $B_X(x, r)$ of *radius* r about x in X is defined by

$$B_X(x, r) \equiv \{x' \in X : d(x', x) < r\}.$$

Definition Let (X, d) be a metric space. A subset V of X is said to be an *open set* if and only if the following condition is satisfied:

- given any point v of V there exists some $\delta > 0$ such that $B_X(v, \delta) \subset V$.

By convention, we regard the empty set $\emptyset$ as being an open subset of X. (The criterion given above is satisfied vacuously in this case.)

Lemma 1.1 *Let X be a metric space with distance function d, and let x_0 be a point of X. Then, for any $r > 0$, the open ball $B_X(x_0, r)$ of radius r about x_0 is an open set in X.*

Proof Let $x \in B_X(x_0, r)$. We must show that there exists some $\delta > 0$ such that $B_X(x, \delta) \subset B_X(x_0, r)$. Now $d(x, x_0) < r$, and hence $\delta > 0$, where $\delta = r - d(x, x_0)$. Moreover if $x' \in B_X(x, \delta)$ then

$$d(x', x_0) \leq d(x', x) + d(x, x_0) < \delta + d(x, x_0) = r,$$

by the Triangle Inequality, hence $x' \in B_X(x_0, r)$. Thus $B_X(x, \delta) \subset B_X(x_0, r)$, showing that $B_X(x_0, r)$ is an open set, as required. ∎

Proposition 1.2 *Let X be a metric space. The collection of open sets in X has the following properties:—*

(i) *the empty set $\emptyset$ and the whole set X are both open sets;*

(ii) *the union of any collection of open sets is itself an open set;*

(iii) *the intersection of any* finite *collection of open sets is itself an open set.*

Proof The empty set $\emptyset$ is an open set by convention. Moreover the definition of an open set is satisfied trivially by the whole set X. Thus (i) is satisfied.

Let $\mathcal{A}$ be any collection of open sets in X, and let U denote the union of all the open sets belonging to $\mathcal{A}$. We must show that U is itself an open set. Let $x \in U$. Then $x \in V$ for some open set V belonging to the collection $\mathcal{A}$.

Therefore there exists some $\delta > 0$ such that $B_X(x, \delta) \subset V$. But $V \subset U$, and thus $B_X(x, \delta) \subset U$. This shows that U is open. Thus (ii) is satisfied.

Finally let $V_1, V_2, V_3, \ldots, V_k$ be a *finite* collection of open sets in X, and let $V = V_1 \cap V_2 \cap \cdots \cap V_k$. Let $x \in V$. Now $x \in V_j$ for all j, and therefore there exist strictly positive real numbers $\delta_1, \delta_2, \ldots, \delta_k$ such that $B_X(x, \delta_j) \subset V_j$ for $j = 1, 2, \ldots, k$. Let δ be the minimum of $\delta_1, \delta_2, \ldots, \delta_k$. Then $\delta > 0$. (This is where we need the fact that we are dealing with a finite collection of open sets.) Moreover $B_X(x, \delta) \subset B_X(x, \delta_j) \subset V_j$ for $j = 1, 2, \ldots, k$, and thus $B_X(x, \delta) \subset V$. This shows that the intersection V of the open sets $V_1, V_2, \ldots, V_k$ is itself open. Thus (iii) is satisfied. ∎

Any metric space may be regarded as a topological space. Indeed let X be a metric space with distance function d. We recall that a subset V of X is an *open set* if and only if, given any point v of V, there exists some $\delta > 0$ such that $\{x \in X : d(x, v) < \delta\} \subset V$. Proposition 1.2 shows that the topological space axioms are satisfied by the collection of open sets in any metric space. We refer to this collection of open sets as the *topology* generated by the distance function d on X.

1.5 Further Examples of Topological Spaces

Example Given any set X, one can define a topology on X where every subset of X is an open set. This topology is referred to as the *discrete topology* on X.

Example Given any set X, one can define a topology on X in which the only open sets are the empty set $\emptyset$ and the whole set X.

1.6 Closed Sets

Definition Let X be a topological space. A subset F of X is said to be a *closed set* if and only if its complement $X \setminus F$ is an open set.

We recall that the complement of the union of some collection of subsets of some set X is the intersection of the complements of those sets, and the complement of the intersection of some collection of subsets of X is the union of the complements of those sets. The following result therefore follows directly from the definition of a topological space.

Proposition 1.3 *Let X be a topological space. Then the collection of closed sets of X has the following properties:—*

(i) *the empty set $\emptyset$ and the whole set X are closed sets,*

(ii) *the intersection of any collection of closed sets is itself a closed set,*

(iii) *the union of any finite collection of closed sets is itself a closed set.*

1.7 Hausdorff Spaces

Definition A topological space X is said to be a *Hausdorff space* if and only if it satisfies the following *Hausdorff Axiom*:

- if x and y are distinct points of X then there exist open sets U and V such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$.

Lemma 1.4 *All metric spaces are Hausdorff spaces.*

Proof Let X be a metric space with distance function d, and let x and y be points of X, where $x \neq y$. Let $\varepsilon = \frac{1}{2}d(x,y)$. Then the open balls $B_X(x,\varepsilon)$ and $B_X(y,\varepsilon)$ of radius ε centred on the points x and y are open sets (see Lemma 1.1). If $B_X(x,\varepsilon) \cap B_X(y,\varepsilon)$ were non-empty then there would exist $z \in X$ satisfying $d(x,z) < \varepsilon$ and $d(z,y) < \varepsilon$. But this is impossible, since it would then follow from the Triangle Inequality that $d(x,y) < 2\varepsilon$, contrary to the choice of ε. Thus $x \in B_X(x,\varepsilon)$, $y \in B_X(y,\varepsilon)$, $B_X(x,\varepsilon) \cap B_X(y,\varepsilon) = \emptyset$. This shows that the metric space X is a Hausdorff space. ∎

We now give an example of a topological space which is not a Hausdorff space.

Example The *Zariski topology* on the set $\mathbb{R}$ of real numbers is defined as follows: a subset U of $\mathbb{R}$ is open (with respect to the Zariski topology) if and only if either $U = \emptyset$ or else $\mathbb{R} \setminus U$ is finite. It is a straightforward exercise to verify that the topological space axioms are satisfied, so that the set $\mathbb{R}$ of real numbers is a topological space with respect to this Zariski topology. Now the intersection of any two non-empty open sets in this topology is always non-empty. (Indeed if U and V are non-empty open sets then $U = \mathbb{R} \setminus F_1$ and $V = \mathbb{R} \setminus F_2$, where F_1 and F_2 are finite sets of real numbers. But then $U \cap V = \mathbb{R} \setminus (F_1 \cup F_2)$, which is non-empty, since $F_1 \cup F_2$ is finite and $\mathbb{R}$ is infinite.) It follows immediately from this that $\mathbb{R}$, with the Zariski topology, is not a Hausdorff space.

1.8 Subspace Topologies

Let X be a topological space with topology τ, and let A be a subset of X. Let τ_A be the collection of all subsets of A that are of the form $V \cap A$ for $V \in \tau$. Then τ_A is a topology on the set A. (It is a straightforward exercise to verify that the topological space axioms are satisfied.) The topology τ_A on A is referred to as the *subspace topology* on A.

Any subset of a Hausdorff space is itself a Hausdorff space (with respect to the subspace topology).

Lemma 1.5 *Let X be a metric space with distance function d, and let A be a subset of X. A subset W of A is open with respect to the subspace topology on A if and only if, given any point w of W, there exists some $\delta > 0$ such that*

$$\{a \in A : d(a, w) < \delta\} \subset W.$$

Thus the subspace topology on A coincides with the topology on A obtained on regarding A as a metric space (with respect to the distance function d).

Proof Suppose that W is open with respect to the subspace topology on A. Then there exists some open set U in X such that $W = U \cap A$. Let w be a point of W. Then there exists some $\delta > 0$ such that

$$\{x \in X : d(x, w) < \delta\} \subset U.$$

But then

$$\{a \in A : d(a, w) < \delta\} \subset U \cap A = W.$$

Conversely, suppose that W is a subset of A with the property that, for any $w \in W$, there exists some $\delta_w > 0$ such that

$$\{a \in A : d(a, w) < \delta_w\} \subset W.$$

Define U to be the union of the open balls $B_X(w, \delta_w)$ as w ranges over all points of W, where

$$B_X(w, \delta_w) = \{x \in X : d(x, w) < \delta_w\}.$$

The set U is an open set in X, since each open ball $B_X(w, \delta_w)$ is an open set in X (Lemma 1.1), and any union of open sets is itself an open set. Moreover

$$B_X(w, \delta_w) \cap A = \{a \in A : d(a, w) < \delta_w\} \subset W$$

for any $w \in W$. Therefore $U \cap A \subset W$. However $W \subset U \cap A$, since, $W \subset A$ and $\{w\} \subset B_X(w, \delta_w) \subset U$ for any $w \in W$. Thus $W = U \cap A$, where U is an open set in X. We deduce that W is open with respect to the subspace topology on A. ∎

Example Let X be any subset of n-dimensional Euclidean space $\mathbb{R}^n$. Then the subspace topology on X coincides with the topology on X generated by the Euclidean distance function on X. We refer to this topology as the *usual topology* on X.

Let X be a topological space, and let A be a subset of X. One can readily verify the following:—

- a subset B of A is closed in A (relative to the subspace topology on A) if and only if $B = A \cap F$ for some closed subset F of X;
- if A is itself open in X then a subset B of A is open in A if and only if it is open in X;
- if A is itself closed in X then a subset B of A is closed in A if and only if it is closed in X.

1.9 Continuous Functions between Topological Spaces

Definition A function $f: X \to Y$ from a topological space X to a topological space Y is said to be *continuous* if $f^{-1}(V)$ is an open set in X for every open set V in Y, where

$$f^{-1}(V) \equiv \{x \in X : f(x) \in V\}.$$

A continuous function from X to Y is often referred to as a *map* from X to Y.

Lemma 1.6 *Let X, Y and Z be topological spaces, and let $f: X \to Y$ and $g: Y \to Z$ be continuous functions. Then the composition $g \circ f: X \to Z$ of the functions f and g is continuous.*

Proof Let V be an open set in Z. Then $g^{-1}(V)$ is open in Y (since g is continuous), and hence $f^{-1}(g^{-1}(V))$ is open in X (since f is continuous). But $f^{-1}(g^{-1}(V)) = (g \circ f)^{-1}(V)$. Thus the composition function $g \circ f$ is continuous. ∎

Lemma 1.7 *Let X and Y be topological spaces, and let $f: X \to Y$ be a function from X to Y. The function f is continuous if and only if $f^{-1}(G)$ is closed in X for every closed subset G of Y.*

Proof If G is any subset of Y then $X \setminus f^{-1}(G) = f^{-1}(Y \setminus G)$ (i.e., the complement of the preimage of G is the preimage of the complement of G). The result therefore follows immediately from the definitions of continuity and closed sets. ∎

1.10 Continuous Functions between Metric Spaces

The following definition of continuity for functions between metric spaces generalizes that for functions of a real or complex variable.

Definition Let X and Y be metric spaces with distance functions d_X and d_Y respectively. A function $f\colon X \to Y$ from X to Y is said to be *continuous* at a point x of X if and only if the following criterion is satisfied:—

- given any real number ε satisfying $\varepsilon > 0$ there exists some $\delta > 0$ such that $d_Y(f(x), f(x')) < \varepsilon$ for all points x' of X satisfying $d_X(x, x') < \delta$.

The function $f\colon X \to Y$ is said to be continuous on X if and only if it is continuous at x for every point x of X.

This definition can be rephrased in terms of open balls: a function $f\colon X \to Y$ from a metric space X to a metric space Y is continuous at a point x of X if and only if, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that f maps $B_X(x, \delta)$ into $B_Y(f(x), \varepsilon)$ (where $B_X(x, \delta)$ and $B_Y(f(x), \varepsilon)$ denote the open balls of radius δ and ε about x and $f(x)$ respectively).

Let $f\colon X \to Y$ be a function from a set X to a set Y. Given any subset V of Y, we denote by $f^{-1}(V)$ the *preimage* of V under the map f, defined by

$$f^{-1}(V) = \{x \in X : f(x) \in V\}.$$

The following result shows that the definition of continuity given above for functions between metric spaces is consistent with the more general definition of continuity for functions between topological spaces.

Proposition 1.8 *Let X and Y be metric spaces, and let $f\colon X \to Y$ be a function from X to Y. The function f is continuous if and only if $f^{-1}(V)$ is an open set in X for every open set V of Y.*

Proof Suppose that $f\colon X \to Y$ is continuous. Let V be an open set in Y. We must show that $f^{-1}(V)$ is open in X. Let x be a point belonging to $f^{-1}(V)$. We must show that there exists some $\delta > 0$ with the property that $B_X(x, \delta) \subset f^{-1}(V)$. Now $f(x)$ belongs to V. But V is open, hence there exists some $\varepsilon > 0$ with the property that $B_Y(f(x), \varepsilon) \subset V$. But f is continuous at x. Therefore there exists some $\delta > 0$ such that f maps the open ball $B_X(x, \delta)$ into $B_Y(f(x), \varepsilon)$ (see the remarks above). Thus $f(x') \in V$ for all $x' \in B_X(x, \delta)$, showing that $B_X(x, \delta) \subset f^{-1}(V)$. We have thus shown that if $f\colon X \to Y$ is continuous then $f^{-1}(V)$ is open in X for every open set V in Y.

Conversely suppose that $f: X \to Y$ has the property that $f^{-1}(V)$ is open in X for every open set V in Y. Let x be any point of X. We must show that f is continuous at x. Let $\varepsilon > 0$ be given. The open ball $B_Y(f(x), \varepsilon)$ is an open set in Y, by Lemma 1.1, hence $f^{-1}(B_Y(f(x), \varepsilon))$ is an open set in X which contains x. It follows that there exists some $\delta > 0$ such that $B_X(x, \delta) \subset f^{-1}(B_Y(f(x), \varepsilon))$. We have thus shown that, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that f maps the open ball $B_X(x, \delta)$ into $B_Y(f(x), \varepsilon)$. We conclude that f is continuous at x, as required. ∎

1.11 A Criterion for Continuity

We now show that, if a topological space X is the union of a finite collection of closed sets, and if a function from X to some topological space is continuous on each of these closed sets, then that function is continuous on X.

Lemma 1.9 *Let X and Y be topological spaces, let $f: X \to Y$ be a function from X to Y, and let $X = A_1 \cup A_2 \cup \cdots \cup A_k$, where $A_1, A_2, \ldots, A_k$ are closed sets in X. Suppose that the restriction of f to the closed set A_i is continuous for $i = 1, 2, \ldots, k$. Then $f: X \to Y$ is continuous.*

Proof Let V be an open set in Y. We must show that $f^{-1}(V)$ is open in X. Now the preimage of the open set V under the restriction $f|A_i$ of f to A_i is $f^{-1}(V) \cap A_i$. It follows from the continuity of $f|A_i$ that $f^{-1}(V) \cap A_i$ is relatively open in A_i for each i, and hence there exist open sets $U_1, U_2, \ldots, U_k$ in X such that $f^{-1}(V) \cap A_i = U_i \cap A_i$ for $i = 1, 2, \ldots, k$. Let $W_i = U_i \cup (X \setminus A_i)$ for $i = 1, 2, \ldots, k$. Then W_i is an open set in X (as it is the union of the open sets U_i and $X \setminus A_i$), and $W_i \cap A_i = U_i \cap A_i = f^{-1}(V) \cap A_i$ for each i. We claim that $f^{-1}(V) = W_1 \cap W_2 \cap \cdots \cap W_k$.

Let $W = W_1 \cap W_2 \cap \cdots \cap W_k$. Then $f^{-1}(V) \subset W$, since $f^{-1}(V) \subset W_i$ for each i. Also

$$W = \bigcup_{i=1}^{k} (W \cap A_i) \subset \bigcup_{i=1}^{k} (W_i \cap A_i) = \bigcup_{i=1}^{k} (f^{-1}(V) \cap A_i) \subset f^{-1}(V),$$

since $X = A_1 \cup A_2 \cup \cdots \cup A_k$ and $W_i \cap A_i = f^{-1}(V) \cap A_i$ for each i. Therefore $f^{-1}(V) = W$. But W is open in X, since it is the intersection of a finite collection of open sets. We have thus shown that $f^{-1}(V)$ is open in X for any open set V in Y. Thus $f: X \to Y$ is continuous, as required. ∎

Alternative Proof A function $f: X \to Y$ is continuous if and only if $f^{-1}(G)$ is closed in X for every closed set G in Y (Lemma 1.7). Let G be an closed set in Y. Then $f^{-1}(G) \cap A_i$ is relatively closed in A_i for $i = 1, 2, \ldots, k$, since the restriction of f to A_i is continuous for each i. But A_i is closed in X, and therefore a subset of A_i is relatively closed in A_i if and only if it is closed in X. Therefore $f^{-1}(G) \cap A_i$ is closed in X for $i = 1, 2, \ldots, k$. Now $f^{-1}(G)$ is the union of the sets $f^{-1}(G) \cap A_i$ for $i = 1, 2, \ldots, k$. It follows that $f^{-1}(G)$, being a finite union of closed sets, is itself closed in X. It now follows from Lemma 1.7 that $f: X \to Y$ is continuous. ∎

Example Let Y be a topological space, and let $\alpha: [0,1] \to Y$ and $\beta: [0,1] \to Y$ be continuous functions defined on the interval $[0,1]$, where $\alpha(1) = \beta(0)$. Let $\gamma: [0,1] \to Y$ be defined by

$$\gamma(t) = \begin{cases} \alpha(2t) & \text{if } 0 \leq t \leq \frac{1}{2}; \\ \beta(2t-1) & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

Now $\gamma|[0, \frac{1}{2}] = \alpha \circ \rho$ where $\rho: [0, \frac{1}{2}] \to [0,1]$ is the continuous function defined by $\rho(t) = 2t$ for all $t \in [0, \frac{1}{2}]$. Thus $\gamma|[0, \frac{1}{2}]$ is continuous, being a composition of two continuous functions. Similarly $\gamma|[\frac{1}{2}, 1]$ is continuous. The subintervals $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$ are closed in $[0,1]$, and $[0,1]$ is the union of these two subintervals. It follows from Lemma 1.9 that $\gamma: [0,1] \to Y$ is continuous.

1.12 Homeomorphisms

Definition Let X and Y be topological spaces. A function $h: X \to Y$ is said to be a *homeomorphism* if and only if the following conditions are satisfied:

- the function $h: X \to Y$ is both injective and surjective (so that the function $h: X \to Y$ has a well-defined inverse $h^{-1}: Y \to X$),
- the function $h: X \to Y$ and its inverse $h^{-1}: Y \to X$ are both continuous.

Two topological spaces X and Y are said to be *homeomorphic* if there exists a homeomorphism $h: X \to Y$ from X to Y.

If $h: X \to Y$ is a homeomorphism between topological spaces X and Y then h induces a one-to-one correspondence between the open sets of X and the open sets of Y. Thus the topological spaces X and Y can be regarded as being identical as topological spaces.

1.13 Neighbourhoods, Closures and Interiors

Definition Let X be a topological space, and let x be a point of X. Let N be a subset of X which contains the point x. Then N is said to be a *neighbourhood* of the point x if and only if there exists an open set U for which $x \in U$ and $U \subset N$.

One can readily verify that this definition of neighbourhoods in topological spaces is consistent with that for neighbourhoods in metric spaces.

Lemma 1.10 *Let X be a topological space. A subset V of X is open in X if and only if V is a neighbourhood of each point belonging to V.*

Proof It follows directly from the definition of neighbourhoods that an open set V is a neighbourhood of any point belonging to V. Conversely, suppose that V is a subset of X which is a neighbourhood of each $v \in V$. Then, given any point v of V, there exists an open set U_v such that $v \in U_v$ and $U_v \subset V$. Thus V is an open set, since it is the union of the open sets U_v as v ranges over all points of V. ∎

Definition Let X be a topological space and let A be a subset of X. The *closure* $\overline{A}$ of A in X is defined to be the intersection of all of the closed subsets of X that contain A. The *interior* A^0 of A in X is defined to be the union of all of the open subsets of X that are contained in A.

Let X be a topological space and let A be a subset of X. It follows directly from the definition of $\overline{A}$ that the closure $\overline{A}$ of A is uniquely characterized by the following two properties:

(i) the closure $\overline{A}$ of A is a closed set containing A,

(ii) if F is any closed set containing A then F contains $\overline{A}$.

Similarly the interior A^0 of A is uniquely characterized by the following two properties:

(i) the interior A^0 of A is an open set contained in A,

(ii) if U is any open set contained in A then U is contained in A^0.

Moreover a point x of A belongs to the interior A^0 of A if and only if A is a neighbourhood of x.

1.14 Bases for Topologies

Proposition 1.11 *Let X be a set, let β be a collection of subsets of X, and let τ be the collection consisting of the empty set, together with all subsets of X that are unions of sets belonging to the collection β. Then τ is a topology on X if and only if the following conditions are satisfied:—*

(i) *the set X is the union of the subsets belonging to the collection β;*

(ii) *given subsets* $B_1, B_2 \in \beta$*, and given any point* p *of* $B_1 \cap B_2$*, there exists some* $B \in \beta$ *such that* $p \in B$ *and* $B \subset B_1 \cap B_2$*.*

Proof First suppose that τ is a topology on X. Then $X \in \tau$. But any subset of X that belongs to τ is a union of sets belonging to β. Therefore X is a union of subsets belonging to the collection β, and thus condition (i) is satisfied.

Moreover the intersection of any two open subsets of a topological space is required to be open. Thus if τ is a topology on X, and if $B_1, B_2 \in \beta$, then $B_1, B_2 \in \tau$ and therefore $B_1 \cap B_2 \in \tau$. It follows that $B_1 \cap B_2$ is a union of subsets of X that belong to β, and therefore, given any $p \in B_1 \cap B_2$, there exists $B \in \beta$ such that $p \in B$ and $B \subset B_1 \cap B_2$. Thus condition (ii) is satisfied.

Conversely we must prove that if the collection β of subsets of a set X satisfies conditions (i) and (ii) then the collection τ of unions of sets belonging to β is a topology on X.

The empty set belongs to τ. Condition (ii) ensures that the whole set X belongs to τ. It follows directly from the definition of τ that any union of sets belonging to τ is a union of sets belonging to β, and therefore itself belongs to τ.

It therefore only remains to show that the intersection of any finite collection of sets belonging to τ belongs to τ. It suffices to prove that the intersection of two sets belonging to τ belongs to τ. Let $V_1, V_2 \in \tau$, and let $p \in V_1 \cap V_2$. Then V_1 and V_2 are union of sets belonging to β, and therefore there exist $B_1, B_2 \in \beta$ such that $p \in B_1$, $p \in B_2$, $B_1 \subset V_1$, and $B_2 \subset V_2$. Now condition (ii) ensures the existence of $B_p \in \beta$ such that $p \in B_p$ and $B_p \subset B_1 \cap B_2$. Then $B_p \subset B_1 \subset V_1$ and $B_p \subset B_2 \subset V_2$. We have thus shown that, given any point p of $V_1 \cap V_2$, there exists some member B_p of the collection β such that $p \in B_p$ and $B_p \subset V_1 \cap V_2$. It follows that $V_1 \cap V_2$ is the union of the sets B_p as p ranges over all points of the intersection $V_1 \cap V_2$, and therefore $V_1 \cap V_2 \in \tau$. We have thus proved that the intersection of two sets belonging to τ belongs to τ. It follows directly by induction on the number of sets involved that the intersection of any finite collection of sets belonging to τ must belong to τ. Thus τ is a topology on the set X, as required. ∎

Definition Let X be a set. A collection β of subsets of X is said to be a *basis* for a topology on X if the following conditions are satisfied:—

(i) the set X is the union of the subsets belonging to the collection β;

(ii) given subsets $B_1, B_2 \in \beta$, and given any point p of $B_1 \cap B_2$, there exists some $B \in \beta$ such that $p \in B$ and $B \subset B_1 \cap B_2$.

If β is a basis for a topology on X then the topology generated by β is the topology whose open sets are those subsets of X that are unions of sets belonging to the basis β.

Lemma 1.12 *Let X be a set, and let β be a basis for a topology on X. A non-empty subset V is open in X with respect to the topology generated by β if and only if, given any point v of V, there exists $B \in \beta$ such that $v \in B$ and $B \subset V$.*

Proof This result follows directly from the fact that the non-empty open sets in X are those subsets of X that are unions of sets belonging to the basis β. ∎

Example Let X be a metric space. Then the collection of open balls in X (determined by the distance function d on X) is a basis for a topology on X. Indeed let x and y be points of X that are the centres of open balls $B_X(x,r)$ and $B_X(y,s)$ of radii r and s respectively, and let $z \in B_X(x,r) \cap B_X(y,s)$. Then $d(z,x) < r$ and $d(z,y) < s$. Let δ be the minimum of $r - d(z,x)$ and $s - d(z,y)$. Then $\delta > 0$ and $z \in B_X(z,\delta)$, and it follows from the Triangle Inequality that $B_X(z,\delta) \subset B_X(x,r) \cap B_X(y,s)$. This ensures that the collection of open balls in this metric space is a basis for a topology on the metric space. The topology generated by this basis is the standard topology carried by the metric space when we consider metric spaces to be topological spaces.

The following result is an immediate consequence of the requirements for a collection of subsets of a given set to be a basis for a topology of that set.

Lemma 1.13 *Let X be a topological space, and let β be a collection of subsets of X. Suppose that $X \in \beta$ and that the intersection of any finite collection of sets belonging to β itself belongs to β. Then β is a basis for a topology on X.*

1.15 Subbases for Topologies

Let X be a set, and let σ be a collection of subsets of X. Let β be the collection of subsets of X consisting of the set X itself together with all subsets of X that are finite intersections of sets belonging to the collection σ. It follows from an immediate application of Lemma 1.13 that β is a basis for a topology τ on X. We refer to τ as the topology on X *generated* by the collection σ of subsets of X, and we refer to the collection σ as a *subbasis* for the topology τ which it generates.

Lemma 1.14 *Let σ be a collection of subsets of a set X which is a subbasis for a topology τ on X, and let $\tilde{\tau}$ be a topology on X. Suppose that $\sigma \subset \tilde{\tau}$. Then $\tau \subset \tilde{\tau}$. Thus the topology τ generated by the subbasis σ is the smallest topology on X that contains σ.*

Proof The definition of topologies ensures that $X \in \tilde{\tau}$. It also ensures that any finite intersection of subsets of X that belong to $\tilde{\tau}$ must itself belong to $\tilde{\tau}$. It follows that if $\sigma \subset \tilde{\tau}$ then $\beta \subset \tilde{\tau}$, where β is the collection of subsets of X consisting of the set X itself together with all subsets of X that are finite intersections of sets belonging to the collection σ. Now every non-empty subset of X that belongs to the topology τ is a union of subsets of X that belong to β. These subsets belong to $\tilde{\tau}$, and any union of subsets of X belonging to $\tilde{\tau}$ must itself belong to $\tilde{\tau}$. Therefore $\tau \subset \tilde{\tau}$. This shows that the topology τ generated by a subbasis σ is the smallest topology on X that contains the members of the subbasis σ. ∎

It follows immediately from Lemma 1.14 that the topology τ on a set X generated by a subbasis σ is the intersection of all topologies $\tilde{\tau}$ on X for which $\sigma \subset \tilde{\tau}$.

1.16 Product Topologies

The *Cartesian product* $X_1 \times X_2 \times \cdots \times X_n$ of sets $X_1, X_2, \ldots, X_n$ is defined to be the set of all ordered n-tuples $(x_1, x_2, \ldots, x_n)$, where $x_i \in X_i$ for $i = 1, 2, \ldots, n$.

The sets $\mathbb{R}^2$ and $\mathbb{R}^3$ are the Cartesian products $\mathbb{R} \times \mathbb{R}$ and $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ respectively.

Lemma 1.15 *Let $X_1, X_2, \ldots, X_n$ be topological spaces, let X be the Cartesian product $X_1 \times X_2 \times \cdots \times X_n$ of the sets $X_1, X_2, \ldots, X_n$, and let β be the collection consisting of all subsets of the Cartesian product X that are of the form $V_1 \times V_2 \times \cdots \times V_n$, where V_j is an open set in X_j for $j = 1, 2, \ldots, n$. Then β is a basis for a topology on X.*

Proof The definition of topological spaces ensures that each of the sets $X_1, X_2, \ldots, X_n$ is an open set in itself. It follows that $X \in \beta$.

Let $U_1, U_2, \ldots, U_m$ be subsets of X, where $U_i \in \beta$ for $i = 1, 2, \ldots, m$. Then there exist subsets $V_{i,j}$ of X_j, where $V_{i,j}$ is an open set in X_j for $j = 1, 2, \ldots, n$ and $i = 1, 2, \ldots, n$, such that

$$U_i = V_{i,1} \times V_{i,2} \times \cdots \times V_{i,n} \quad \text{for } i = 1, 2, \ldots, m.$$

Then

$$\bigcap_{i=1}^{m} U_i = V_1 \times V_2 \times \cdots \times V_n$$

where

$$V_j = \bigcap_{i=1}^{m} V_{i,j} \quad \text{for } j = 1, 2, \ldots, n.$$

Indeed let $(x_1, x_2, \ldots, x_n) \in X$, where $x_j \in X_j$ for $j = 1, 2, \ldots, n$. Then

$$\begin{aligned}
(x_1, x_2, \ldots, x_n) \in \bigcap_{i=1}^{m} U_i \quad & \iff \quad (x_1, x_2, \ldots, x_n) \in U_i \text{ for } i = 1, 2, \ldots, m \\
& \iff \quad x_j \in V_{i,j} \text{ for } i = 1, 2, \ldots, m \\
& \qquad\quad \text{and } j = 1, 2, \ldots, n \\
& \iff \quad x_j \in \bigcap_{i=1}^{m} V_{i,j} \text{ for } j = 1, 2, \ldots, n \\
& \iff \quad x_j \in V_j \text{ for } j = 1, 2, \ldots, n \\
& \iff \quad (x_1, x_2, \ldots, x_n) \in V_1 \times V_2 \times \cdots \times V_n.
\end{aligned}$$

But V_j is an open set in X_j for $j = 1, 2, \ldots, n$, since any finite intersection of open sets in a topological space must itself be an open set in that space, and therefore $\bigcap_{i=1}^{m} U_i \in \beta$. We have now shown that $X \in \beta$, and also that any finite intersection of subsets of X that belong to β must itself belong to β. It now follows from Lemma 1.13 that β is a basis for a topology on X. ∎

Definition Let $X_1, X_2, \ldots, X_n$ be topological spaces. The *product topology* on the Cartesian product $X_1 \times X_2 \times \cdots \times X_n$ is the topology on this Cartesian product of sets that is generated by the collection consisting of all subsets of the product set that are of the form

$$V_1 \times V_2 \times \cdots \times V_n,$$

where V_j is an open set in X_j for $j = 1, 2, \ldots, n$.

Lemma 1.16 *Let $X_1, X_2, \ldots, X_n$ be topological spaces. A non-empty subset U of the Cartesian product $X_1 \times X_2 \times \cdots \times X_n$ is open with respect to the product topology if and only if, given any point p of U, there exist open sets V_i in X_i for $i = 1, 2, \ldots, n$ such that*

$$p \in V_1 \times V_2 \times \cdots \times V_n \quad \textit{and} \quad V_1 \times V_2 \times \cdots \times V_n \subset U.$$

Proof The product topology is the topology generated by the collection β consisting of those subsets of the Cartesian product that are of the form $V_1 \times V_2 \times \cdots \times V_n$, where V_j is an open set in X_j for $j = 1, 2, \ldots, n$. This collection of subsets is a basis for the product topology (Lemma 1.15). It follows that a non-empty subset of X is open with respect to the product topology if and only if it is a union of sets that belong to the basis β. The result follows directly (see Lemma 1.12). ∎

Lemma 1.17 *Let $X_1, X_2, \ldots, X_n$ and Z be topological spaces. Then a function $f\colon X_1 \times X_2 \times \cdots \times X_n \to Z$ is continuous if and only if, given any point p of $X_1 \times X_2 \times \cdots \times X_n$, and given any open set U in Z containing $f(p)$, there exist open sets V_i in X_i for $i = 1, 2, \ldots, n$ such that $p \in V_1 \times V_2 \cdots \times V_n$ and $f(V_1 \times V_2 \times \cdots \times V_n) \subset U$.*

Proof Let V_i be an open set in X_i for $i = 1, 2, \ldots, n$, and let U be an open set in Z. Then $V_1 \times V_2 \times \cdots \times V_n \subset f^{-1}(U)$ if and only if $f(V_1 \times V_2 \times \cdots \times V_n) \subset U$. It follows that $f^{-1}(U)$ is open in the product topology on $X_1 \times X_2 \times \cdots \times X_n$ if and only if, given any point p of $X_1 \times X_2 \times \cdots \times X_n$ satisfying $f(p) \in U$, there exist open sets V_i in X_i for $i = 1, 2, \ldots, n$ such that $f(V_1 \times V_2 \times \cdots \times V_n) \subset U$. The required result now follows from the definition of continuity. ∎

Let $X_1, X_2, \ldots, X_n$ be topological spaces, and let V_i be an open set in X_i for $i = 1, 2, \ldots, n$. It follows directly from the definition of the product topology that $V_1 \times V_2 \times \cdots \times V_n$ is open in $X_1 \times X_2 \times \cdots \times X_n$.

Theorem 1.18 *Let $X = X_1 \times X_2 \times \cdots \times X_n$, where $X_1, X_2, \ldots, X_n$ are topological spaces and X is given the product topology, and for each i, let $p_i\colon X \to X_i$ denote the projection function which sends $(x_1, x_2, \ldots, x_n) \in X$ to x_i. Then the functions $p_1, p_2, \ldots, p_n$ are continuous. Moreover a function $f\colon Z \to X$ mapping a topological space Z into X is continuous if and only if $p_i \circ f\colon Z \to X_i$ is continuous for $i = 1, 2, \ldots, n$.*

Proof Let V be an open set in X_i. Then

$$p_i^{-1}(V) = X_1 \times \cdots \times X_{i-1} \times V \times X_{i+1} \times \cdots \times X_n,$$

and therefore $p_i^{-1}(V)$ is open in X. Thus $p_i\colon X \to X_i$ is continuous for all i.

Let $f\colon Z \to X$ be continuous. Then, for each i, $p_i \circ f\colon Z \to X_i$ is a composition of continuous functions, and is thus itself continuous.

Conversely suppose that $f\colon Z \to X$ is a function with the property that $p_i \circ f$ is continuous for all i. Let U be an open set in X. We must show that $f^{-1}(U)$ is open in Z.

Let z be a point of $f^{-1}(U)$, and let $f(z) = (u_1, u_2, \ldots, u_n)$. Now U is open in X, and therefore there exist open sets $V_1, V_2, \ldots, V_n$ in $X_1, X_2, \ldots, X_n$ respectively such that $u_i \in V_i$ for all i and $V_1 \times V_2 \times \cdots \times V_n \subset U$. Let

$$N_z = f_1^{-1}(V_1) \cap f_2^{-1}(V_2) \cap \cdots \cap f_n^{-1}(V_n),$$

where $f_i = p_i \circ f$ for $i = 1, 2, \ldots, n$. Now $f_i^{-1}(V_i)$ is an open subset of Z for $i = 1, 2, \ldots, n$, since V_i is open in X_i and $f_i \colon Z \to X_i$ is continuous. Thus N_z, being a finite intersection of open sets, is itself open in Z. Moreover

$$f(N_z) \subset V_1 \times V_2 \times \cdots \times V_n \subset U,$$

so that $N_z \subset f^{-1}(U)$. It follows that $f^{-1}(U)$ is the union of the open sets N_z as z ranges over all points of $f^{-1}(U)$. Therefore $f^{-1}(U)$ is open in Z. This shows that $f \colon Z \to X$ is continuous, as required. ∎

Proposition 1.19 *The usual topology on $\mathbb{R}^n$ coincides with the product topology on $\mathbb{R}^n$ obtained on regarding $\mathbb{R}^n$ as the Cartesian product $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ of n copies of the real line $\mathbb{R}$.*

Proof We must show that a subset U of $\mathbb{R}^n$ is open with respect to the usual topology if and only if it is open with respect to the product topology.

Let U be a subset of $\mathbb{R}^n$ that is open with respect to the usual topology, and let $\mathbf{u} \in U$. Then there exists some $\delta > 0$ such that $B(\mathbf{u}, \delta) \subset U$, where

$$B(\mathbf{u}, \delta) = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{u}| < \delta\}.$$

Let $I_1, I_2, \ldots, I_n$ be the open intervals in $\mathbb{R}$ defined by

$$I_i = \{t \in \mathbb{R} : u_i - \frac{\delta}{\sqrt{n}} < t < u_i + \frac{\delta}{\sqrt{n}}\} \qquad (i = 1, 2, \ldots, n),$$

Then $I_1, I_2, \ldots, I_n$ are open sets in $\mathbb{R}$. Moreover

$$\{\mathbf{u}\} \subset I_1 \times I_2 \times \cdots \times I_n \subset B(\mathbf{u}, \delta) \subset U,$$

since

$$|\mathbf{x} - \mathbf{u}|^2 = \sum_{i=1}^{n} (x_i - u_i)^2 < n \left(\frac{\delta}{\sqrt{n}}\right)^2 = \delta^2$$

for all $\mathbf{x} \in I_1 \times I_2 \times \cdots \times I_n$. This shows that any subset U of $\mathbb{R}^n$ that is open with respect to the usual topology on $\mathbb{R}^n$ is also open with respect to the product topology on $\mathbb{R}^n$.

Conversely suppose that U is a subset of $\mathbb{R}^n$ that is open with respect to the product topology on $\mathbb{R}^n$, and let $\mathbf{u} \in U$. Then there exist open sets $V_1, V_2, \ldots, V_n$ in $\mathbb{R}$ containing $u_1, u_2, \ldots, u_n$ respectively such that $V_1 \times V_2 \times \cdots \times V_n \subset U$. Now we can find $\delta_1, \delta_2, \ldots, \delta_n$ such that $\delta_i > 0$ and $(u_i - \delta_i, u_i + \delta_i) \subset V_i$ for all i. Let $\delta > 0$ be the minimum of $\delta_1, \delta_2, \ldots, \ldots, \delta_n$. Then

$$B(\mathbf{u}, \delta) \subset V_1 \times V_2 \times \cdots V_n \subset U,$$

for if $\mathbf{x} \in B(\mathbf{u}, \delta)$ then $|x_i - u_i| < \delta_i$ for $i = 1, 2, \ldots, n$. This shows that any subset U of $\mathbb{R}^n$ that is open with respect to the product topology on $\mathbb{R}^n$ is also open with respect to the usual topology on $\mathbb{R}^n$. ∎

The following result is now an immediate corollary of Proposition 1.19 and Theorem 1.18.

Corollary 1.20 *Let X be a topological space and let $f: X \to \mathbb{R}^n$ be a function from X to $\mathbb{R}^n$. Let us write*

$$f(x) = (f_1(x), f_2(x), \ldots, f_n(x))$$

for all $x \in X$, where the components $f_1, f_2, \ldots, f_n$ of f are functions from X to $\mathbb{R}$. The function f is continuous if and only if its components $f_1, f_2, \ldots, f_n$ are all continuous.

Let $f: X \to \mathbb{R}$ and $g: X \to \mathbb{R}$ be continuous real-valued functions on some topological space X. We claim that $f + g$, $f - g$ and $f.g$ are continuous. Now it is a straightforward exercise to verify that the sum and product functions $s: \mathbb{R}^2 \to \mathbb{R}$ and $p: \mathbb{R}^2 \to \mathbb{R}$ defined by $s(x, y) = x + y$ and $p(x, y) = xy$ are continuous, and $f + g = s \circ h$ and $f.g = p \circ h$, where $h: X \to \mathbb{R}^2$ is defined by $h(x) = (f(x), g(x))$. Moreover it follows from Corollary 1.20 that the function h is continuous, and compositions of continuous functions are continuous. Therefore $f + g$ and $f.g$ are continuous, as claimed. Also $-g$ is continuous, and $f - g = f + (-g)$, and therefore $f - g$ is continuous. If in addition the continuous function g is non-zero everywhere on X then $1/g$ is continuous (since $1/g$ is the composition of g with the reciprocal function $t \mapsto 1/t$), and therefore f/g is continuous.

Lemma 1.21 *The Cartesian product $X_1 \times X_2 \times \ldots X_n$ of Hausdorff spaces $X_1, X_2, \ldots, X_n$ is Hausdorff.*

Proof Let $X = X_1 \times X_2 \times \ldots, X_n$, and let u and v be distinct points of X, where $u = (x_1, x_2, \ldots, x_n)$ and $v = (y_1, y_2, \ldots, y_n)$. Then $x_i \neq y_i$ for some integer i between 1 and n. But then there exist open sets U and V in X_i

such that $x_i \in U$, $y_i \in V$ and $U \cap V = \emptyset$ (since X_i is a Hausdorff space). Let $p_i: X \to X_i$ denote the projection function. Then $p_i^{-1}(U)$ and $p_i^{-1}(V)$ are open sets in X, since p_i is continuous. Moreover $u \in p_i^{-1}(U)$, $v \in p_i^{-1}(V)$, and $p_i^{-1}(U) \cap p_i^{-1}(V) = \emptyset$. Thus X is Hausdorff, as required. ∎

1.17 Identification Maps and Quotient Topologies

Definition Let X and Y be topological spaces and let $q: X \to Y$ be a function from X to Y. The function q is said to be an *identification map* if and only if the following conditions are satisfied:

- the function $q: X \to Y$ is surjective,
- a subset U of Y is open in Y if and only if $q^{-1}(U)$ is open in X.

It follows directly from the definition that any identification map is continuous. Moreover, in order to show that a continuous surjection $q: X \to Y$ is an identification map, it suffices to prove that if V is a subset of Y with the property that $q^{-1}(V)$ is open in X then V is open in Y.

Lemma 1.22 *Let X be a topological space, let Y be a set, and let $q: X \to Y$ be a surjection. Then there is a unique topology on Y for which the function $q: X \to Y$ is an identification map.*

Proof Let τ be the collection consisting of all subsets U of Y for which $q^{-1}(U)$ is open in X. Now $q^{-1}(\emptyset) = \emptyset$, and $q^{-1}(Y) = X$, so that $\emptyset \in \tau$ and $Y \in \tau$. If $\{V_\alpha : \alpha \in A\}$ is any collection of subsets of Y indexed by a set A, then it is a straightforward exercise to verify that

$$\bigcup_{\alpha \in A} q^{-1}(V_\alpha) = q^{-1}\left(\bigcup_{\alpha \in A} V_\alpha\right), \qquad \bigcap_{\alpha \in A} q^{-1}(V_\alpha) = q^{-1}\left(\bigcap_{\alpha \in A} V_\alpha\right)$$

(i.e., given any collection of subsets of Y, the union of the preimages of the sets is the preimage of the union of those sets, and the intersection of the preimages of the sets is the preimage of the intersection of those sets). It follows easily from this that unions and finite intersections of sets belonging to τ must themselves belong to τ. Thus τ is a topology on Y, and the function $q: X \to Y$ is an identification map with respect to the topology τ. Clearly τ is the unique topology on Y for which the function $q: X \to Y$ is an identification map. ∎

Let X be a topological space, let Y be a set, and let $q: X \to Y$ be a surjection. The unique topology on Y for which the function q is an identification map is referred to as the *quotient topology* (or *identification topology*) on Y.

Lemma 1.23 *Let X and Y be topological spaces and let $q: X \to Y$ be an identification map. Let Z be a topological space, and let $f: Y \to Z$ be a function from Y to Z. Then the function f is continuous if and only if the composition function $f \circ q: X \to Z$ is continuous.*

Proof Suppose that f is continuous. Then the composition function $f \circ q$ is a composition of continuous functions and hence is itself continuous.

Conversely suppose that $f \circ q$ is continuous. Let U be an open set in Z. Then $q^{-1}(f^{-1}(U))$ is open in X (since $f \circ q$ is continuous), and hence $f^{-1}(U)$ is open in Y (since the function q is an identification map). Therefore the function f is continuous, as required. ∎

Example Let S^1 be the unit circle in $\mathbb{R}^2$, and let $q: [0,1] \to S^1$ be the map that sends $t \in [0,1]$ to $(\cos 2\pi t, \sin 2\pi t)$. Then $q: [0,1] \to S^1$ is an identification map, and therefore a function $f: S^1 \to Z$ from S^1 to some topological space Z is continuous if and only if $f \circ q: [0,1] \to Z$ is continuous.

Example Let S^n be the n-sphere, consisting of all points $\mathbf{x}$ in $\mathbb{R}^{n+1}$ satisfying $|\mathbf{x}| = 1$. Let $\mathbb{R}P^n$ be the set of all lines in $\mathbb{R}^{n+1}$ passing through the origin (i.e., $\mathbb{R}P^n$ is the set of all one-dimensional vector subspaces of $\mathbb{R}^{n+1}$). Let $q: S^n \to \mathbb{R}P^n$ denote the function which sends a point $\mathbf{x}$ of S^n to the element of $\mathbb{R}P^n$ represented by the line in $\mathbb{R}^{n+1}$ that passes through both $\mathbf{x}$ and the origin. Note that each element of $\mathbb{R}P^n$ is the image (under q) of exactly two antipodal points $\mathbf{x}$ and $-\mathbf{x}$ of S^n. The function q induces a corresponding quotient topology on $\mathbb{R}P^n$ such that $q: S^n \to \mathbb{R}P^n$ is an identification map. The set $\mathbb{R}P^n$, with this topology, is referred to as *real projective n-dimensional space*. In particular $\mathbb{R}P^2$ is referred to as the *real projective plane*. It follows from Lemma 1.23 that a function $f: \mathbb{R}P^n \to Z$ from $\mathbb{R}P^n$ to any topological space Z is continuous if and only if the composition function $f \circ q: S^n \to Z$ is continuous.

1.18 Compact Topological Spaces

Let X be a topological space, and let A be a subset of X. A collection of subsets of X in X is said to *cover* A if and only if every point of A belongs to at least one of these subsets. In particular, an *open cover* of X is collection of open sets in X that covers X.

If $\mathcal{U}$ and $\mathcal{V}$ are open covers of some topological space X then $\mathcal{V}$ is said to be a *subcover* of $\mathcal{U}$ if and only if every open set belonging to $\mathcal{V}$ also belongs to $\mathcal{U}$.

Definition A topological space X is said to be *compact* if and only if every open cover of X possesses a finite subcover.

Lemma 1.24 *Let X be a topological space. A subset A of X is compact (with respect to the subspace topology on A) if and only if, given any collection $\mathcal{U}$ of open sets in X covering A, there exists a finite collection $V_1, V_2, \ldots, V_r$ of open sets belonging to $\mathcal{U}$ such that $A \subset V_1 \cup V_2 \cup \cdots \cup V_r$.*

Proof A subset B of A is open in A (with respect to the subspace topology on A) if and only if $B = A \cap V$ for some open set V in X. The desired result therefore follows directly from the definition of compactness. ∎

We now show that any closed bounded interval in the real line is compact. This result is known as the *Heine-Borel Theorem*. The proof of this theorem uses the *least upper bound principle* which states that, given any non-empty set S of real numbers which is bounded above, there exists a *least upper bound* (or *supremum*) $\sup S$ for the set S.

Theorem 1.25 (Heine-Borel) *Let a and b be real numbers satisfying $a < b$. Then the closed bounded interval $[a, b]$ is a compact subset of $\mathbb{R}$.*

Proof Let $\mathcal{U}$ be a collection of open sets in $\mathbb{R}$ with the property that each point of the interval $[a, b]$ belongs to at least one of these open sets. We must show that $[a, b]$ is covered by finitely many of these open sets.

Let S be the set of all $\tau \in [a, b]$ with the property that $[a, \tau]$ is covered by some finite collection of open sets belonging to $\mathcal{U}$, and let $s = \sup S$. Now $s \in W$ for some open set W belonging to $\mathcal{U}$. Moreover W is open in $\mathbb{R}$, and therefore there exists some $\delta > 0$ such that $(s - \delta, s + \delta) \subset W$. Moreover $s - \delta$ is not an upper bound for the set S, hence there exists some $\tau \in S$ satisfying $\tau > s - \delta$. It follows from the definition of S that $[a, \tau]$ is covered by some finite collection $V_1, V_2, \ldots, V_r$ of open sets belonging to $\mathcal{U}$.

Let $t \in [a, b]$ satisfy $\tau \leq t < s + \delta$. Then

$$[a, t] \subset [a, \tau] \cup (s - \delta, s + \delta) \subset V_1 \cup V_2 \cup \cdots \cup V_r \cup W,$$

and thus $t \in S$. In particular $s \in S$, and moreover $s = b$, since otherwise s would not be an upper bound of the set S. Thus $b \in S$, and therefore $[a, b]$ is covered by a finite collection of open sets belonging to $\mathcal{U}$, as required. ∎

Lemma 1.26 *Let A be a closed subset of some compact topological space X. Then A is compact.*

Proof Let $\mathcal{U}$ be any collection of open sets in X covering A. On adjoining the open set $X \setminus A$ to $\mathcal{U}$, we obtain an open cover of X. This open cover of X possesses a finite subcover, since X is compact. Moreover A is covered by the open sets in the collection $\mathcal{U}$ that belong to this finite subcover. It follows from Lemma 1.24 that A is compact, as required. ∎

Lemma 1.27 *Let $f: X \to Y$ be a continuous function between topological spaces X and Y, and let A be a compact subset of X. Then $f(A)$ is a compact subset of Y.*

Proof Let $\mathcal{V}$ be a collection of open sets in Y which covers $f(A)$. Then A is covered by the collection of all open sets of the form $f^{-1}(V)$ for some $V \in \mathcal{V}$. It follows from the compactness of A that there exists a finite collection $V_1, V_2, \ldots, V_k$ of open sets belonging to $\mathcal{V}$ such that

$$A \subset f^{-1}(V_1) \cup f^{-1}(V_2) \cup \cdots \cup f^{-1}(V_k).$$

But then $f(A) \subset V_1 \cup V_2 \cup \cdots \cup V_k$. This shows that $f(A)$ is compact. ∎

Lemma 1.28 *Let $f: X \to \mathbb{R}$ be a continuous real-valued function on a compact topological space X. Then f is bounded above and below on X.*

Proof The range $f(X)$ of the function f is covered by some finite collection $I_1, I_2, \ldots, I_k$ of open intervals of the form $(-m, m)$, where $m \in \mathbb{N}$, since $f(X)$ is compact (Lemma 1.27) and $\mathbb{R}$ is covered by the collection of all intervals of this form. It follows that $f(X) \subset (-M, M)$, where $(-M, M)$ is the largest of the intervals $I_1, I_2, \ldots, I_k$. Thus the function f is bounded above and below on X, as required. ∎

Proposition 1.29 *Let $f: X \to \mathbb{R}$ be a continuous real-valued function on a compact topological space X. Then there exist points u and v of X such that $f(u) \leq f(x) \leq f(v)$ for all $x \in X$.*

Proof Let $m = \inf\{f(x) : x \in X\}$ and $M = \sup\{f(x) : x \in X\}$. There must exist $v \in X$ satisfying $f(v) = M$, for if $f(x) < M$ for all $x \in X$ then the function $x \mapsto 1/(M - f(x))$ would be a continuous real-valued function on X that was not bounded above, contradicting Lemma 1.28. Similarly there must exist $u \in X$ satisfying $f(u) = m$, since otherwise the function $x \mapsto 1/(f(x) - m)$ would be a continuous function on X that was not bounded above, again contradicting Lemma 1.28. But then $f(u) \leq f(x) \leq f(v)$ for all $x \in X$, as required. ∎

Proposition 1.30 *Let A be a compact subset of a metric space X. Then A is closed in X.*

Proof Let p be a point of X that does not belong to A, and let $f(x) = d(x,p)$, where d is the distance function on X. It follows from Proposition 1.29 that there is a point q of A such that $f(a) \geq f(q)$ for all $a \in A$, since A is compact. Now $f(q) > 0$, since $q \neq p$. Let δ satisfy $0 < \delta \leq f(q)$. Then the open ball of radius δ about the point p is contained in the complement of A, since $f(x) < f(q)$ for all points x of this open ball. It follows that the complement of A is an open set in X, and thus A itself is closed in X. ∎

Proposition 1.31 *Let X be a Hausdorff topological space, and let K be a compact subset of X. Let x be a point of $X \setminus K$. Then there exist open sets V and W in X such that $x \in V$, $K \subset W$ and $V \cap W = \emptyset$.*

Proof For each point $y \in K$ there exist open sets $V_{x,y}$ and $W_{x,y}$ such that $x \in V_{x,y}$, $y \in W_{x,y}$ and $V_{x,y} \cap W_{x,y} = \emptyset$ (since X is a Hausdorff space). But then there exists a finite set $\{y_1, y_2, \ldots, y_r\}$ of points of K such that K is contained in $W_{x,y_1} \cup W_{x,y_2} \cup \cdots \cup W_{x,y_r}$, since K is compact. Define

$$V = V_{x,y_1} \cap V_{x,y_2} \cap \cdots \cap V_{x,y_r}, \qquad W = W_{x,y_1} \cup W_{x,y_2} \cup \cdots \cup W_{x,y_r}.$$

Then V and W are open sets, $x \in V$, $K \subset W$ and $V \cap W = \emptyset$, as required. ∎

Corollary 1.32 *A compact subset of a Hausdorff topological space is closed.*

Proof Let K be a compact subset of a Hausdorff topological space X. It follows immediately from Proposition 1.31 that, for each $x \in X \setminus K$, there exists an open set V_x such that $x \in V_x$ and $V_x \cap K = \emptyset$. But then $X \setminus K$ is equal to the union of the open sets V_x as x ranges over all points of $X \setminus K$, and any set that is a union of open sets is itself an open set. We conclude that $X \setminus K$ is open, and thus K is closed. ∎

Proposition 1.33 *Let X be a Hausdorff topological space, and let K_1 and K_2 be compact subsets of X, where $K_1 \cap K_2 = \emptyset$. Then there exist open sets U_1 and U_2 such that $K_1 \subset U_1$, $K_2 \subset U_2$ and $U_1 \cap U_2 = \emptyset$.*

Proof It follows from Proposition 1.31 that, for each point x of K_1, there exist open sets V_x and W_x such that $x \in V_x$, $K_2 \subset W_x$ and $V_x \cap W_x = \emptyset$. But then there exists a finite set $\{x_1, x_2, \ldots, x_r\}$ of points of K_1 such that

$$K_1 \subset V_{x_1} \cup V_{x_2} \cup \cdots \cup V_{x_r},$$

since K_1 is compact. Define

$$U_1 = V_{x_1} \cup V_{x_2} \cup \cdots \cup V_{x_r}, \qquad U_2 = W_{x_1} \cap W_{x_2} \cap \cdots \cap W_{x_r}.$$

Then U_1 and U_2 are open sets, $K_1 \subset U_1$, $K_2 \subset U_2$ and $U_1 \cap U_2 = \emptyset$, as required. ∎

Lemma 1.34 *Let $f: X \to Y$ be a continuous function from a compact topological space X to a Hausdorff space Y. Then $f(K)$ is closed in Y for every closed set K in X.*

Proof If K is a closed set in X, then K is compact (Lemma 1.26), and therefore $f(K)$ is compact (Lemma 1.27). But any compact subset of a Hausdorff space is closed (Corollary 1.32). Thus $f(K)$ is closed in Y, as required. ∎

Remark If the Hausdorff space Y in Lemma 1.34 is a metric space, then Proposition 1.30 may be used in place of Corollary 1.32 in the proof of the lemma.

Theorem 1.35 *A continuous bijection $f: X \to Y$ from a compact topological space X to a Hausdorff space Y is a homeomorphism.*

Proof Let $g: Y \to X$ be the inverse of the bijection $f: X \to Y$. If U is open in X then $X \setminus U$ is closed in X, and hence $f(X \setminus U)$ is closed in Y, by Lemma 1.34. But $f(X \setminus U) = g^{-1}(X \setminus U) = Y \setminus g^{-1}(U)$. It follows that $g^{-1}(U)$ is open in Y for every open set U in X. Therefore $g: Y \to X$ is continuous, and thus $f: X \to Y$ is a homeomorphism. ∎

We recall that a function $f: X \to Y$ from a topological space X to a topological space Y is said to be an *identification map* if it is surjective and satisfies the following condition: a subset U of Y is open in Y if and only if $f^{-1}(U)$ is open in X.

Proposition 1.36 *A continuous surjection $f: X \to Y$ from a compact topological space X to a Hausdorff space Y is an identification map.*

Proof Let U be a subset of Y. We claim that $Y \setminus U = f(K)$, where $K = X \setminus f^{-1}(U)$. Clearly $f(K) \subset Y \setminus U$. Also, given any $y \in Y \setminus U$, there exists $x \in X$ satisfying $y = f(x)$, since $f: X \to Y$ is surjective. Moreover $x \in K$, since $f(x) \notin U$. Thus $Y \setminus U \subset f(K)$, and hence $Y \setminus U = f(K)$, as claimed.

We must show that the set U is open in Y if and only if $f^{-1}(U)$ is open in X. First suppose that $f^{-1}(U)$ is open in X. Then K is closed in X, and hence $f(K)$ is closed in Y, by Lemma 1.34. It follows that U is open in Y. Conversely if U is open in Y then $f^{-1}(U)$ is open in X, since $f: X \to Y$ is continuous. Thus the surjection $f: X \to Y$ is an identification map. ∎

Example Let S^1 be the unit circle in $\mathbb{R}^2$, defined by $S^1 = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, and let $q\colon [0,1] \to S^1$ be defined by $q(t) = (\cos 2\pi t, \sin 2\pi t)$ for all $t \in [0,1]$. It has been shown that the map q is an identification map. This also follows directly from the fact that $q\colon [0,1] \to S^1$ is a continuous surjection from the compact space $[0,1]$ to the Hausdorff space S^1.

We shall show that a finite Cartesian product of compact spaces is compact. To prove this, we apply the following result, known as the *Tube Lemma*.

Lemma 1.37 *Let X and Y be topological spaces, let K be a compact subset of Y, and U be an open set in $X \times Y$. Let $V = \{x \in X : \{x\} \times K \subset U\}$. Then V is an open set in X.*

Proof Let $x \in V$. For each $y \in K$ there exist open subsets D_y and E_y of X and Y respectively such that $(x,y) \in D_y \times E_y$ and $D_y \times E_y \subset U$. Now there exists a finite set $\{y_1, y_2, \ldots, y_k\}$ of points of K such that $K \subset E_{y_1} \cup E_{y_2} \cup \cdots \cup E_{y_k}$, since K is compact. Set $N_x = D_{y_1} \cap D_{y_2} \cap \cdots \cap D_{y_k}$. Then N_x is an open set in X. Moreover

$$N_x \times K \subset \bigcup_{i=1}^{k} (N_x \times E_{y_i}) \subset \bigcup_{i=1}^{k} (D_{y_i} \times E_{y_i}) \subset U,$$

so that $N_x \subset V$. It follows that V is the union of the open sets N_x for all $x \in V$. Thus V is itself an open set in X, as required. ∎

Theorem 1.38 *A Cartesian product of a finite number of compact spaces is itself compact.*

Proof It suffices to prove that the product of two compact topological spaces X and Y is compact, since the general result then follows easily by induction on the number of compact spaces in the product.

Let $\mathcal{U}$ be an open cover of $X \times Y$. We must show that this open cover possesses a finite subcover.

Let x be a point of X. The set $\{x\} \times Y$ is a compact subset of $X \times Y$, since it is the image of the compact space Y under the continuous map from Y to $X \times Y$ which sends $y \in Y$ to (x,y), and the image of any compact set under a continuous map is itself compact (Lemma 1.27). Therefore there exists a finite collection $U_1, U_2, \ldots, U_r$ of open sets belonging to the open cover $\mathcal{U}$ such that $\{x\} \times Y$ is contained in $U_1 \cup U_2 \cup \cdots \cup U_r$. Let V_x denote the set of all points x' of X for which $\{x'\} \times Y$ is contained in $U_1 \cup U_2 \cup \cdots \cup U_r$. Then $x \in V_x$, and Lemma 1.37 ensures That V_x is an open set in X. Note that

$V_x \times Y$ is covered by finitely many of the open sets belonging to the open cover $\mathcal{U}$.

Now $\{V_x : x \in X\}$ is an open cover of the space X. It follows from the compactness of X that there exists a finite set $\{x_1, x_2, \ldots, x_r\}$ of points of X such that $X = V_{x_1} \cup V_{x_2} \cup \cdots \cup V_{x_r}$. Now $X \times Y$ is the union of the sets $V_{x_j} \times Y$ for $j = 1, 2, \ldots, r$, and each of these sets can be covered by a finite collection of open sets belonging to the open cover $\mathcal{U}$. On combining these finite collections, we obtain a finite collection of open sets belonging to $\mathcal{U}$ which covers $X \times Y$. This shows that $X \times Y$ is compact. ∎

Theorem 1.39 *Let K be a subset of $\mathbb{R}^n$. Then K is compact if and only if K is both closed and bounded.*

Proof Suppose that K is compact. Then K is closed, since $\mathbb{R}^n$ is Hausdorff, and a compact subset of a Hausdorff space is closed (by Corollary 1.32). For each natural number m, let B_m be the open ball of radius m about the origin, given by $B_m = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| < m\}$. Then $\{B_m : m \in \mathbb{N}\}$ is an open cover of $\mathbb{R}^n$. It follows from the compactness of K that there exist natural numbers $m_1, m_2, \ldots, m_k$ such that $K \subset B_{m_1} \cup B_{m_2} \cup \cdots \cup B_{m_k}$. But then $K \subset B_M$, where M is the maximum of $m_1, m_2, \ldots, m_k$, and thus K is bounded.

Conversely suppose that K is both closed and bounded. Then there exists some real number L such that K is contained within the closed cube C given by

$$C = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : -L \leq x_j \leq L \text{ for } j = 1, 2, \ldots, n\}.$$

Now the closed interval $[-L, L]$ is compact by the Heine-Borel Theorem (Theorem 1.25), and C is the Cartesian product of n copies of the compact set $[-L, L]$. It follows from Theorem 1.38 that C is compact. But K is a closed subset of C, and a closed subset of a compact topological space is itself compact, by Lemma 1.26. Thus K is compact, as required. ∎

1.19 The Lebesgue Lemma and Uniform Continuity

Definition Let X be a metric space with distance function d. A subset A of X is said to be *bounded* if there exists a non-negative real number K such that $d(x, y) \leq K$ for all $x, y \in A$. The smallest real number K with this property is referred to as the *diameter* of A, and is denoted by $\operatorname{diam} A$. (Note that $\operatorname{diam} A$ is the supremum of the values of $d(x, y)$ as x and y range over all points of A.)

Lemma 1.40 (Lebesgue Lemma) *Let (X, d) be a compact metric space. Let $\mathcal{U}$ be an open cover of X. Then there exists a positive real number δ such that* *every subset of X whose diameter is less than δ is contained wholly within one of the open sets belonging to the open cover $\mathcal{U}$.*

Proof Every point of X is contained in at least one of the open sets belonging to the open cover $\mathcal{U}$. It follows from this that, for each point x of X, there exists some $\delta_x > 0$ such that the open ball $B(x, 2\delta_x)$ of radius $2\delta_x$ about the point x is contained wholly within one of the open sets belonging to the open cover $\mathcal{U}$. But then the collection consisting of the open balls $B(x, \delta_x)$ of radius δ_x about the points x of X forms an open cover of the compact space X. Therefore there exists a finite set $x_1, x_2, \ldots, x_r$ of points of X such that

$$B(x_1, \delta_1) \cup B(x_2, \delta_2) \cup \cdots \cup B(x_r, \delta_r) = X,$$

where $\delta_i = \delta_{x_i}$ for $i = 1, 2, \ldots, r$. Let $\delta > 0$ be given by

$$\delta = \text{minimum}(\delta_1, \delta_2, \ldots, \delta_r).$$

Suppose that A is a subset of X whose diameter is less than δ. Let u be a point of A. Then u belongs to $B(x_i, \delta_i)$ for some integer i between 1 and r. But then it follows that $A \subset B(x_i, 2\delta_i)$, since, for each point v of A,

$$d(v, x_i) \leq d(v, u) + d(u, x_i) < \delta + \delta_i \leq 2\delta_i.$$

But $B(x_i, 2\delta_i)$ is contained wholly within one of the open sets belonging to the open cover $\mathcal{U}$. Thus A is contained wholly within one of the open sets belonging to $\mathcal{U}$, as required. ∎

Let $\mathcal{U}$ be an open cover of a compact metric space X. A *Lebesgue number* for the open cover $\mathcal{U}$ is a positive real number δ such that every subset of X whose diameter is less than δ is contained wholly within one of the open sets belonging to the open cover $\mathcal{U}$. The Lebesgue Lemma thus states that there exists a Lebesgue number for every open cover of a compact metric space.

Let X and Y be metric spaces with distance functions d_X and d_Y respectively, and let $f: X \to Y$ be a function from X to Y. The function f is said to be *uniformly continuous* on X if and only if, given $\varepsilon > 0$, there exists some $\delta > 0$ such that $d_Y(f(x), f(x')) < \varepsilon$ for all points x and x' of X satisfying $d_X(x, x') < \delta$. (The value of δ should be independent of both x and x'.)

Theorem 1.41 *Let X and Y be metric spaces. Suppose that X is compact. Then every continuous function from X to Y is uniformly continuous.*

Proof Let d_X and d_Y denote the distance functions for the metric spaces X and Y respectively. Let $f: X \to Y$ be a continuous function from X to Y. We must show that f is uniformly continuous.

Let $\varepsilon > 0$ be given. For each $y \in Y$, define

$$V_y = \{x \in X : d_Y(f(x), y) < \tfrac{1}{2}\varepsilon\}.$$

Note that $V_y = f^{-1}\left(B_Y(y, \frac{1}{2}\varepsilon)\right)$, where $B_Y(y, \frac{1}{2}\varepsilon)$ denotes the open ball of radius $\frac{1}{2}\varepsilon$ about y in Y. Now the open ball $B_Y(y, \frac{1}{2}\varepsilon)$ is an open set in Y, and f is continuous. Therefore V_y is open in X for all $y \in Y$. Note that $x \in V_{f(x)}$ for all $x \in X$.

Now $\{V_y : y \in Y\}$ is an open cover of the compact metric space X. It follows from the Lebesgue Lemma (Lemma 1.40) that there exists some $\delta > 0$ such that every subset of X whose diameter is less than δ is a subset of some set V_y. Let x and x' be points of X satisfying $d_X(x, x') < \delta$. The diameter of the set $\{x, x'\}$ is $d_X(x, x')$, which is less than δ. Therefore there exists some $y \in Y$ such that $x \in V_y$ and $x' \in V_y$. But then $d_Y(f(x), y) < \frac{1}{2}\varepsilon$ and $d_Y(f(x'), y) < \frac{1}{2}\varepsilon$, and hence

$$d_Y(f(x), f(x')) \leq d_Y(f(x), y) + d_Y(y, f(x')) < \varepsilon.$$

This shows that $f: X \to Y$ is uniformly continuous, as required. ■

Let K be a closed bounded subset of $\mathbb{R}^n$. It follows from Theorem 1.39 and Theorem 1.41 that any continuous function $f: K \to \mathbb{R}^k$ is uniformly continuous.

1.20 Connected Topological Spaces

Definition A topological space X is said to be *connected* if the empty set $\emptyset$ and the whole space X are the only subsets of X that are both open and closed.

Lemma 1.42 *A topological space X is connected if and only if it has the following property: if U and V are non-empty open sets in X such that $X = U \cup V$, then $U \cap V$ is non-empty.*

Proof If U is a subset of X that is both open and closed, and if $V = X \setminus U$, then U and V are both open, $U \cup V = X$ and $U \cap V = \emptyset$. Conversely if U and V are open subsets of X satisfying $U \cup V = X$ and $U \cap V = \emptyset$, then $U = X \setminus V$, and hence U is both open and closed. Thus a topological space X is connected if and only if there do not exist non-empty open sets U and V such that $U \cup V = X$ and $U \cap V = \emptyset$. The result follows. ■

Let $\mathbb{Z}$ be the set of integers with the usual topology (i.e., the subspace topology on $\mathbb{Z}$ induced by the usual topology on $\mathbb{R}$). Then $\{n\}$ is open for all $n \in \mathbb{Z}$, since

$$\{n\} = \mathbb{Z} \cap \{t \in \mathbb{R} : |t - n| < \tfrac{1}{2}\}.$$

It follows that every subset of $\mathbb{Z}$ is open (since it is a union of sets consisting of a single element, and any union of open sets is open). It follows that a function $f: X \to \mathbb{Z}$ on a topological space X is continuous if and only if $f^{-1}(V)$ is open in X for any subset V of $\mathbb{Z}$. We use this fact in the proof of the next theorem.

Proposition 1.43 *A topological space X is connected if and only if every continuous function $f: X \to \mathbb{Z}$ from X to the set $\mathbb{Z}$ of integers is constant.*

Proof Suppose that X is connected. Let $f: X \to \mathbb{Z}$ be a continuous function. Choose $n \in f(X)$, and let

$$U = \{x \in X : f(x) = n\}, \qquad V = \{x \in X : f(x) \neq n\}.$$

Then U and V are the preimages of the open subsets $\{n\}$ and $\mathbb{Z} \setminus \{n\}$ of $\mathbb{Z}$, and therefore both U and V are open in X. Moreover $U \cap V = \emptyset$, and $X = U \cup V$. It follows that $V = X \setminus U$, and thus U is both open and closed. Moreover U is non-empty, since $n \in f(X)$. It follows from the connectedness of X that $U = X$, so that $f: X \to \mathbb{Z}$ is constant, with value n.

Conversely suppose that every continuous function $f: X \to \mathbb{Z}$ is constant. Let S be a subset of X which is both open and closed. Let $f: X \to \mathbb{Z}$ be defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in S; \\ 0 & \text{if } x \notin S. \end{cases}$$

Now the preimage of any subset of $\mathbb{Z}$ under f is one of the open sets $\emptyset$, S, $X \setminus S$ and X. Therefore the function f is continuous. But then the function f is constant, so that either $S = \emptyset$ or $S = X$. This shows that X is connected. ∎

Lemma 1.44 *The closed interval $[a, b]$ is connected, for all real numbers a and b satisfying $a \leq b$.*

Proof Let $f: [a, b] \to \mathbb{Z}$ be a continuous integer-valued function on $[a, b]$. We show that f is constant on $[a, b]$. Indeed suppose that f were not constant. Then $f(\tau) \neq f(a)$ for some $\tau \in [a, b]$. But the Intermediate Value Theorem would then ensure that, given any real number c between $f(a)$ and $f(\tau)$, there would exist some $t \in [a, \tau]$ for which $f(t) = c$, and this is clearly impossible, since f is integer-valued. Thus f must be constant on $[a, b]$. We now deduce from Proposition 1.43 that $[a, b]$ is connected. ∎

Example Let $X = \{(x, y) \in \mathbb{R}^2 : x \neq 0\}$. The topological space X is not connected. Indeed if $f: X \to \mathbb{Z}$ is defined by

$$f(x, y) = \begin{cases} 1 & \text{if } x > 0, \\ -1 & \text{if } x < 0, \end{cases}$$

then f is continuous on X but is not constant.

A concept closely related to that of connectedness is *path-connectedness*. Let x_0 and x_1 be points in a topological space X. A *path* in X from x_0 to x_1 is defined to be a continuous function $\gamma: [0, 1] \to X$ such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$. A topological space X is said to be *path-connected* if and only if, given any two points x_0 and x_1 of X, there exists a path in X from x_0 to x_1.

Proposition 1.45 *Every path-connected topological space is connected.*

Proof Let X be a path-connected topological space, and let $f: X \to \mathbb{Z}$ be a continuous integer-valued function on X. If x_0 and x_1 are any two points of X then there exists a path $\gamma: [0, 1] \to X$ such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$. But then $f \circ \gamma: [0, 1] \to \mathbb{Z}$ is a continuous integer-valued function on $[0, 1]$. But $[0, 1]$ is connected (Lemma 1.44), therefore $f \circ \gamma$ is constant (Proposition 1.43). It follows that $f(x_0) = f(x_1)$. Thus every continuous integer-valued function on X is constant. Therefore X is connected, by Proposition 1.43. ∎

The topological spaces $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{R}^n$ are all path-connected. Indeed, given any two points of one of these spaces, the straight line segment joining these two points is a continuous path from one point to the other. Also the n-sphere S^n is path-connected for all $n > 0$. We conclude that these topological spaces are connected.

Let A be a subset of a topological space X. Using Lemma 1.42 and the definition of the subspace topology, we see that A is connected if and only if the following condition is satisfied:

- if U and V are open sets in X such that $A \cap U$ and $A \cap V$ are non-empty and $A \subset U \cup V$ then $A \cap U \cap V$ is also non-empty.

Lemma 1.46 *Let X be a topological space and let A be a connected subset of X. Then the closure $\overline{A}$ of A is connected.*

Proof It follows from the definition of the closure of A that $\overline{A} \subset F$ for any closed subset F of X for which $A \subset F$. On taking F to be the complement of some open set U, we deduce that $\overline{A} \cap U = \emptyset$ for any open set U for which

$A \cap U = \emptyset$. Thus if U is an open set in X and if $\overline{A} \cap U$ is non-empty then $A \cap U$ must also be non-empty.

Now let U and V be open sets in X such that $\overline{A} \cap U$ and $\overline{A} \cap V$ are non-empty and $\overline{A} \subset U \cup V$. Then $A \cap U$ and $A \cap V$ are non-empty, and $A \subset U \cup V$. But A is connected. Therefore $A \cap U \cap V$ is non-empty, and thus $\overline{A} \cap U \cap V$ is non-empty. This shows that $\overline{A}$ is connected. ∎

Lemma 1.47 *Let $f: X \to Y$ be a continuous function between topological spaces X and Y, and let A be a connected subset of X. Then $f(A)$ is connected.*

Proof Let $g: f(A) \to \mathbb{Z}$ be any continuous integer-valued function on $f(A)$. Then $g \circ f: A \to \mathbb{Z}$ is a continuous integer-valued function on A. It follows from Proposition 1.43 that $g \circ f$ is constant on A. Therefore g is constant on $f(A)$. We deduce from Proposition 1.43 that $f(A)$ is connected. ∎

Lemma 1.48 *The Cartesian product $X \times Y$ of connected topological spaces X and Y is itself connected.*

Proof Let $f: X \times Y \to \mathbb{Z}$ be a continuous integer-valued function from $X \times Y$ to Z. Choose $x_0 \in X$ and $y_0 \in Y$. The function $x \mapsto f(x, y_0)$ is continuous on X, and is thus constant. Therefore $f(x, y_0) = f(x_0, y_0)$ for all $x \in X$. Now fix x. The function $y \mapsto f(x, y)$ is continuous on Y, and is thus constant. Therefore

$$f(x, y) = f(x, y_0) = f(x_0, y_0)$$

for all $x \in X$ and $y \in Y$. We deduce from Proposition 1.43 that $X \times Y$ is connected. ∎

We deduce immediately that a finite Cartesian product of connected topological spaces is connected.

Proposition 1.49 *Let X be a topological space. For each $x \in X$, let S_x be the union of all connected subsets of X that contain x. Then*

(i) *S_x is connected,*

(ii) *S_x is closed,*

(iii) *if $x, y \in X$, then either $S_x = S_y$, or else $S_x \cap S_y = \emptyset$.*

Proof Let $f\colon S_x \to \mathbb{Z}$ be a continuous integer-valued function on S_x, for some $x \in X$. Let y be any point of S_x. Then, by definition of S_x, there exists some connected set A containing both x and y. But then f is constant on A, and thus $f(x) = f(y)$. This shows that the function f is constant on S_x. We deduce that S_x is connected. This proves (i). Moreover the closure $\overline{S_x}$ is connected, by Lemma 1.46. Therefore $\overline{S_x} \subset S_x$. This shows that S_x is closed, proving (ii).

Finally, suppose that x and y are points of X for which $S_x \cap S_y \neq \emptyset$. Let $f\colon S_x \cup S_y \to \mathbb{Z}$ be any continuous integer-valued function on $S_x \cup S_y$. Then f is constant on both S_x and S_y. Moreover the value of f on S_x must agree with that on S_y, since $S_x \cap S_y$ is non-empty. We deduce that f is constant on $S_x \cup S_y$. Thus $S_x \cup S_y$ is a connected set containing both x and y, and thus $S_x \cup S_y \subset S_x$ and $S_x \cup S_y \subset S_y$, by definition of S_x and S_y. We conclude that $S_x = S_y$. This proves (iii). ∎

Given any topological space X, the connected subsets S_x of X defined as in the statement of Proposition 1.49 are referred to as the *connected components* of X. We see from Proposition 1.49, part (iii) that the topological space X is the disjoint union of its connected components.

Example The connected components of $\{(x, y) \in \mathbb{R}^2 : x \neq 0\}$ are

$$\{(x, y) \in \mathbb{R}^2 : x > 0\} \quad \text{and} \quad \{(x, y) \in \mathbb{R}^2 : x < 0\}.$$

Example The connected components of

$$\{t \in \mathbb{R} : |t - n| < \tfrac{1}{2} \text{ for some integer } n\}.$$

are the sets J_n for all $n \in \mathbb{Z}$, where $J_n = (n - \frac{1}{2}, n + \frac{1}{2})$.

2 Covering Maps and the Monodromy Theorem

2.1 Covering Maps

Definition Let X and $\tilde{X}$ be topological spaces and let $p\colon \tilde{X} \to X$ be a continuous map. An open subset U of X is said to be *evenly covered* by the map p if and only if $p^{-1}(U)$ is a disjoint union of open sets of $\tilde{X}$ each of which is mapped homeomorphically onto U by p. The map $p\colon \tilde{X} \to X$ is said to be a *covering map* if $p\colon \tilde{X} \to X$ is surjective and in addition every point of X is contained in some open set that is evenly covered by the map p.

If $p\colon \tilde{X} \to X$ is a covering map, then we say that $\tilde{X}$ is a *covering space* of X.

Example Let S^1 be the unit circle in $\mathbb{R}^2$. Then the map $p\colon \mathbb{R} \to S^1$ defined by

$$p(t) = (\cos 2\pi t, \sin 2\pi t)$$

is a covering map. Indeed let $\mathbf{n}$ be a point of S^1. Consider the open set U in S^1 containing $\mathbf{n}$ defined by $U = S^1 \setminus \{-\mathbf{n}\}$. Now $\mathbf{n} = (\cos 2\pi t_0, \sin 2\pi t_0)$ for some $t_0 \in \mathbb{R}$. Then $p^{-1}(U)$ is the union of the disjoint open sets J_n for all integers n, where

$$J_n = \{t \in \mathbb{R} : t_0 + n - \tfrac{1}{2} < t < t_0 + n + \tfrac{1}{2}\}.$$

Each of the open sets J_n is mapped homeomorphically onto U by the map p. This shows that $p\colon \mathbb{R} \to S^1$ is a covering map.

Example The map $p\colon \mathbb{C} \to \mathbb{C} \setminus \{0\}$ defined by $p(z) = \exp(z)$ is a covering map. Indeed, given any $\theta \in [-\pi, \pi]$ let us define

$$U_\theta = \{z \in \mathbb{C} \setminus \{0\} : \arg(-z) \neq \theta\}.$$

Then $p^{-1}(U_\theta)$ is the disjoint union of the open sets

$$\{z \in \mathbb{C} : |\operatorname{Im} z - \theta - 2\pi n| < \pi\},$$

for all integers n, and p maps each of these open sets homeomorphically onto U_θ. Thus U_θ is evenly covered by the map p.

Example Consider the map $\alpha\colon (-2, 2) \to S^1$, where $\alpha(t) = (\cos 2\pi t, \sin 2\pi t)$ for all $t \in (-2, 2)$. It can easily be shown that there is no open set U containing the point $(1, 0)$ that is evenly covered by the map α. Indeed suppose that there were to exist such an open set U. Then there would exist some δ satisfying $0 < \delta < \frac{1}{2}$ such that $U_\delta \subset U$, where

$$U_\delta = \{(\cos 2\pi t, \sin 2\pi t) : -\delta < t < \delta\}.$$

The open set U_δ would then be evenly covered by the map α. However the connected components of $\alpha^{-1}(U_\delta)$ are $(-2, -2+\delta)$, $(-1-\delta, -1+\delta)$, $(-\delta, \delta)$, $(1-\delta, 1+\delta)$ and $(2-\delta, 2)$, and neither $(-2, -2+\delta)$ nor $(2-\delta, 2)$ is mapped homeomorphically onto U_δ by α.

Lemma 2.1 *Let $p: \tilde{X} \to X$ be a covering map. Then $p(V)$ is open in X for every open set V in $\tilde{X}$. In particular, a covering map $p: \tilde{X} \to X$ is a homeomorphism if and only if it is a bijection.*

Proof Let V be open in $\tilde{X}$, and let $x \in p(V)$. Then $x = p(v)$ for some $v \in V$. Now there exists an open set U containing the point x which is evenly covered by the covering map p. Then $p^{-1}(U)$ is a disjoint union of open sets, each of which is mapped homeomorphically onto U by the covering map p. One of these open sets contains v; let $\tilde{U}$ be this open set, and let $N_x = p(V \cap \tilde{U})$. Now N_x is open in X, since $V \cap \tilde{U}$ is open in $\tilde{U}$ and $p|\tilde{U}$ is a homeomorphism from $\tilde{U}$ to U. Also $x \in N_x$ and $N_x \subset p(V)$. It follows that $p(V)$ is the union of the open sets N_x as x ranges over all points of $p(V)$, and thus $p(V)$ is itself an open set, as required. The result that a bijective covering map is a homeomorphism then follows directly from the fact that a continuous bijection is a homeomorphism if and only if it maps open sets to open sets. ∎

2.2 Path Lifting and the Monodromy Theorem

Let $p: \tilde{X} \to X$ be a covering map over a topological space X. Let Z be a topological space, and let $f: Z \to X$ be a continuous map from Z to X. A continuous map $\tilde{f}: Z \to \tilde{X}$ is said to be a *lift* of the map $f: Z \to X$ if and only if $p \circ \tilde{f} = f$. We shall prove various results concerning the existence and uniqueness of such lifts.

Proposition 2.2 *Let $p: \tilde{X} \to X$ be a covering map, let Z be a connected topological space, and let $g: Z \to \tilde{X}$ and $h: Z \to \tilde{X}$ be continuous maps. Suppose that $p \circ g = p \circ h$ and that $g(z) = h(z)$ for some $z \in Z$. Then $g = h$.*

Proof Let $Z_0 = \{z \in Z : g(z) = h(z)\}$. Note that Z_0 is non-empty, by hypothesis. We show that Z_0 is both open and closed in Z.

Let z be a point of Z. There exists an open set U in X containing the point $p(g(z))$ which is evenly covered by the covering map p. Then $p^{-1}(U)$ is a disjoint union of open sets, each of which is mapped homeomorphically onto U by the covering map p. One of these open sets contains $g(z)$; let this set be denoted by $\tilde{U}$. Also one of these open sets contains $h(z)$; let this open set be denoted by $\tilde{V}$. Let $N_z = g^{-1}(\tilde{U}) \cap h^{-1}(\tilde{V})$. Then N_z is an open set in Z containing z.

Consider the case when $z \in Z_0$. Then $g(z) = h(z)$, and therefore $\tilde{V} = \tilde{U}$. It follows from this that both g and h map the open set N_z into $\tilde{U}$. But $p \circ g = p \circ h$, and $p|\tilde{U}: \tilde{U} \to U$ is a homeomorphism. Therefore $g|N_z = h|N_z$, and thus $N_z \subset Z_0$. We have thus shown that, for each $z \in Z_0$, there exists an open set N_z such that $z \in N_z$ and $N_z \subset Z_0$. We conclude that Z_0 is open.

Next consider the case when $z \in Z \setminus Z_0$. In this case $\tilde{U} \cap \tilde{V} = \emptyset$, since $g(z) \neq h(z)$. But $g(N_z) \subset \tilde{U}$ and $h(N_z) \subset \tilde{V}$. Therefore $g(z') \neq h(z')$ for all $z' \in N_z$, and thus $N_z \subset Z \setminus Z_0$. We have thus shown that, for each $z \in Z \setminus Z_0$, there exists an open set N_z such that $z \in N_z$ and $N_z \subset Z \setminus Z_0$. We conclude that $Z \setminus Z_0$ is open.

The subset Z_0 of Z is therefore both open and closed. Also Z_0 is non-empty by hypothesis. We deduce that $Z_0 = Z$, since Z is connected. Thus $g = h$, as required. ∎

Lemma 2.3 *Let $p: \tilde{X} \to X$ be a covering map, let Z be a topological space, let A be a connected subset of Z, and let $f: Z \to X$ and $g: A \to \tilde{X}$ be continuous maps with the property that $p \circ g = f|A$. Suppose that $f(Z) \subset U$, where U is an open subset of X that is evenly covered by the covering map p. Then there exists a continuous map $\tilde{f}: Z \to \tilde{X}$ such that $\tilde{f}|A = g$ and $p \circ \tilde{f} = f$.*

Proof The open set U is evenly covered by the covering map p, and therefore $p^{-1}(U)$ is a disjoint union of open sets, each of which is mapped homeomorphically onto U by the covering map p. One of these open sets contains $g(a)$ for some $a \in A$; let this set be denoted by $\tilde{U}$. Let $\sigma: U \to \tilde{U}$ be the inverse of the homeomorphism $p|\tilde{U}: \tilde{U} \to U$, and let $\tilde{f} = \sigma \circ f$. Then $p \circ \tilde{f} = f$. Also $p \circ \tilde{f}|A = p \circ g$ and $\tilde{f}(a) = g(a)$. It follows from Proposition 2.2 that $\tilde{f}|A = g$, since A is connected. Thus $\tilde{f}: Z \to \tilde{X}$ is the required map. ∎

Theorem 2.4 (Path Lifting Theorem) *Let $p: \tilde{X} \to X$ be a covering map, let $\gamma: [0,1] \to X$ be a continuous path in X, and let w be a point of $\tilde{X}$ satisfying $p(w) = \gamma(0)$. Then there exists a unique continuous path $\tilde{\gamma}: [0,1] \to \tilde{X}$ such that $\tilde{\gamma}(0) = w$ and $p \circ \tilde{\gamma} = \gamma$.*

Proof The map $p: \tilde{X} \to X$ is a covering map; therefore there exists an open cover $\mathcal{U}$ of X such that each open set U belonging to X is evenly covered by the map p. Now the collection consisting of the preimages $\gamma^{-1}(U)$ of the open sets U belonging to $\mathcal{U}$ is an open cover of the interval $[0,1]$. But $[0,1]$ is compact, by the Heine-Borel Theorem. It follows from the Lebesgue Lemma that there exists some $\delta > 0$ such that every subinterval of length less than δ is mapped by γ into one of the open sets belonging to $\mathcal{U}$. Partition the interval $[0,1]$ into subintervals $[t_{i-1}, t_i]$, where $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$, and where the length of each subinterval is less than δ. Then each subinterval $[t_{i-1}, t_i]$ is mapped by γ into some open set in X that is evenly covered by the map p. It follows from Lemma 2.3 that once $\tilde{\gamma}(t_{i-1})$ has been determined, we can extend $\tilde{\gamma}$ continuously over the ith subinterval $[t_{i-1}, t_i]$. Thus by extending $\tilde{\gamma}$ successively over $[t_0, t_1]$, $[t_1, t_2]$,..., $[t_{n-1}, t_n]$, we can lift the path $\gamma: [0,1] \to X$ to a path $\tilde{\gamma}: [0,1] \to \tilde{X}$ starting at w. The uniqueness of $\tilde{\gamma}$ follows from Proposition 2.2. ∎

Theorem 2.5 (The Monodromy Theorem) *Let $p: \tilde{X} \to X$ be a covering map, let $H: [0,1] \times [0,1] \to X$ be a continuous map, and let w be a point of $\tilde{X}$ satisfying $p(w) = H(0,0)$. Then there exists a unique continuous map $\tilde{H}: [0,1] \times [0,1] \to \tilde{X}$ such that $\tilde{H}(0,0) = w$ and $p \circ \tilde{H} = H$.*

Proof The unit square $[0,1] \times [0,1]$ is compact. By applying the Lebesgue Lemma to an open cover of the square by preimages of evenly covered open sets in X (as in the proof of Theorem 2.4), we see that there exists some $\delta > 0$ with the property that any square contained in $[0,1] \times [0,1]$ whose sides have length less than δ is mapped by H into some open set in X which is evenly covered by the covering map p. It follows from Lemma 2.3 that if the lift $\tilde{H}$ of H has already been determined over a corner, or along one side, or along two adjacent sides of a square whose sides have length less than δ, then $\tilde{H}$ can be extended over the whole of that square. Thus if we subdivide $[0,1] \times [0,1]$ into squares $S_{j,k}$, where

$$S_{j,k} = \left\{ (s,t) \in [0,1] \times [0,1] : \frac{j-1}{n} \leq s \leq \frac{j}{n} \quad \text{and} \quad \frac{k-1}{n} \leq t \leq \frac{k}{n} \right\},$$

and $1/n < \delta$, then we can construct a lift $\tilde{H}$ of H by defining $\tilde{H}(0,0) = w$, and then successively extending $\tilde{H}$ in turn over each of these smaller squares. (Indeed the map $\tilde{H}$ can be extended successively over the squares

$$S_{1,1}, S_{1,2}, \ldots, S_{1,n}, S_{2,1}, S_{2,2}, \ldots, S_{2,n}, S_{3,1}, \ldots, S_{n-1,n}, \ldots, S_{n,1}, S_{n,2}, \ldots, S_{n,n}.)$$

The uniqueness of $\tilde{H}$ follows from Proposition 2.2. ∎

3 Homotopies and the Fundamental Group

3.1 Homotopies

Definition Let $f: X \to Y$ and $g: X \to Y$ be continuous maps between topological spaces X and Y. The maps f and g are said to be *homotopic* if there exists a continuous map $H: X \times [0,1] \to Y$ such that $H(x,0) = f(x)$ and $H(x,1) = g(x)$ for all $x \in X$. If the maps f and g are homotopic then we denote this fact by writing $f \simeq g$. The map H with the properties stated above is referred to as a *homotopy* between f and g.

Continuous maps f and g from X to Y are homotopic if and only if it is possible to 'continuously deform' the map f into the map g.

Lemma 3.1 *Let X and Y be topological spaces. The homotopy relation $\simeq$ is an equivalence relation on the set of all continuous maps from X to Y.*

Proof Clearly $f \simeq f$, since $(x,t) \mapsto f(x)$ is a homotopy between f and itself. Thus the relation is reflexive. If $f \simeq g$ then there exists a homotopy $H: X \times [0,1] \to Y$ between f and g (so that $H(x,0) = f(x)$ and $H(x,1) = g(x)$ for all $x \in X$). But then $(x,t) \mapsto H(x, 1-t)$ is a homotopy between g and f. Therefore $f \simeq g$ if and only if $g \simeq f$. Thus the relation is symmetric. Finally, suppose that $f \simeq g$ and $g \simeq h$. Then there exist homotopies $H_1: X \times [0,1] \to Y$ and $H_2: X \times [0,1] \to Y$ such that $H_1(x,0) = f(x)$, $H_1(x,1) = g(x) = H_2(x,0)$ and $H_2(x,1) = h(x)$ for all $x \in X$. Define $H: X \times [0,1] \to Y$ by

$$H(x,t) = \begin{cases} H_1(x,2t) & \text{if } 0 \leq t \leq \frac{1}{2}; \\ H_2(x,2t-1) & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

Now $H|X \times [0, \frac{1}{2}]$ and $H|X \times [\frac{1}{2}, 1]$ are continuous. It follows from elementary point set topology that H is continuous on $X \times [0,1]$. Moreover $H(x,0) = f(x)$ and $H(x,1) = h(x)$ for all $x \in X$. Thus $f \simeq h$. Thus the relation is transitive. The relation $\simeq$ is therefore an equivalence relation. ∎

Definition Let X and Y be topological spaces, and let A be a subset of X. Let $f: X \to Y$ and $g: X \to Y$ be continuous maps from X to some topological space Y, where $f|A = g|A$ (i.e., $f(a) = g(a)$ for all $a \in A$). We say that f and g are *homotopic relative to* A (denoted by $f \simeq g \text{ rel } A$) if and only if there exists a (continuous) homotopy $H: X \times [0,1] \to Y$ such that $H(x,0) = f(x)$ and $H(x,1) = g(x)$ for all $x \in X$ and $H(a,t) = f(a) = g(a)$ for all $a \in A$.

Homotopy relative to a chosen subset of X is also an equivalence relation on the set of all continuous maps between topological spaces X and Y.

3.2 The Fundamental Group of a Topological Space

Definition Let X be a topological space, and let x_0 and x_1 be points of X. A *path* in X from x_0 to x_1 is defined to be a continuous map $\gamma: [0,1] \to X$ for which $\gamma(0) = x_0$ and $\gamma(1) = x_1$. A *loop* in X based at x_0 is defined to be a continuous map $\gamma: [0,1] \to X$ for which $\gamma(0) = \gamma(1) = x_0$.

We can concatenate paths. Let $\gamma_1: [0,1] \to X$ and $\gamma_2: [0,1] \to X$ be paths in some topological space X. Suppose that $\gamma_1(1) = \gamma_2(0)$. We define the *product path* $\gamma_1.\gamma_2: [0,1] \to X$ by

$$(\gamma_1.\gamma_2)(t) = \begin{cases} \gamma_1(2t) & \text{if } 0 \leq t \leq \frac{1}{2}; \\ \gamma_2(2t-1) & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

(The continuity of $\gamma_1.\gamma_2$ may be deduced from Lemma 3.1.)

If $\gamma: [0,1] \to X$ is a path in X then we define the *inverse path* $\gamma^{-1}: [0,1] \to X$ by $\gamma^{-1}(t) = \gamma(1-t)$. (Thus if γ is a path from the point x_0 to the point x_1 then γ^{-1} is the path from x_1 to x_0 obtained by traversing γ in the reverse direction.)

Let X be a topological space, and let $x_0 \in X$ be some chosen point of X. We define an equivalence relation on the set of all (continuous) loops based at the basepoint x_0 of X, where two such loops γ_0 and γ_1 are equivalent if and only if $\gamma_0 \simeq \gamma_1 \text{ rel } \{0,1\}$. We denote the equivalence class of a loop $\gamma: [0,1] \to X$ based at x_0 by $[\gamma]$. This equivalence class is referred to as the *based homotopy class* of the loop γ. The set of equivalence classes of loops based at x_0 is denoted by $\pi_1(X, x_0)$. Thus two loops γ_0 and γ_1 represent the same element of $\pi_1(X, x_0)$ if and only if $\gamma_0 \simeq \gamma_1 \text{ rel } \{0,1\}$ (i.e., there exists a homotopy $F: [0,1] \times [0,1] \to X$ between γ_0 and γ_1 which maps $(0,\tau)$ and $(1,\tau)$ to x_0 for all $\tau \in [0,1]$).

Theorem 3.2 *Let X be a topological space, let x_0 be some chosen point of X, and let $\pi_1(X, x_0)$ be the set of all based homotopy classes of loops based at the point x_0. Then $\pi_1(X, x_0)$ is a group, the group multiplication on $\pi_1(X, x_0)$ being defined according to the rule $[\gamma_1][\gamma_2] = [\gamma_1.\gamma_2]$ for all loops γ_1 and γ_2 based at x_0.*

Proof First we show that the group operation on $\pi_1(X, x_0)$ is well-defined. Let γ_1, γ_1', γ_2 and γ_2' be loops in X based at the point x_0. Suppose that $[\gamma_1] = [\gamma_1']$ and $[\gamma_2] = [\gamma_2']$. Let the map $F: [0,1] \times [0,1] \to X$ be defined by

$$F(t,\tau) = \begin{cases} F_1(2t,\tau) & \text{if } 0 \leq t \leq \frac{1}{2}, \\ F_2(2t-1,\tau) & \text{if } \frac{1}{2} \leq t \leq 1, \end{cases}$$

where $F_1\colon [0,1] \times [0,1] \to X$ is a homotopy between γ_1 and γ_1', $F_2\colon [0,1] \times [0,1] \to X$ is a homotopy between γ_2 and γ_2', and where the homotopies F_1 and F_2 map $(0,\tau)$ and $(1,\tau)$ to x_0 for all $\tau \in [0,1]$. Then F is itself a homotopy from $\gamma_1.\gamma_2$ to $\gamma_1'.\gamma_2'$, and maps $(0,\tau)$ and $(1,\tau)$ to x_0 for all $\tau \in [0,1]$. Thus $[\gamma_1.\gamma_2] = [\gamma_1'.\gamma_2']$, showing that the group operation on $\pi_1(X,x_0)$ is well-defined.

Next we show that the group operation on $\pi_1(X,x_0)$ is associative. Let γ_1, γ_2 and γ_3 be loops based at x_0, and let $\alpha = (\gamma_1.\gamma_2).\gamma_3$. Then $\gamma_1.(\gamma_2.\gamma_3) = \alpha\circ\theta$, where

$$\theta(t) = \begin{cases} \frac{1}{2}t & \text{if } 0 \leq t \leq \frac{1}{2}; \\ t - \frac{1}{4} & \text{if } \frac{1}{2} \leq t \leq \frac{3}{4}; \\ 2t - 1 & \text{if } \frac{3}{4} \leq t \leq 1. \end{cases}$$

Thus the map $G\colon [0,1]\times[0,1] \to X$ defined by $G(t,\tau) = \alpha((1-\tau)t + \tau\theta(t))$ is a homotopy between $(\gamma_1.\gamma_2).\gamma_3$ and $\gamma_1.(\gamma_2.\gamma_3)$, and moreover this homotopy maps $(0,\tau)$ and $(1,\tau)$ to x_0 for all $\tau \in [0,1]$. It follows that $(\gamma_1.\gamma_2).\gamma_3 \simeq \gamma_1.(\gamma_2.\gamma_3)$ rel $\{0,1\}$ and hence $([\gamma_1][\gamma_2])[\gamma_3] = [\gamma_1]([\gamma_2][\gamma_3])$. This shows that the group operation on $\pi_1(X,x_0)$ is associative.

Let $\varepsilon\colon [0,1] \to X$ denote the constant loop at x_0, defined by $\varepsilon(t) = x_0$ for all $t \in [0,1]$. Then $\varepsilon.\gamma = \gamma\circ\theta_0$ and $\gamma.\varepsilon = \gamma\circ\theta_1$ for any loop γ based at x_0, where

$$\theta_0(t) = \begin{cases} 0 & \text{if } 0 \leq t \leq \frac{1}{2}, \\ 2t-1 & \text{if } \frac{1}{2} \leq t \leq 1, \end{cases} \qquad \theta_1(t) = \begin{cases} 2t & \text{if } 0 \leq t \leq \frac{1}{2}, \\ 1 & \text{if } \frac{1}{2} \leq t \leq 1, \end{cases}$$

for all $t \in [0,1]$. But the continuous map $(t,\tau) \mapsto \gamma((1-\tau)t + \tau\theta_j(t))$ is a homotopy between γ and $\gamma\circ\theta_j$ for $j = 0,1$ which sends $(0,\tau)$ and $(1,\tau)$ to x_0 for all $\tau \in [0,1]$. Therefore $\varepsilon.\gamma \simeq \gamma \simeq \gamma.\varepsilon$ rel $\{0,1\}$, and hence $[\varepsilon][\gamma] = [\gamma] = [\gamma][\varepsilon]$. We conclude that $[\varepsilon]$ represents the identity element of $\pi_1(X,x_0)$.

It only remains to verify the existence of inverses. Now the map $K\colon [0,1]\times[0,1] \to X$ defined by

$$K(t,\tau) = \begin{cases} \gamma(2\tau t) & \text{if } 0 \leq t \leq \frac{1}{2}; \\ \gamma(2\tau(1-t)) & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

is a homotopy between the loops $\gamma.\gamma^{-1}$ and ε, and moreover this homotopy sends $(0,\tau)$ and $(1,\tau)$ to x_0 for all $\tau \in [0,1]$. Therefore $\gamma.\gamma^{-1} \simeq \varepsilon \operatorname{rel}\{0,1\}$, and thus $[\gamma][\gamma^{-1}] = [\gamma.\gamma^{-1}] = [\varepsilon]$. On replacing γ by γ^{-1}, we see also that $[\gamma^{-1}][\gamma] = [\varepsilon]$, and thus $[\gamma^{-1}] = [\gamma]^{-1}$, as required. ∎

Let x_0 be a point of some topological space X. The group $\pi_1(X,x_0)$ is referred to as the *fundamental group* of X based at the point x_0.

Let $f: X \to Y$ be a continuous map between topological spaces X and Y, and let x_0 be a point of X. Then f induces a homomorphism $f_\#: \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$, where $f_\#([\gamma]) = [f \circ \gamma]$ for all loops $\gamma: [0, 1] \to X$ based at x_0. If x_0, y_0 and z_0 are points belonging to topological spaces X, Y and Z, and if $f: X \to Y$ and $g: Y \to Z$ are continuous maps satisfying $f(x_0) = y_0$ and $g(y_0) = z_0$, then the induced homomorphisms $f_\#: \pi_1(X, x_0) \to \pi_1(Y, y_0)$ and $g_\#: \pi_1(Y, x_0) \to \pi_1(Z, z_0)$ satisfy $g_\# \circ f_\# = (g \circ f)_\#$. It follows easily from this that any homeomorphism of topological spaces induces a corresponding isomorphism of fundamental groups, and thus the fundamental group is a topological invariant.

3.3 Simply-Connected Topological Spaces

Definition A topological space X is said to be *simply-connected* if it is path-connected, and any continuous map $f: \partial D \to X$ mapping the boundary circle ∂D of a closed disc D into X can be extended continuously over the whole of the disk.

Example $\mathbb{R}^n$ is simply-connected for all n. Indeed any continuous map $f: \partial D \to \mathbb{R}^n$ defined over the boundary ∂D of the closed unit disk D can be extended to a continuous map $F: D \to \mathbb{R}^n$ over the whole disk by setting $F(r\mathbf{x}) = rf(\mathbf{x})$ for all $\mathbf{x} \in \partial D$ and $r \in [0, 1]$.

Let E be a topological space that is homeomorphic to the closed disk D, and let $\partial E = h(\partial D)$, where ∂D is the boundary circle of the disk D and $h: D \to E$ is a homeomorphism from D to E. Then any continuous map $g: \partial E \to X$ mapping ∂E into a simply-connected space X extends continuously to the whole of E. Indeed there exists a continuous map $F: D \to X$ which extends $g \circ h: \partial D \to X$, and the map $F \circ h^{-1}: E \to X$ then extends the map g.

Theorem 3.3 *A path-connected topological space X is simply-connected if and only if $\pi_1(X, x)$ is trivial for all $x \in X$.*

Proof Suppose that the space X is simply-connected. Let $\gamma: [0, 1] \to X$ be a loop based at some point x of X. Now the unit square is homeomorphic to the unit disk, and therefore any continuous map defined over the boundary of the square can be continuously extended over the whole of the square. It follows that there exists a continuous map $F: [0, 1] \times [0, 1] \to X$ such that $F(t, 0) = \gamma(t)$ and $F(t, 1) = x$ for all $t \in [0, 1]$, and $F(0, \tau) = F(1, \tau) = x$ for all $\tau \in [0, 1]$. Thus $\gamma \simeq \varepsilon_x \,\mathrm{rel}\{0, 1\}$, where ε_x is the constant loop at x, and hence $[\gamma] = [\varepsilon_x]$ in $\pi_1(X, x)$. This shows that $\pi_1(X, x)$ is trivial.

Conversely suppose that X is path-connected and $\pi_1(X, x)$ is trivial for all $x \in X$. Let $f: \partial D \to X$ be a continuous function defined on the boundary circle ∂D of the closed unit disk D in $\mathbb{R}^2$. We must show that f can be extended continuously over the whole of D. Let $x = f(1, 0)$. There exists a continuous map $G: [0, 1]\times[0, 1] \to X$ such that $G(t, 0) = f(\cos(2\pi t), \sin(2\pi t))$ and $G(t, 1) = x$ for all $t \in [0, 1]$ and $G(0, \tau) = G(1, \tau) = x$ for all $\tau \in [0, 1]$, since $\pi_1(X, x)$ is trivial. Moreover $G(t_1, \tau_1) = G(t_2, \tau_2)$ whenever $q(t_1, \tau_1) = q(t_2, \tau_2)$, where

$$q(t, \tau) = ((1 - \tau) \cos(2\pi t) + \tau, (1 - \tau) \sin(2\pi t))$$

for all $t, \tau \in [0, 1]$. It follows that there is a well-defined function $F: D \to X$ such that $F \circ q = G$. However $q: [0, 1] \times [0, 1] \to D$ is a continuous surjection from a compact space to a Hausdorff space and is therefore an identification map. It follows that $F: D \to X$ is continuous (since a basic property of identification maps ensures that a function $F: D \to X$ is continuous if and only if $F \circ q: [0, 1] \times [0, 1] \to X$ is continuous). Moreover $F: D \to X$ extends the map f. We conclude that the space X is simply-connected, as required. ∎

One can show that, if two points x_1 and x_2 in a topological space X can be joined by a path in X then $\pi_1(X, x_1)$ and $\pi_1(X, x_2)$ are isomorphic. On combining this result with Theorem 3.3, we see that a path-connected topological space X is simply-connected if and only if $\pi_1(X, x)$ is trivial for some $x \in X$.

Theorem 3.4 *Let X be a topological space, and let U and V be open subsets of X, with $U \cup V = X$. Suppose that U and V are simply-connected, and that $U \cap V$ is non-empty and path-connected. Then X is itself simply-connected.*

Proof We must show that any continuous function $f: \partial D \to X$ defined on the unit circle ∂D can be extended continuously over the closed unit disk D. Now the preimages $f^{-1}(U)$ and $f^{-1}(V)$ of U and V are open in ∂D (since f is continuous), and $\partial D = f^{-1}(U) \cup f^{-1}(V)$. It follows from the Lebesgue Lemma that there exists some $\delta > 0$ such that any arc in ∂D whose length is less than δ is entirely contained in one or other of the sets $f^{-1}(U)$ and $f^{-1}(V)$. Choose points $z_1, z_2, \ldots, z_n$ around ∂D such that the distance from z_i to z_{i+1} is less than δ for $i = 1, 2, \ldots, n-1$ and the distance from z_n to z_1 is also less than δ. Then, for each i, the short arc joining z_{i-1} to z_i is mapped by f into one or other of the open sets U and V.

Let x_0 be some point of $U \cap V$. Now the sets U, V and $U \cap V$ are all path-connected. Therefore we can choose paths $\alpha_i: [0, 1] \to X$ for $i = 1, 2, \ldots, n$

such that $\alpha_i(0) = x_0$, $\alpha_i(1) = f(z_i)$, $\alpha_i([0,1]) \subset U$ whenever $f(z_i) \in U$, and $\alpha_i([0,1]) \subset V$ whenever $f(z_i) \in V$. For convenience let $\alpha_0 = \alpha_n$.

Now, for each i, consider the sector T_i of the closed unit disk bounded by the line segments joining the centre of the disk to the points z_{i-1} and z_i and by the short arc joining z_{i-1} to z_i. Now this sector is homeomorphic to the closed unit disk, and therefore any continuous function mapping the boundary ∂T_i of T_i into a simply-connected space can be extended continuously over the whole of T_i. In particular, let F_i be the function on ∂T_i defined by

$$F_i(z) = \begin{cases} f(z) & \text{if } z \in T_i \cap \partial D, \\ \alpha_{i-1}(t) & \text{if } z = tz_{i-1} \text{ for some } t \in [0,1], \\ \alpha_i(t) & \text{if } z = tz_i \text{ for any } t \in [0,1], \end{cases}$$

Note that $F_i(\partial T_i) \subset U$ whenever the short arc joining z_{i-1} to z_i is mapped by f into U, and $F_i(\partial T_i) \subset V$ whenever this short arc is mapped into V. But U and V are both simply-connected. It follows that each of the functions F_i can be extended continuously over the whole of the sector T_i. Moreover the functions defined in this fashion on each of the sectors T_i agree with one another wherever the sectors intersect, and can therefore be pieced together to yield a continuous map defined over the the whole of the closed disk D which extends the map f, as required. ▮

Example The n-dimensional sphere S^n is simply-connected for all $n > 1$, where $S^n = \{\mathbf{x} \in \mathbb{R}^{n+1} : |\mathbf{x}| = 1\}$. Indeed let $U = \{\mathbf{x} \in S^n : x_{n+1} > -\frac{1}{2}\}$ and $V = \{\mathbf{x} \in S^n : x_{n+1} < \frac{1}{2}\}$. Then U and V are homeomorphic to an n-dimensional ball, and are therefore simply-connected. Moreover $U \cap V$ is path-connected, provided that $n > 1$. It follows that S^n is simply-connected for all $n > 1$.

4 Covering Maps and Discontinuous Group Actions

4.1 Covering Maps and Induced Homomorphisms of the Fundamental Group

Let $p: \tilde{X} \to X$ be a covering map and let $\alpha: [0,1] \to X$ and $\beta: [0,1] \to X$ be paths in the base space X which both start at some point x_0 of X and finish at some point x_1 of X, so that

$$\alpha(0) = \beta(0) = x_0 \quad \text{and} \quad \alpha(1) = \beta(1) = x_1.$$

Let $\tilde{x}_0$ be some point of the covering space $\tilde{X}$ that projects down to x_0, so that $p(\tilde{x}_0) = x_0$. It follows from the Path Lifting Theorem (Theorem 2.4) that there exist paths $\tilde{\alpha}: [0,1] \to \tilde{X}$ and $\tilde{\beta}: [0,1] \to \tilde{X}$ in the covering space $\tilde{X}$ that both start at $\tilde{x}_0$ and that are lifts of the paths α and β respectively. Thus

$$\tilde{\alpha}(0) = \tilde{\beta}(0) = \tilde{x}_0,$$
$$p(\tilde{\alpha}(t) = \alpha(t) \quad \text{and} \quad p(\tilde{\beta}(t) = \beta(t) \quad \text{for all } t \in [0,1].$$

These lifts $\tilde{\alpha}$ and $\tilde{\beta}$ of the paths α and β are uniquely determined by their starting point $\tilde{x}_0$ (see Proposition 2.2).

Now, though the lifts $\tilde{\alpha}$ and $\tilde{\beta}$ of the paths α and β have been chosen such that they start at the same point $\tilde{x}_0$ of the covering space $\tilde{X}$, they need not in general end at the same point of $\tilde{X}$. However we shall prove that if $\alpha \simeq \beta$ rel $\{0,1\}$, then the lifts $\tilde{\alpha}$ and $\tilde{\beta}$ of α and β respectively that both start at some point $\tilde{x}_0$ of $\tilde{X}$ will both finish at some point $\tilde{x}_1$ of $\tilde{x}$, so that $\tilde{\alpha}(1) = \tilde{\beta}(1) = \tilde{x}_1$. This result is established in Proposition 4.1 below.

Proposition 4.1 *Let $p: \tilde{X} \to X$ be a covering map over a topological space X, let $\alpha: [0,1] \to X$ and $\beta: [0,1] \to X$ be paths in X, where $\alpha(0) = \beta(0)$ and $\alpha(1) = \beta(1)$, and let $\tilde{\alpha}: [0,1] \to \tilde{X}$ and $\tilde{\beta}: [0,1] \to \tilde{X}$ be paths in $\tilde{X}$ such that $p \circ \tilde{\alpha} = \alpha$ and $p \circ \tilde{\beta} = \beta$. Suppose that $\tilde{\alpha}(0) = \tilde{\beta}(0)$ and that $\alpha \simeq \beta$ rel $\{0,1\}$. Then $\tilde{\alpha}(1) = \tilde{\beta}(1)$ and $\tilde{\alpha} \simeq \tilde{\beta}$ rel $\{0,1\}$.*

Proof Let x_0 and x_1 be the points of X given by

$$x_0 = \alpha(0) = \beta(0), \qquad x_1 = \alpha(1) = \beta(1).$$

Now $\alpha \simeq \beta$ rel $\{0,1\}$, and therefore there exists a homotopy $F: [0,1] \times [0,1] \to X$ such that

$$F(t,0) = \alpha(t) \quad \text{and} \quad F(t,1) = \beta(t) \quad \text{for all } t \in [0,1],$$

$$F(0,\tau) = x_0 \quad \text{and} \quad F(1,\tau) = x_1 \quad \text{for all } \tau \in [0,1].$$

It then follows from the Monodromy Theorem (Theorem 2.5) that there exists a continuous map $G: [0,1] \times [0,1] \to \tilde{X}$ such that $p \circ G = F$ and $G(0,0) = \tilde{\alpha}(0)$. Then $p(G(0,\tau)) = x_0$ and $p(G(1,\tau)) = x_1$ for all $\tau \in [0,1]$. A straightforward application of Proposition 2.2 shows that any continuous lift of a constant path must itself be a constant path. Therefore $G(0,\tau) = \tilde{x}_0$ and $G(1,\tau) = \tilde{x}_1$ for all $\tau \in [0,1]$, where

$$\tilde{x}_0 = G(0,0) = \tilde{\alpha}(0), \qquad \tilde{x}_1 = G(1,0).$$

However

$$G(0,0) = G(0,1) = \tilde{x}_0 = \tilde{\alpha}(0) = \tilde{\beta}(0),$$

$$p(G(t,0)) = F(t,0) = \alpha(t) = p(\tilde{\alpha}(t))$$

and

$$p(G(t,1)) = F(t,1) = \beta(t) = p(\tilde{\beta}(t))$$

for all $t \in [0,1]$. It follows that the map that sends $t \in [0,1]$ to $G(t,0)$ is a lift of the path α that starts at $\tilde{x}_0$, and the map that sends $t \in [0,1]$ to $G(t,1)$ is a lift of the path β that also starts at $\tilde{x}_0$. However Proposition 2.2 ensures that the lifts $\tilde{\alpha}$ and $\tilde{\beta}$ of the paths α and β are uniquely determined by their starting points. It follows that $G(t,0) = \tilde{\alpha}(t)$ and $G(t,1) = \tilde{\beta}(t)$ for all $t \in [0,1]$. In particular,

$$\tilde{\alpha}(1) = G(1,0) = \tilde{x}_1 = G(1,1) = \tilde{\beta}(1).$$

Moreover the map $G: [0,1] \times [0,1] \to \tilde{X}$ is a homotopy between the paths $\tilde{\alpha}$ and $\tilde{\beta}$ which satisfies $G(0,\tau) = \tilde{x}_0$ and $G(1,\tau) = \tilde{x}_1$ for all $\tau \in [0,1]$. It follows that $\tilde{\alpha} \simeq \tilde{\beta}$ rel $\{0,1\}$, as required. ∎

4.2 The Fundamental Group of the Circle

Theorem 4.2 *The fundamental group $\pi_1(S^1, b)$ of the circle S^1 is isomorphic to the group $\mathbb{Z}$ of integers under the operation of addition, where $b \in S^1$.*

Proof We regard S^1 as the unit circle in $\mathbb{R}^2$. Without loss of generality, we can take $b = (1,0)$. Now the map $p: \mathbb{R} \to S^1$ which sends $t \in \mathbb{R}$ to $(\cos 2\pi t, \sin 2\pi t)$ is a covering map, and $b = p(0)$. Moreover $p(t_1) = p(t_2)$ if and only if $t_1 - t_2$ is an integer; in particular $p(t) = b$ if and only if t is an integer.

Let α and β be loops in S^1 based at b, and let $\tilde{\alpha}$ and $\tilde{\beta}$ be paths in the real line $\mathbb{R}$ that start at 0 and are lifts of the loops α and β respectively, so that

$$\tilde{\alpha}(0) = \tilde{\beta}(0) = 0, \quad p \circ \tilde{\alpha} = \alpha \quad \text{and} \quad p \circ \tilde{\beta} = \beta.$$

Suppose that α and β represent the same element of $\pi_1(S^1, b)$. Then $\alpha \simeq \beta$ rel $\{0, 1\}$. It follows from Proposition 4.1 that $\tilde{\alpha}(1) = \tilde{\beta}(1)$. We conclude from this that there exists a well-defined function $\lambda\colon \pi_1(S^1, b) \to \mathbb{Z}$ characterized by the property that $\lambda([\alpha]) = \tilde{\alpha}(1)$ for all loops α based at b, where $\tilde{\alpha}\colon [0, 1] \to \mathbb{R}$ is the unique path in $\mathbb{R}$ satisfying $\tilde{\alpha}(0) = 0$ and $p \circ \tilde{\alpha} = \alpha$.

Next we show that $\lambda\colon \pi_1(S^1, b) \to \mathbb{Z}$ is a homomorphism.

Let α and β be any loops based at b, and let $\tilde{\alpha}$ and $\tilde{\beta}$ be lifts of α and β, where $\tilde{\alpha}(0) = \tilde{\beta}(0) = 0$. The element $[\alpha][\beta]$ of $\pi_1(S^1, b)$ is represented by the product path $\alpha.\beta$, where

$$(\alpha.\beta)(t) = \begin{cases} \alpha(2t) & \text{if } 0 \le t \le \frac{1}{2}; \\ \beta(2t - 1) & \text{if } \frac{1}{2} \le t \le 1. \end{cases}$$

Define a continuous path $\sigma\colon [0, 1] \to \mathbb{R}$ by

$$\sigma(t) = \begin{cases} \tilde{\alpha}(2t) & \text{if } 0 \le t \le \frac{1}{2}; \\ \tilde{\beta}(2t - 1) + \tilde{\alpha}(1) & \text{if } \frac{1}{2} \le t \le 1. \end{cases}$$

(Note that $\sigma(t)$ is well-defined when $t = \frac{1}{2}$, since $\tilde{\beta}(0) = 0$.) Then $p \circ \sigma = \alpha.\beta$ and thus

$$\begin{aligned} \lambda([\alpha][\beta]) &= \lambda([\alpha.\beta]) = \sigma(1) = \tilde{\alpha}(1) + \tilde{\beta}(1) \\ &= \lambda([\alpha]) + \lambda([\beta]). \end{aligned}$$

Thus $\lambda\colon \pi_1(S^1, b) \to \mathbb{Z}$ is a homomorphism.

Now suppose that $\lambda([\alpha]) = \lambda([\beta])$. Let $F\colon [0, 1] \times [0, 1] \to S^1$ be the homotopy between α and β defined by

$$F(t, \tau) = p\left((1 - \tau)\tilde{\alpha}(t) + \tau\tilde{\beta}(t)\right),$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are the lifts of α and β respectively satisfying $\tilde{\alpha}(0) = \tilde{\beta}(0) = 0$. Now $\tilde{\beta}(1) = \lambda([\beta]) = \lambda([\alpha]) = \tilde{\alpha}(1)$, and $\tilde{\beta}(0) = \tilde{\alpha}(0) = 0$. Therefore $F(0, \tau) = b = p(\tilde{\alpha}(1)) = F(1, \tau)$ for all $\tau \in [0, 1]$. Thus $\alpha \simeq \beta$ rel $\{0, 1\}$, and therefore $[\alpha] = [\beta]$. This shows that $\lambda\colon \pi_1(S^1, b) \to \mathbb{Z}$ is injective.

The homomorphism λ is surjective, since $n = \lambda([\gamma_n])$ for all $n \in \mathbb{Z}$, where the loop $\gamma_n\colon [0, 1] \to S^1$ is given by $\gamma_n(t) = p(nt) = (\cos 2\pi nt, \sin 2\pi nt)$ for all $t \in [0, 1]$. We conclude that $\lambda\colon \pi_1(S^1, b) \to \mathbb{Z}$ is an isomorphism. ∎

The method used to prove Theorem 4.2 can be adapted in order to find the fundamental groups of other topological spaces such as the torus, the Klein bottle and the real projective plane. All these spaces are the base spaces of covering maps for which the covering space is simply-connected. In the case of the torus and the Klein bottle, the simply-connected covering space if homeomorphic to the plane $\mathbb{R}^2$. In the case of the real projective plane, the simply-connected covering space is homeomorphic to the 2-dimensional sphere S^2.

4.3 Homomorphisms of Fundamental Groups induced by Covering Maps

Proposition 4.3 *Let $p: \tilde{X} \to X$ be a covering map, and let $\tilde{x}_0$ be a point of the covering space $\tilde{X}$. Then the homomorphism*

$$p_\#: \pi_1(\tilde{X}, \tilde{x}_0) \to \pi_1(X, p(\tilde{x}_0))$$

of fundamental groups induced by the covering map p is injective.

Proof Let σ_0 and σ_1 be loops in $\tilde{X}$ based at the point $\tilde{x}_0$, representing elements $[\sigma_0]$ and $[\sigma_1]$ of $\pi_1(\tilde{X}, \tilde{x}_0)$. Suppose that $p_\#[\sigma_0] = p_\#[\sigma_1]$. Then $p \circ \sigma_0 \simeq p \circ \sigma_1$ rel $\{0, 1\}$. Also $\sigma_0(0) = \tilde{x}_0 = \sigma_1(0)$. Therefore $\sigma_0 \simeq \sigma_1$ rel $\{0, 1\}$, by Proposition 4.1, and thus $[\sigma_0] = [\sigma_1]$. We conclude that the homomorphism $p_\#: \pi_1(\tilde{X}, \tilde{x}_0) \to \pi_1(X, p(\tilde{x}_0))$ is injective. ∎

Proposition 4.4 *Let $p: \tilde{X} \to X$ be a covering map, let $\tilde{x}_0$ be a point of the covering space $\tilde{X}$, and let γ be a loop in X based at $p(\tilde{x}_0)$. Then $[\gamma] \in p_\#(\pi_1(\tilde{X}, \tilde{x}_0))$ if and only if there exists a loop $\tilde{\gamma}$ in $\tilde{X}$, based at the point $\tilde{x}_0$, such that $p \circ \tilde{\gamma} = \gamma$.*

Proof If $\gamma = p \circ \tilde{\gamma}$ for some loop $\tilde{\gamma}$ in $\tilde{X}$ based at $\tilde{x}_0$ then $[\gamma] = p_\#[\tilde{\gamma}]$, and therefore $[\gamma] \in p_\#(\pi_1(\tilde{X}, \tilde{x}_0))$.

Conversely suppose that $[\gamma] \in p_\#(\pi_1(\tilde{X}, \tilde{x}_0))$. We must show that there exists some loop $\tilde{\gamma}$ in $\tilde{X}$ based at $\tilde{x}_0$ such that $\gamma = p \circ \tilde{\gamma}$. Now there exists a loop σ in $\tilde{X}$ based at the point $\tilde{x}_0$ such that $[\gamma] = p_\#([\sigma])$ in $\pi_1(X, p(\tilde{x}_0))$. Then $\gamma \simeq p \circ \sigma$ rel $\{0, 1\}$. It follows from the Path Lifting Theorem for covering maps (Theorem 2.4) that there exists a unique path $\tilde{\gamma}: [0, 1] \to \tilde{X}$ in $\tilde{X}$ for which $\tilde{\gamma}(0) = \tilde{x}_0$ and $p \circ \tilde{\gamma} = \gamma$. It then follows from Proposition 4.1 that $\tilde{\gamma}(1) = \sigma(1)$ and $\tilde{\gamma} \simeq \sigma$ rel $\{0, 1\}$. But $\sigma(1) = \tilde{x}_0$. Therefore the path $\tilde{\gamma}$ is the required loop in $\tilde{X}$ based the point $\tilde{x}_0$ which satisfies $p \circ \tilde{\gamma} = \gamma$. ∎

Corollary 4.5 *Let $p: \tilde{X} \to X$ be a covering map over a topological space X, let w_0 and w_1 be points of $\tilde{X}$ satisfying $p(w_0) = p(w_1)$, and let $\alpha: [0, 1] \to \tilde{X}$ be a path in $\tilde{X}$ from w_0 to w_1. Suppose that $[p \circ \alpha] \in p_\#(\pi_1(\tilde{X}, w_0))$. Then the path α is a loop in $\tilde{X}$, and thus $w_0 = w_1$.*

Proof It follows from Proposition 4.4 that there exists a loop β based at w_0 satisfying $p \circ \beta = p \circ \alpha$. Then $\alpha(0) = \beta(0)$. Now Proposition 2.2 ensures that the lift to $\tilde{X}$ of any path in X is uniquely determined by its starting point. It follows that $\alpha = \beta$. But then the path α must be a loop in $\tilde{X}$, and therefore $w_0 = w_1$, as required. ∎

Corollary 4.6 *Let $p: \tilde{X} \to X$ be a covering map over a topological space X. Let $\alpha: [0,1] \to X$ and $\beta: [0,1] \to X$ be paths in X such that $\alpha(0) = \beta(0)$ and $\alpha(1) = \beta(1)$, and let $\alpha.\beta^{-1}$ be the loop in X defined such that*

$$(\alpha.\beta^{-1})(t) = \begin{cases} \alpha(2t) & \textit{if } 0 \le t \le \frac{1}{2}; \\ \beta(2-2t) & \textit{if } \frac{1}{2} \le t \le 1. \end{cases}$$

Let $\tilde{\alpha}: [0,1] \to \tilde{X}$ and $\tilde{\beta}: [0,1] \to \tilde{X}$ be paths in $\tilde{X}$ such that $p \circ \tilde{\alpha} = \alpha$, and $p \circ \tilde{\beta} = \beta$. Suppose that $\tilde{\alpha}(0) = \tilde{\beta}(0)$. Then $\tilde{\alpha}(1) = \tilde{\beta}(1)$ if and only if $[\alpha.\beta^{-1}] \in p_\#(\pi_1(\tilde{X}, \tilde{x}_0))$, where $\tilde{x}_0 = \tilde{\alpha}(0) = \tilde{\beta}(0)$.

Proof Suppose that $\tilde{\alpha}(1) = \tilde{\beta}(1)$. Then the concatenation $\tilde{\alpha}.\tilde{\beta}^{-1}$ is a loop in $\tilde{X}$ based at $\tilde{x}_0$, and $[\alpha.\beta^{-1}] = p_\#([\tilde{\alpha}.\tilde{\beta}^{-1}])$, and therefore $[\alpha.\beta^{-1}] \in p_\#(\pi_1(\tilde{X}, \tilde{x}_0))$.

Conversely suppose that $\tilde{\alpha}$ and $\tilde{\beta}$ are paths in $\tilde{X}$ satisfying $p \circ \tilde{\alpha} = \alpha$, $p \circ \tilde{\beta} = \beta$ and $\tilde{\alpha}(0) = \tilde{\beta}(0) = \tilde{x}_0$, and that $[\alpha.\beta^{-1}] \in p_\#(\pi_1(\tilde{X}, \tilde{x}_0))$. We must show that $\tilde{\alpha}(1) = \tilde{\beta}(1)$. Let $\gamma: [0,1] \to X$ be the loop based at $p(\tilde{x}_0)$ given by $\gamma = \alpha.\beta^{-1}$. Thus

$$\gamma(t) = \begin{cases} \alpha(2t) & \text{if } 0 \le t \le \frac{1}{2}; \\ \beta(2-2t) & \text{if } \frac{1}{2} \le t \le 1. \end{cases}$$

Then $[\gamma] \in p_\#(\pi_1(\tilde{X}, \tilde{x}_0))$. It follows from Proposition 4.4 that there exists a loop $\tilde{\gamma}$ in $\tilde{X}$ based at $\tilde{x}_0$ such that $p \circ \tilde{\gamma} = \gamma$. Let $\hat{\alpha}: [0,1] \to \tilde{X}$ and Let $\hat{\beta}: [0,1] \to \tilde{X}$ be the paths in $\tilde{X}$ defined such that $\hat{\alpha}(t) = \tilde{\gamma}(\frac{1}{2}t)$ and $\hat{\beta}(t) = \tilde{\gamma}(1 - \frac{1}{2}t)$ for all $t \in [0,1]$. Then

$$\tilde{\alpha}(0) = \hat{\alpha}(0) = \tilde{\beta}(0) = \hat{\beta}(0) = \tilde{x}_0,$$

$p \circ \hat{\alpha} = \alpha = p \circ \tilde{\alpha}$ and $p \circ \hat{\beta} = \beta = p \circ \tilde{\beta}$. But Proposition 2.2 ensures that the lift to $\tilde{X}$ of any path in X is uniquely determined by its starting point. Therefore $\tilde{\alpha} = \hat{\alpha}$ and $\tilde{\beta} = \hat{\beta}$. It follows that

$$\tilde{\alpha}(1) = \hat{\alpha}(1) = \tilde{\gamma}(\tfrac{1}{2}) = \hat{\beta}(1) = \tilde{\beta}(1),$$

as required. ∎

Theorem 4.7 *Let $p: \tilde{X} \to X$ be a covering map over a topological space X. Suppose that $\tilde{X}$ is path-connected and that X is simply-connected. Then the covering map $p: \tilde{X} \to X$ is a homeomorphism.*

Proof We show that the map $p: \tilde{X} \to X$ is a bijection. This map is surjective (since covering maps are by definition surjective). We must show that it is injective. Let w_0 and w_1 be points of $\tilde{X}$ with the property that $p(w_0) = p(w_1)$. Then there exists a path $\alpha: [0,1] \to \tilde{X}$ with $\alpha(0) = w_0$ and $\alpha(1) = w_1$, since $\tilde{X}$ is path-connected. Then $p \circ \alpha$ is a loop in X based at the point x_0, where $x_0 = p(w_0)$. However $\pi_1(X, p(w_0))$ is the trivial group, since X is simply-connected. It follows from Corollary 4.5 that the path α is a loop in $\tilde{X}$ based at w_0, and therefore $w_0 = w_1$. This shows that the the covering map $p: \tilde{X} \to X$ is injective. Thus the map $p: \tilde{X} \to X$ is a bijection, and thus has a well-defined inverse $p^{-1}: X \to \tilde{X}$. It now follows from Lemma 2.1 that $p: \tilde{X} \to X$ is a homeomorphism, as required. ∎

Let $p: \tilde{X} \to X$ be a covering map over some topological space X, and let x_0 be some chosen basepoint of X. We shall investigate the dependence of the subgroup $p_\#(\pi_1(\tilde{X}, \tilde{x}))$ of $\pi_1(X, x_0)$ on the choice of the point $\tilde{x}$ in $\tilde{X}$, where $\tilde{x}$ is chosen such that $p(\tilde{x}) = x_0$. We first introduce some concepts from group theory.

Let G be a group, and let H be a subgroup of G. Given any $g \in G$, let gHg^{-1} denote the subset of G defined by

$$gHg^{-1} = \{g' \in G : g' = ghg^{-1} \text{ for some } h \in H\}.$$

It is easy to verify that gHg^{-1} is a subgroup of G.

Definition Let G be a group, and let H and H' be subgroups of G. We say that H and H' are *conjugate* if and only if there exists some $g \in G$ for which $H' = gHg^{-1}$.

Note that if $H' = gHg^{-1}$ then $H = g^{-1}H'g$. The relation of conjugacy is an equivalence relation on the set of all subgroups of the group G. Moreover conjugate subgroups of G are isomorphic, since the homomorphism sending $h \in H$ to ghg^{-1} is an isomorphism from H to gHg^{-1} whose inverse is the homorphism sending $h' \in gHg^{-1}$ to $g^{-1}h'g$.

A subgroup H of a group G is said to be a *normal subgroup* of G if $ghg^{-1} \in H$ for all $h \in H$ and $g \in G$. If H is a normal subgroup of G then $gHg^{-1} \subset H$ for all $g \in G$. But then $g^{-1}Hg \subset H$ and $H = g(g^{-1}Hg)g^{-1}$ for all $g \in G$, and therefore $H \subset gHg^{-1}$ for all $g \in G$. It follows from this that a subgroup H of G is a normal subgroup if and only if $gHg^{-1} = H$ for all $g \in G$. Thus a subgroup H of G is a normal subgroup if and only if there is no other subgroup of G conjugate to H.

Lemma 4.8 *Let $p: \tilde{X} \to X$ be a covering map over a topological space X. Let x_0 be a point of X, and let w_0 and w_1 be points of $\tilde{X}$ for which $p(w_0) = x_0 = p(w_1)$. Let H_0 and H_1 be the subgroups of $\pi_1(X, x_0)$ defined by*

$$H_0 = p_\#(\pi_1(\tilde{X}, w_0)), \quad H_1 = p_\#(\pi_1(\tilde{X}, w_1)).$$

Suppose that the covering space $\tilde{X}$ is path-connected. Then the subgroups H_0 and H_1 of $\pi_1(X, x_0)$ are conjugate. Moreover if H is any subgroup of $\pi_1(X, x_0)$ which is conjugate to H_0 then there exists an element w of $\tilde{X}$ for which $p(w) = x$ and $p_\#(\pi_1(\tilde{X}, w)) = H$.

Proof Let $\alpha: [0,1] \to \tilde{X}$ be a path in $\tilde{X}$ for which $\alpha(0) = w_0$ and $\alpha(1) = w_1$. (Such a path exists since $\tilde{X}$ is path-connected.) Then each loop σ in $\tilde{X}$ based at w_1 determines a corresponding loop $\alpha.\sigma.\alpha^{-1}$ in $\tilde{X}$ based at w_0, where

$$(\alpha.\sigma.\alpha^{-1})(t) \equiv \begin{cases} \alpha(3t) & \text{if } 0 \leq t \leq \frac{1}{3}; \\ \sigma(3t-1) & \text{if } \frac{1}{3} \leq t \leq \frac{2}{3}; \\ \alpha(3-3t) & \text{if } \frac{2}{3} \leq t \leq 1. \end{cases}$$

(This loop traverses the path α from w_0 to w_1, then continues round the loop σ, and traverses the path α in the reverse direction in order to return from w_1 to w_0.) Let $\eta: [0,1] \to X$ be the loop in X based at the point x_0 given by $\eta = p \circ \alpha$, and let $\varphi: \pi_1(X, x_0) \to \pi_1(X, x_0)$ be the automorphism of the group $\pi_1(X, x_0)$ defined such that $\varphi([\gamma]) = [\eta][\gamma][\eta]^{-1}$ for all loops γ in X based at the point x_0. Then $p \circ (\alpha.\sigma.\alpha^{-1}) = \eta.(p \circ \sigma).\eta^{-1}$, and therefore $p_\#([\alpha.\sigma.\alpha^{-1}]) = [\eta] p_\#([\sigma])[\eta]^{-1} = \varphi(p_\#([\sigma]))$ in $\pi_1(X, x_0)$. It follows that $\varphi(H_1) \subset H_0$. Similarly $\varphi^{-1}(H_0) \subset H_1$, where $\varphi^{-1}([\gamma]) = [\eta]^{-1}[\gamma][\eta]$ for all loops γ in X based at the point x_0. It follows that $\varphi(H_1) = H_0$, and thus the subgroups H_0 and H_1 are conjugate

Now let H be a subgroup of $\pi_1(X, x_0)$ which is conjugate to H_0. Then $H_0 = [\eta] H [\eta]^{-1}$ for some loop η in X based at the point x_0. It follows from the Path Lifting Theorem for covering maps (Theorem 2.4) that there exists a path $\alpha: [0,1] \to \tilde{X}$ in $\tilde{X}$ for which $\alpha(0) = w_0$ and $p \circ \alpha = \eta$. Let $w = \alpha(1)$. Then

$$p_\#(\pi_1(\tilde{X}, w)) = [\eta]^{-1} H_0 [\eta] = H,$$

as required. ∎

4.4 Discontinuous Group Actions

Definition Let G be a group, and let X be a set. The group G is said to *act* on the set X (on the left) if each element g of G determines a corresponding function $\theta_g: X \to X$ from the set X to itself, where

(i) $\theta_{gh} = \theta_g \circ \theta_h$ for all $g, h \in G$;

(ii) the function θ_e determined by the identity element e of G is the identity function of X.

Let G be a group acting on a set X. Given any element x of X, the *orbit* $[x]_G$ of x (under the group action) is defined to be the subset $\{\theta_g(x) : g \in G\}$ of X, and the *stabilizer* of x is defined to the the subgroup $\{g \in G : \theta_g(x) = x\}$ of the group G. Thus the orbit of an element x of X is the set consisting of all points of X to which x gets mapped under the action of elements of the group G. The stabilizer of x is the subgroup of G consisting of all elements of this group that fix the point x. The group G is said to act *freely* on X if $\theta_g(x) \neq x$ for all $x \in X$ and $g \in G$ satisfying $g \neq e$. Thus the group G acts freely on X if and only if the stabilizer of every element of X is the trivial subgroup of G.

Let e be the identity element of G. Then $x = \theta_e(x)$ for all $x \in X$, and therefore $x \in [x]_G$ for all $x \in X$, where $[x]_G = \{\theta_g(x) : g \in G\}$.

Let x and y be elements of G for which $[x]_G \cap [y]_G$ is non-empty, and let $z \in [x]_G \cap [y]_G$. Then there exist elements h and k of G such that $z = \theta_h(x) = \theta_k(y)$. Then $\theta_g(z) = \theta_{gh}(x) = \theta_{gk}(y)$, $\theta_g(x) = \theta_{gh^{-1}}(z)$ and $\theta_g(y) = \theta_{gk^{-1}}(z)$ for all $g \in G$, and therefore $[x]_G = [z]_G = [y]_G$. It follows from this that the group action partitions the set X into orbits, so that each element of X determines an orbit which is the unique orbit for the action of G on X to which it belongs. We denote by X/G the set of orbits for the action of G on X.

Now suppose that the group G acts on a topological space X. Then there is a surjective function $q: X \to X/G$, where $q(x) = [x]_G$ for all $x \in X$. This surjective function induces a quotient topology on the set of orbits: a subset U of X/G is open in this quotient topology if and only if $q^{-1}(U)$ is an open set in X (see Lemma 1.22). We define the *orbit space* X/G for the action of G on X to be the topological space whose underlying set is the set of orbits for the action of G on X, the topology on X/G being the quotient topology induced by the function $q: X \to X/G$. This function $q: X \to X/G$ is then an identification map: we shall refer to it as the *quotient map* from X to X/G.

We shall be concerned here with situations in which a group action on a topological space gives rise to a covering map. The relevant group actions are those where the group acts *freely and properly discontinuously* on the topological space.

Definition Let G be a group with identity element e, and let X be a topological space. The group G is said to act *freely and properly discontinuously* on X if each element g of G determines a corresponding continuous map $\theta_g \colon X \to X$, where the following conditions are satisfied:

(i) $\theta_{gh} = \theta_g \circ \theta_h$ for all $g, h \in G$;

(ii) the continuous map θ_e determined by the identity element e of G is the identity map of X;

(iii) given any point x of X, there exists an open set U in X such that $x \in U$ and $\theta_g(U) \cap U = \emptyset$ for all $g \in G$ satisfying $g \neq e$.

Let G be a group which acts freely and properly discontinuously on a topological space X. Given any element g of G, the corresponding continuous function $\theta_g \colon X \to X$ determined by X is a homeomorphism. Indeed it follows from conditions (i) and (ii) in the above definition that $\theta_{g^{-1}} \circ \theta_g$ and $\theta_g \circ \theta_{g^{-1}}$ are both equal to the identity map of X, and therefore $\theta_g \colon X \to X$ is a homeomorphism with inverse $\theta_{g^{-1}} \colon X \to X$.

Remark The terminology 'freely and properly discontinuously' is traditional, but is hardly ideal. The adverb 'freely' refers to the requirement that $\theta_g(x) \neq x$ for all $x \in X$ and for all $g \in G$ satisfying $g \neq e$. The adverb 'discontinuously' refers to the fact that, given any point x of G, the elements of the orbit $\{\theta_g(x) : g \in G\}$ of x are separated; it does not signify that the functions defining the action are in any way discontinuous or badly-behaved. The adverb 'properly' refers to the fact that, given any compact subset K of X, the number of elements of g for which $K \cap \theta_g(K) \neq \emptyset$ is finite. Moreover the definitions of *properly discontinuous actions* in textbooks and in sources of reference are not always in agreement: some say that an action of a group G on a topological space X (where each group element determines a corresponding homeomorphism of the topological space) is *properly discontinuous* if, given any $x \in X$, there exists an open set U in X such that the number of elements g of the group for which $g(U) \cap U \neq \emptyset$ is finite; others say that the action is *properly discontinuous* if it satisfies the conditions given in the definition above for a group acting freely and properly discontinuously on the set. William Fulton, in his textbook *Algebraic topology: a first course* (Springer, 1995), introduced the term 'evenly' in place of 'freely and properly discontinuously', but this change in terminology does not appear to have been generally adopted.

Proposition 4.9 *Let G be a group acting freely and properly discontinuously on a topological space X. Then the quotient map $q \colon X \to X/G$ from X to the corresponding orbit space X/G is a covering map.*

Proof The quotient map $q: X \to X/G$ is surjective. Let V be an open set in X. Then $q^{-1}(q(V))$ is the union $\bigcup_{g \in G} \theta_g(V)$ of the open sets $\theta_g(V)$ as g ranges over the group G, since $q^{-1}(q(V))$ is the subset of X consisting of all elements of X that belong to the orbit of some element of V. But any union of open sets in a topological space is an open set. We conclude therefore that if V is an open set in X then $q(V)$ is an open set in X/G.

Let x be a point of X. Then there exists an open set U in X such that $x \in U$ and $\theta_g(U) \cap U = \emptyset$ for all $g \in G$ satisfying $g \neq e$. Now $q^{-1}(q(U)) = \bigcup_{g \in G} \theta_g(U)$. We claim that the sets $\theta_g(U)$ are disjoint. Let g and h be elements of G. Suppose that $\theta_g(U) \cap \theta_h(U) \neq \emptyset$. Then $\theta_{h^{-1}}(\theta_g(U) \cap \theta_h(U)) \neq \emptyset$. But $\theta_{h^{-1}}: X \to X$ is a bijection, and therefore

$$\theta_{h^{-1}}(\theta_g(U) \cap \theta_h(U)) = \theta_{h^{-1}}(\theta_g(U)) \cap \theta_{h^{-1}}(\theta_h(U)) = \theta_{h^{-1}g}(U) \cap U,$$

and therefore $\theta_{h^{-1}g}(U) \cap U \neq \emptyset$. It follows that $h^{-1}g = e$, where e denotes the identity element of G, and therefore $g = h$. Thus if g and h are elements of g, and if $g \neq h$, then $\theta_g(U) \cap \theta_h(U) = \emptyset$. We conclude therefore that the preimage $q^{-1}(q(U))$ of $q(U)$ is the disjoint union of the sets $\theta_g(U)$ as g ranges over the group G. Moreover each these sets $\theta_g(U)$ is an open set in X.

Now $U \cap [u]_G = \{u\}$ for all $u \in U$, since $[u]_G = \{\theta_g(u) : g \in G\}$ and $U \cap \theta_g(U) = \emptyset$ when $g \neq e$. Thus if u and v are elements of U, and if $q(u) = q(v)$ then $[u]_G = [v]_G$ and therefore $u = v$. It follows that the restriction $q|U: U \to X/G$ of the quotient map q to U is injective, and therefore q maps U bijectively onto $q(U)$. But q maps open sets onto open sets, and any continuous bijection that maps open sets onto open sets is a homeomorphism. We conclude therefore that the restriction of $q: X \to X/G$ to the open set U maps U homeomorphically onto $q(U)$. Moreover, given any element g of G, the quotient map q satisfies $q = q \circ \theta_{g^{-1}}$, and the homeomorphism $\theta_{g^{-1}}$ maps $\theta_g(U)$ homeomorphically onto U. It follows that the quotient map q maps $\theta_g(U)$ homeomorphically onto $q(U)$ for all $g \in U$. We conclude therefore that $q(U)$ is an evenly covered open set in X/G whose preimage $q^{-1}(q(U))$ is the disjoint union of the open sets $\theta_g(U)$ as g ranges over the group G. It follows that the quotient map $q: X \to X/G$ is a covering map, as required. ∎

Theorem 4.10 *Let G be a group acting freely and properly discontinuously on a path-connected topological space X, let $q: X \to X/G$ be the quotient map from X to the orbit space X/G, and let x_0 be a point of X. Then there exists a surjective homomorphism $\lambda: \pi_1(X/G, q(x_0)) \to G$ with the property that $\tilde{\gamma}(1) = \theta_{\lambda([\gamma])}(x_0)$ for any loop γ in X/G based at $q(x_0)$, where $\tilde{\gamma}$ denotes the unique path in X for which $\tilde{\gamma}(0) = x_0$ and $q \circ \tilde{\gamma} = \gamma$. The kernel of this homomorphism is the subgroup $q_\#(\pi_1(X, x_0))$ of $\pi_1(X/G, q(x_0))$.*

Proof Let $\gamma\colon [0,1] \to X/G$ be a loop in the orbit space with $\gamma(0) = \gamma(1) = q(x_0)$. It follows from the Path Lifting Theorem for covering maps (Theorem 2.4) that there exists a unique path $\tilde{\gamma}\colon [0,1] \to X$ for which $\tilde{\gamma}(0) = x_0$ and $q \circ \tilde{\gamma} = \gamma$. Now $\tilde{\gamma}(0)$ and $\tilde{\gamma}(1)$ must belong to the same orbit, since $q(\tilde{\gamma}(0)) = \gamma(0) = \gamma(1) = q(\tilde{\gamma}(1))$. Therefore there exists some element g of G such that $\tilde{\gamma}(1) = \theta_g(x_0)$. This element g is uniquely determined, since the group G acts freely on X. Moreover the value of g is determined by the based homotopy class $[\gamma]$ of γ in $\pi_1(X/G, q(x_0))$. Indeed it follows from Proposition 4.1 that if σ is a loop in X/G based at $q(x_0)$, if $\tilde{\sigma}$ is the lift of σ starting at x_0 (so that $q \circ \tilde{\sigma} = \sigma$ and $\tilde{\sigma}(0) = x_0$), and if $[\gamma] = [\sigma]$ in $\pi_1(X/G, q(x_0))$ (so that $\gamma \simeq \sigma$ rel $\{0,1\}$), then $\tilde{\gamma}(1) = \tilde{\sigma}(1)$. We conclude therefore that there exists a well-defined function

$$\lambda\colon \pi_1(X/G, q(x_0)) \to G,$$

which is characterized by the property that $\tilde{\gamma}(1) = \theta_{\lambda([\gamma])}(x_0)$ for any loop γ in X/G based at $q(x_0)$, where $\tilde{\gamma}$ denotes the unique path in X for which $\tilde{\gamma}(0) = x_0$ and $q \circ \tilde{\gamma} = \gamma$.

Now let $\alpha\colon [0,1] \to X/G$ and $\beta\colon [0,1] \to X/G$ be loops in X/G based at x_0, and let $\tilde{\alpha}\colon [0,1] \to X$ and $\tilde{\beta}\colon [0,1] \to X$ be the lifts of α and β respectively starting at x_0, so that $q \circ \tilde{\alpha} = \alpha$, $q \circ \tilde{\beta} = \beta$ and $\tilde{\alpha}(0) = \tilde{\beta}(0) = x_0$. Then $\tilde{\alpha}(1) = \theta_{\lambda([\alpha])}(x_0)$ and $\tilde{\beta}(1) = \theta_{\lambda([\beta])}(x_0)$. Then the path $\theta_{\lambda([\alpha])} \circ \tilde{\beta}$ is also a lift of the loop β, and is the unique lift of β starting at $\tilde{\alpha}(1)$. Let $\alpha.\beta$ be the concatenation of the loops α and β, where

$$(\alpha.\beta)(t) = \begin{cases} \alpha(2t) & \text{if } 0 \le t \le \frac{1}{2}; \\ \beta(2t-1) & \text{if } \frac{1}{2} \le t \le 1. \end{cases}$$

Then the unique lift of $\alpha.\beta$ to X starting at x_0 is the path $\sigma\colon [0,1] \to X$, where

$$\sigma(t) = \begin{cases} \tilde{\alpha}(2t) & \text{if } 0 \le t \le \frac{1}{2}; \\ \theta_{\lambda([\alpha])}(\tilde{\beta}(2t-1)) & \text{if } \frac{1}{2} \le t \le 1. \end{cases}$$

It follows that

$$\begin{aligned} \theta_{\lambda([\alpha][\beta])}(x_0) &= \theta_{\lambda([\alpha.\beta])}(x_0) = \sigma(1) = \theta_{\lambda([\alpha])}(\tilde{\beta}(1)) \\ &= \theta_{\lambda([\alpha])}(\theta_{\lambda([\beta])}(x_0)) = \theta_{\lambda([\alpha])\lambda([\beta])}(x_0) \end{aligned}$$

and therefore $\lambda([\alpha][\beta]) = \lambda([\alpha])\lambda([\beta])$. Therefore the function

$$\lambda\colon \pi_1(X/G, q(x_0)) \to G$$

is a homomorphism.

Let $g \in G$. Then there exists a path α in X from x_0 to $\theta_g(x_0)$, since the space X is path-connected. Then $q \circ \alpha$ is a loop in X/G based at $q(x_0)$, and $g = \lambda([q \circ \alpha])$. This shows that the homomorphism λ is surjective.

Let $\gamma\colon [0,1] \to X/G$ be a loop in X/G based at $q(x_0)$. Suppose that $[\gamma] \in \ker \lambda$. Then $\tilde{\gamma}(1) = \theta_e(x_0) = x_0$, and therefore $\tilde{\gamma}$ is a loop in X based at x_0. Moreover $[\gamma] = q_\#[\tilde{\gamma}]$, and therefore $[\gamma] \in q_\#(\pi_1(X, x_0))$. On the other hand, if $[\gamma] \in q_\#(\pi_1(X, x_0))$ then $\gamma = q \circ \tilde{\gamma}$ for some loop $\tilde{\gamma}$ in X based at x_0 (see Proposition 4.4). But then $x_0 = \tilde{\gamma}(1) = \theta_{\lambda([\gamma])}(x_0)$, and therefore $\lambda([\gamma]) = e$, where e is the identity element of G. Thus $\ker \lambda = q_\#(\pi_1(X, x_0))$, as required. ∎

Corollary 4.11 *Let G be a group acting freely and properly discontinuously on a path-connected topological space X, let $q\colon X \to X/G$ be the quotient map from X to the orbit space X/G, and let x_0 be a point of X. Then $q_\#(\pi_1(X, x_0))$ is a normal subgroup of the fundamental group $\pi_1(X/G, q(x_0))$ of the orbit space, and*

$$\frac{\pi_1(X/G, q(x_0))}{q_\#(\pi_1(X, x_0))} \cong G.$$

Proof The subgroup $q_\#(\pi_1(X, x_0))$ is the kernel of the homomorphism

$$\lambda\colon \pi_1(X/G, q(x_0)) \to G$$

described in the statement of Theorem 4.10. It is therefore a normal subgroup of $\pi_1(X/G, q(x_0))$, since the kernel of any homomorphism is a normal subgroup. The homomorphism λ is surjective, and the image of any group homomorphism is isomorphism of the quotient of its domain by its kernel. The result follows. ∎

Corollary 4.12 *Let G be a group acting freely and properly discontinuously on a simply-connected topological space X, let $q\colon X \to X/G$ be the quotient map from X to the orbit space X/G, and let x_0 be a point of X. Then $\pi_1(X/G, q(x_0)) \cong G$.*

Proof This is a special case of Corollary 4.11. ∎

Example The group $\mathbb{Z}$ of integers under addition acts freely and properly discontinuously on the real line $\mathbb{R}$. Indeed each integer n determines a corresponding homeomorphism $\theta_n\colon \mathbb{R} \to \mathbb{R}$, where $\theta_n(x) = x + n$ for all $x \in \mathbb{R}$. Moreover $\theta_m \circ \theta_n = \theta_{m+n}$ for all $m, n \in \mathbb{Z}$, and θ_0 is the identity map of $\mathbb{R}$. If $U = (-\frac{1}{2}, \frac{1}{2})$ then $\theta_n(U) \cap U = \emptyset$ for all non-zero integers n. The real line $\mathbb{R}$ is simply-connected. It follows from Corollary 4.12 that $\pi_1(\mathbb{R}/\mathbb{Z}, b) \cong \mathbb{Z}$ for any point b of $\mathbb{R}/\mathbb{Z}$.

Now the orbit space $\mathbb{R}/\mathbb{Z}$ is homeomorphic to a circle. Indeed let $q: \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ be the quotient map. Then the surjective function $p: \mathbb{R} \to S^1$ which sends $t \in \mathbb{R}$ to $(\cos 2\pi t, \sin 2\pi t)$ induces a continuous map $h: \mathbb{R}/\mathbb{Z} \to S^1$ defined on the orbit space which satisfies $h \circ q = p$, since the quotient map q is an identification map. Moreover real numbers t_1 and t_2 satisfy $p(t_1) = p(t_2)$ if and only if $q(t_1) = q(t_2)$. It follows that the induced map $h: \mathbb{R}/\mathbb{Z} \to S^1$ is a bijection. This map also maps open sets to open sets, for if W is any open set in the orbit space $\mathbb{R}/\mathbb{Z}$ then $q^{-1}(W)$ is an open set in $\mathbb{R}$, and therefore $p(q^{-1}(W))$ is an open set in S^1, since the covering map $p: \mathbb{R} \to S^1$ maps open sets to open sets (Lemma 2.1). But $p(q^{-1}(W)) = h(W)$ for all open sets W in $\mathbb{R}/\mathbb{Z}$. Thus the continuous bijection $h: \mathbb{R}/\mathbb{Z} \to S^1$ maps open sets to open sets, and is therefore a homeomorphism. It follows from Corollary 4.12 that $\pi_1(S^1, b) \cong \mathbb{Z}$ for any point b of the circle S^1. This shows that Theorem 4.2 concerning the fundamental group of the circle can be obtained as a special case of the more general result Corollary 4.12 concerning fundamental groups of orbit spaces obtained via discontinuous group actions on simply connected topological spaces.

Example The group $\mathbb{Z}^n$ of ordered n-tuples of integers under addition acts freely and properly discontinuously on $\mathbb{R}^n$, where

$$\theta_{(m_1,m_2,\ldots,m_n)}(x_1, x_2, \ldots, x_n) = (x_1 + m_1, x_2 + m_2, \ldots, x_n + m_n)$$

for all $(m_1, m_2, \ldots, m_n) \in \mathbb{Z}^n$ and $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. The orbit space $\mathbb{R}^n/\mathbb{Z}^n$ is an n-dimensional torus, homeomorphic to the product of n circles. It follows from Corollary 4.12 that the fundamental group of this n-dimensional torus is isomorphic to the group $\mathbb{Z}^n$.

Example Let C_2 be the cyclic group of order 2. Then $C_2 = \{e, a\}$ where e is the identity element, $a \neq e$, $a^2 = e$. Then the group C_2 acts freely and properly discontinuously on the n-dimensional sphere S^n for each non-negative integer n. We represent S^n as the unit sphere centred on the origin in $\mathbb{R}^{n+1}$. The homeomorphism θ_e determined by the identity element e of C_2 is the identity map of S^n; the homeomorphism θ_a determined by the element a of C_2 is the antipodal map that sends each point $\mathbf{x}$ of S^n to $-\mathbf{x}$. The orbit space S^n/C_2 is homeomorphic to real projective n-dimensional space $\mathbb{R}P^n$. The n-dimensional sphere is simply-connected if $n > 1$. It follows from Corollary 4.12 that the fundamental group of $\mathbb{R}P^n$ is isomorphic to the cyclic group C_2 when $n > 1$.

Note that S^0 is a pair of points, and $\mathbb{R}P^0$ is a single point. Also S^1 is a circle (which is not simply-connected) and $\mathbb{R}P^1$ is homeomorphic to a circle. Moreover, for any $b \in S^1$, the homomorphism $q_\#: \pi_1(S^1, b) \to \pi_1(\mathbb{R}P^1, q(b))$

corresponds to the homomorphism from $\mathbb{Z}$ to $\mathbb{Z}$ that sends each integer n to $2n$. This is consistent with the conclusions of Corollary 4.11 in this example.

Example Given a pair (m, n) of integers, let $\theta_{m,n}: \mathbb{R}^2 \to \mathbb{R}^2$ be the homeomorphism of the plane $\mathbb{R}^2$ defined such that $\theta_{m,n}(x, y) = (x + m, (-1)^m y + n)$ for all $(x, y) \in \mathbb{R}^2$. Let (m_1, n_1) and (m_2, n_2) be ordered pairs of integers. Then $\theta_{m_1,n_1} \circ \theta_{m_2,n_2} = \theta_{m_1+m_2,n_1+(-1)^{m_1}n_2}$. Let Γ be the group whose elements are represented as ordered pairs of integers, where the group operation $\#$ on Γ is defined such that

$$(m_1, n_1)\#(m_2, n_2) = (m_1 + m_2, n_1 + (-1)^{m_1} n_2)$$

for all $(m_1, n_1), (m_2, n_2) \in \Gamma$. The group Γ is non-Abelian, and its identity element is $(0, 0)$. This group acts on the plane $\mathbb{R}^2$: given $(m, n) \in \Gamma$ the corresponding symmetry $\theta_{m,n}$ is a translation if m is even, and is a glide reflection if m is odd. Given a pair (m, n) of integers, the corresponding homeomorphism $\theta_{m,n}$ maps an open disk about the point (x, y) onto an open disk of the same radius about the point $\theta_{(m,n)}(x, y)$. It follows that if D is the open disk of radius $\frac{1}{2}$ about the point (x, y), and if $D \cap \theta_{m,n}(D)$ is non-empty, then $(m, n) = (0, 0)$. Thus the group Γ maps freely and properly discontinuously on the plane $\mathbb{R}^2$.

The orbit space $\mathbb{R}^2/\Gamma$ is homeomorphic to a Klein bottle. To see this, note each orbit intersects the closed unit square S, where $S = [0, 1] \times [0, 1]$. If $0 < x < 1$ and $0 < y < 1$ then the orbit of (x, y) intersects the square S in one point, namely the point (x, y). If $0 < x < 1$, then the orbit of $(x, 0)$ intersects the square in two points $(x, 0)$ and $(x, 1)$. If $0 < y < 1$ then the orbit of $(0, y)$ intersects the square S in the two points $(0, y)$ and $(1, 1 - y)$. (Note that $(1, 1-y) = \theta_{1,1}(0, y)$.) And the orbit of any corner of the square S intersects the square in the four corners of the square. The restriction $q|S$ of the quotient map $q: \mathbb{R}^2 \to \mathbb{R}^2/\Gamma$ to the square S is a continuous surjection defined on the square: one can readily verify that it is an identification map. It follows that the orbit space $\mathbb{R}^2/\Gamma$ is homeomorphic to the identification space obtained from the closed square S by identifying together the points $(x, 0)$ and $(x, 1)$ where the real number x satisfies $0 < x < 1$, identifying together the points $(0, y)$ and $(1, 1 - y)$ where the real number y satisfies $0 < y < 1$, and identifying together the four corners of the square: this identification space is the Klein bottle.

The plane $\mathbb{R}^2$ is simply-connected. It follows from Corollary 4.12 that the fundamental group of the Klein bottle is isomorphic to the group Γ defined above.

4.5 The Brouwer Fixed Point Theorem in Two Dimensions

Theorem 4.13 *Let $f: D \to D$ be a continuous map which maps the closed disk D into itself. Then $f(\mathbf{x}_0) = \mathbf{x}_0$ for some $\mathbf{x}_0 \in D$.*

Proof Let ∂D denote the boundary circle of D. The inclusion map $i: \partial D \hookrightarrow D$ induces a corresponding homomorphism $i_\#: \pi_1(\partial D, \mathbf{b}) \to \pi_1(D, \mathbf{b})$ of fundamental groups for any $\mathbf{b} \in \partial D$.

Suppose that it were the case that the map f has no fixed point in D. Then one could define a continuous map $r: D \to \partial D$ as follows: for each $\mathbf{x} \in D$, let $r(\mathbf{x})$ be the point on the boundary ∂D of D obtained by continuing the line segment joining $f(\mathbf{x})$ to $\mathbf{x}$ beyond $\mathbf{x}$ until it intersects ∂D at the point $r(\mathbf{x})$. Note that $r|\partial D$ is the identity map of ∂D.

Let $r_\#: \pi_1(D, \mathbf{b}) \to \pi_1(\partial D, \mathbf{b})$ be the homomorphism of fundamental groups induced by $r: D \to \partial D$. Now $(r \circ i)_\#: \pi_1(\partial D, \mathbf{b}) \to \pi_1(\partial D, \mathbf{b})$ is the identity isomorphism of $\pi_1(\partial D, \mathbf{b})$, since $r \circ i: \partial D \to \partial D$ is the identity map. But it follows directly from the definition of induced homomorphisms that $(r \circ i)_\# = r_\# \circ i_\#$. Therefore $i_\#: \pi_1(\partial D, \mathbf{b}) \to \pi_1(D, \mathbf{b})$ is injective, and $r_\#: \pi_1(D, \mathbf{b}) \to \pi_1(\partial D, \mathbf{b})$ is surjective. But this is impossible, since $\pi_1(\partial D, \mathbf{b}) \cong \mathbb{Z}$ (Theorem 4.2) and $\pi_1(D, \mathbf{b})$ is the trivial group. This contradiction shows that the continuous map $f: D \to D$ must have at least one fixed point. ∎

5 The Classification of Surfaces

5.1 Triangulated Closed Surfaces

Let $\mathbf{v}$ and $\mathbf{w}$ be distinct points of some Euclidean space $\mathbb{R}^k$. Then $\mathbf{v}$ and $\mathbf{w}$ are the *vertices* (or *endpoints*) of an *edge* E in $\mathbb{R}^k$, where

$$E = \{r\mathbf{v} + s\mathbf{w} : s, t \in [0, 1] \text{ and } s + t = 1\}.$$

We denote by $\mathbf{v}\,\mathbf{w}$ the edge in $\mathbb{R}^k$ whose vertices (or endpoints) are the points $\mathbf{v}$ and $\mathbf{w}$.

Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be points of some Euclidean space $\mathbb{R}^k$ which are not colinear. These points are then the vertices of a *triangle* T in $\mathbb{R}^k$, where

$$T = \{r\mathbf{u} + s\mathbf{v} + t\mathbf{w}; r, s, t \in [0, 1] \text{ and } r + s + r = 1\}.$$

We denote by $\mathbf{u}\,\mathbf{v}\,\mathbf{w}$ the triangle in $\mathbb{R}^k$ whose vertices are the points $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ and whose edges are $\mathbf{u}\,\mathbf{v}$, $\mathbf{v}\,\mathbf{w}$ and $\mathbf{w}\,\mathbf{u}$.

Any edge or triangle in a Euclidean space is a compact subset of that Euclidean space.

Definition A *two-dimensional simplicial complex* in a Euclidean space consists of a finite collection K of triangles, edges (which are line segments) and vertices (which are points) in that space which contains at least one triangle, and which satisfies the following conditions:

(i) The edges and vertices of any triangle belonging to K themselves belong to K;

(ii) The endpoints of any edge belonging to K are vertices belonging to K;

(iii) if two distinct triangles belonging to K have a non-empty intersection, then that intersection is either a single common edge or a single common vertex of both triangles;

(iv) if two distinct edges belonging to K have a non-empty intersection then that intersection is a common vertex (or endpoint) of both edges.

Definition Let K be a two-dimensional simplicial complex in some Euclidean space. The *polyhedron* $|K|$ of K is the union of all the triangles, edges and vertices belonging to the collection K.

Lemma 5.1 *The polyhedron of a two-dimensional simplicial complex is a compact Hausdorff space.*

Proof The simplicial complex K is a finite collection of triangles, edges and vertices in some ambient Euclidean space, and each triangle, edge and vertex in the collection is a closed bounded subset of this ambient Euclidean space. Now a subset of a Euclidean space is compact if and only if it is both closed and bounded. It follows that each of the triangles, edges and vertices belonging to K is a compact subset of the ambient Euclidean space. Moreover it follows directly from the definition of compactness that any finite union of compact topological spaces is itself compact. Therefore the polyhedron $|K|$ of K is a compact subset of the ambient Euclidean space. This ambient Euclidean space is a Hausdorff space (as it is a metric space, and all metric spaces are Hausdorff spaces), and any subset of a Hausdorff space is itself a Hausdorff space (with the subspace topology). Therefore the polyhedron $|K|$ of K is a compact Hausdorff space, as required. ∎

Definition Let $\mathbf{p}$ be a point of the polyhedron $|K|$ of the two-dimensional simplicial complex K. The *star neighbourhood* $\mathrm{st}_K(\mathbf{p})$ of the point $\mathbf{p}$ in $|K|$ is defined to be the subset of $|K|$ whose complement is the union of all triangles, edges and vertices belonging to K that do not contain the point $\mathbf{p}$.

Lemma 5.2 *Let K be a two-dimensional simplicial complex, and let $\mathbf{p}$ be a point of K. Then the star neighbourhood $\mathrm{st}_K(\mathbf{p})$ of the point $\mathbf{p}$ of $|K|$ is an open subset of $|K|$, and moreover $\mathbf{p} \in \mathrm{st}_K(\mathbf{p})$.*

Proof A two-dimensional simplicial complex is a finite collection of triangles, edges and vertices in some ambient Euclidean space. Each of those triangles, edges and vertices is a closed subset of the ambient Euclidean space, and therefore the union of any finite collection of such triangles, edges and vertices is a closed subset of the ambient Euclidean space.

Now, given any point $\mathbf{p}$ of $|K|$, the complement $|K| \setminus \mathrm{st}_K(\mathbf{p})$ of the star neighbourhood $\mathrm{st}_K(\mathbf{p})$ of $\mathbf{p}$ in $|K|$ is by definition the union of all triangles, edges and vertices belonging to K that do not contain the point $\mathbf{p}$. It follows that $|K| \setminus \mathrm{st}_K(\mathbf{p})$ is closed in $|K|$, and $\mathbf{p} \notin |K| \setminus \mathrm{st}_K(\mathbf{p})$. Therefore $\mathrm{st}_K(\mathbf{p})$ is open in $|K|$, and $\mathbf{p} \in \mathrm{st}_K(\mathbf{p})$, as required. ∎

5.2 Triangulated Closed Surfaces

Definition A *topological closed surface* is a compact Hausdorff space that may be covered by open sets, where each of these open sets is homeomorphic to a open set in the Euclidean plane.

An open set in the Euclidean plane is a union of open disks in that plane. It follows that a compact Hausdorff space is a topological closed surface if and only if it can be covered by open sets, where each of these open sets is homeomorphic to a open disk in the Euclidean plane.

Proposition 5.3 *Let K be a two-dimensional simplicial complex which satisfies the following two conditions:—*

(i) *every edge belonging to K is an edge of exactly two triangles belonging to K;*

(ii) *given any vertex $\mathbf{v}$ belonging to K, the triangles that have $\mathbf{v}$ as vertex can be listed as a finite sequence $T_1, T_2, \ldots, T_m$, where $m > 1$, where T_i and T_{i-1} intersect along a common edge when $1 < i \leq m$, and where T_m and T_1 also intersect along a common edge.*

Then the polyhedron $|K|$ of K is a topological closed surface.

Proof The polyhedron $|K|$ of the two-dimensional simplicial complex K is a compact Hausdorff space. We shall prove that the star neighbourhood of each point of $|K|$ is homeomorphic to an open disk.

Now suppose that the point $\mathbf{p}$ belongs to a triangle T of K with vertices $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ but does not lie on any edge of that triangle. Then the triangle T is the only member of the collection K of triangles, edges and vertices that contains the point $\mathbf{p}$. It follows that the star neighbourhood $\mathrm{st}_K(\mathbf{p})$ consists of all points of the triangle T that do not lie on any edge of T. Thus $\mathrm{st}_K(\mathbf{p})$ is homeomorphic to the interior of a triangle in the Euclidean plane.

Next suppose that the point $\mathbf{p}$ belongs to an edge of K with vertices $\mathbf{v}$ and $\mathbf{w}$ but is not an endpoint of that edge. The edge is an edge of exactly two triangles belonging to K, because K represents a triangulated closed surface. Let these two triangles be $\mathbf{v}\,\mathbf{w}\,\mathbf{x}$ and $\mathbf{v}\,\mathbf{w}\,\mathbf{y}$. The conditions in the definition of two-dimensional complex ensure that the only members of the collection K that contain the point $\mathbf{p}$ are the edge $\mathbf{v}\,\mathbf{w}$ and the two triangles $\mathbf{v}\,\mathbf{w}\,\mathbf{x}$ and $\mathbf{v}\,\mathbf{w}\,\mathbf{y}$. It follows that the star neighbourhood $\mathrm{st}_K(\mathbf{p})$ of the point $\mathbf{p}$ in $|K|$ consists of all points of the union of these two triangles that do not lie on any of the edges $\mathbf{v}\,\mathbf{x}$, $\mathbf{x}\,\mathbf{w}$, $\mathbf{w}\,\mathbf{y}$ and $\mathbf{y}\,\mathbf{v}$. It follows from this that $\mathrm{st}_K(\mathbf{p})$ is homeomorphic to the interior of a quadrilateral in the Euclidean plane.

Finally suppose that $\mathbf{v}$ is a vertex belonging to K. Then the triangles that have $\mathbf{v}$ as vertex can be listed as a finite sequence $T_1, T_2, \ldots, T_m$, where $m > 1$, where T_i and T_{i-1} intersect along a common edge when $1 < i \leq m$, and where T_m and T_1 also intersect along a common edge. Let $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m$ be the vertices of these triangles distinct from $\mathbf{v}$, ordered so that the triangles T_m and T_1 intersect along the edge $\mathbf{v}\,\mathbf{w}_1$ and the triangles T_i and T_{i-1} intersect along the edge $\mathbf{v}\,\mathbf{w}_i$ for $i < i \leq m$. Then T_i is the triangle $\mathbf{v}\,\mathbf{w}_i\,\mathbf{w}_{i+1}$ for

$i = 1, 2, \ldots, m-1$, and T_m is the triangle $\mathbf{v}\,\mathbf{w}_m\,\mathbf{w}_1$. The triangles of K that have $\mathbf{v}$ as a vertex are thus in the configuration depicted in Figure 1. The union of these triangles $T_1, T_2, \ldots, T_m$ is then homeomorphic to a convex

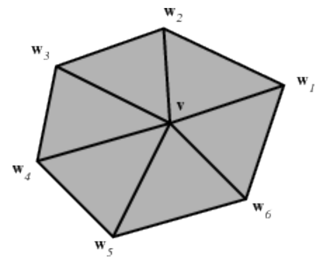

Figure 1

polygon in the Euclidean plane. The union of those edges

$$\mathbf{w}_m\,\mathbf{w}_1, \quad \mathbf{w}_1\,\mathbf{w}_2, \quad \cdots \quad \mathbf{w}_{m-1}\,\mathbf{w}_m$$

of these triangles that do not have $\mathbf{v}$ as one endpoint corresponds under this homeomorphism to the boundary of the convex polygon, and therefore the star neighbourhood $\mathrm{st}_K(\mathbf{v})$ of $\mathbf{v}$ in $|K|$ is homeomorphic to the interior of a convex polygon in the Euclidean plane.

We have thus shown that, given any point $\mathbf{p}$ of the polyhedron of K, the star neighbourhood of the point $\mathbf{p}$ is an open set in $|K|$ which is homeomorphic to the interior of a convex polygon in the Euclidean plane. The interior of such a polygon is homeomorphic to a disk. The result follows. ∎

Lemma 5.4 *Let K be a two-dimensional simplicial complex which satisfies the two conditions listed in the statement of Proposition 5.3 that ensure that the polyhedron $|K|$ of K is a topological closed surface. Then this polyhedron is a connected topological space if and only if, given any two triangles σ and τ of K, we can find a sequence $\sigma_1, \sigma_2, \ldots, \sigma_k$ of triangles of K with $\sigma = \sigma_1$ and $\tau = \sigma_k$, where σ_{i-1} and σ_i intersect in a common edge for $i = 2, 3, \ldots, k$.*

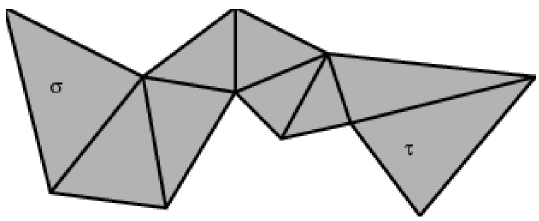

Figure 2

Proof Let σ_0 be a triangle in K, and let F be the subset of the polyhedron $|K|$ of K which is the union of all triangles that can be joined to σ_0 by a finite sequence of triangles belonging to K, where successive triangles in this sequence intersect along a common edge. Then F is a finite union of triangles, and those trianges are closed subsets of $|K|$, and therefore F is itself a closed subset of $|K|$.

Let $\mathbf{p}$ be a point of F. If $\mathbf{p}$ does not lie on any edge belonging to K then the star neighbourhood $\mathrm{st}_K(\mathbf{p})$ belongs to just one triangle belonging to K, and moreover this triangle must then be a subset of F (or else the point $\mathbf{p}$ would not belong to F). Thus if $\mathbf{p} \in F$ does not like on any edge belonging to K then $\mathrm{st}_K(\mathbf{p}) \subset F$.

Next suppose that the point $\mathbf{p}$ of F lies on some edge belonging to K but is not an endpoint of that edge. Then the point $\mathbf{p}$ belongs to exactly two triangles of K that intersect along a common edge (because the two-dimensional simplicial complex represents a closed surface). At least one of these triangles must be contained in the set F (since $\mathbf{p} \in F$) and therefore both triangles are contained in F. But the star neighbourhood of the point $\mathbf{p}$ is contained in the union of those two triangles. Therefore $\mathrm{st}_K(\mathbf{p}) \subset F$ in this case also.

Finally suppose that the point $\mathbf{p}$ is a vertex of K. Then the requirement that the two-dimensional simplicial complex K represent a triangulated closed surface ensures that if at least one of the triangles belonging to K with a vertex at $\mathbf{p}$ is contained in F then every triangle belonging to K with a vertex at $\mathbf{v}$ must be contained in F. It follows that $\mathrm{st}_K(\mathbf{p}) \subset F$.

We have now shown that, given any point $\mathbf{p}$ of F, the star neighbourhood $\mathrm{st}_K(\mathbf{p})$ of $\mathbf{p}$ in $|K|$ is a subset of F. But this star neighbourhood is an open subset of $|K|$ (see Lemma 5.2). Therefore the subset F of $|K|$ is both open and closed in $|K|$. Thus if the topological space $|K|$ is connected then $F = |K|$.

Every point of a topological space belongs to unique connected component which is the union of all connected subsets of the topological space that contain the given point. It follows that every triangle belonging to K is contained in a some connected component of $|K|$, and if two triangles belonging to K intersect along a common edge, or at a common vertex, then both belong to the same connected component of $|K|$. It follows that the set F is contained in some connected component of $|K|$. Thus if the topological space $|K|$ is not connected then F is a proper subset of $|K|$. We deduce that $F = |K|$ if and only if $|K|$ is a connected topological space. The result follows. ∎

Lemma 5.5 *Let K be a triangulated closed surface whose polyhedron $|K|$ is* *a connected topological space. Then $|K|$ is homeomorphic to the topological space obtained from a filled polygon with an even number of edges by identifying edges in pairs (i.e., given any edge with endpoints* $\mathbf{a}$ *and* $\mathbf{b}$*, there exists exactly one other edge with endpoints* $\mathbf{c}$ *and* $\mathbf{d}$ *such that* $(1-t)\mathbf{a}+t\mathbf{b}$ *is identified with* $(1-t)\mathbf{c}+t\mathbf{d}$ *for all* $t \in [0,1]$*).*

Proof Suppose that we have constructed some subcomplex L of K whose polyhedron is homeomorphic to the identification space obtained from a filled polygon P_L by identifying some of the edges of that polygon in pairs. Let $q_L: P_L \to L$ denote the identification map. Suppose that e is an edge of P_L that is not identified to any other edge of L. Then e corresponds under the identification map to some edge e' of P_L. Moreover only one of the two triangles in K adjoining the edge e' belongs to L. Thus there is some triangle σ of $K \setminus L$ which has e' as one of its edges. Let M be the subcomplex of K obtained on adjoining to L the triangle σ, together with all its edges and vertices.

We now extend the polygon P_L by attaching a triangle T along the free edge e to obtain a filled polygon P_M, where $P_M = P_L \cup T$ and $P_L \cap T = e$. We also extend the identification map $q_L: P_L \to |L|$ over this attached triangle to obtain an identification map $q_M: P_M \to |M|$, where $q_M|P_L = q_L$ and $q_M|T$ is a simplicial homeomorphism mapping the triangle T onto σ. Then the new identification map $q_M: P_M \to M$ also identifies some of the edges of the polygon P_M in pairs. Thus, by successively adding triangles in this fashion, we eventually obtain a subcomplex L of K whose polyhedron is homeomorphic to the identification space obtained from a filled polygon on identifying all of the edges of that polygon in pairs. But then, given any two triangles of K that intersect along a common edge, either both triangles belong to L, or else neither triangle belongs to L. It now follows from Lemma 5.4 that $L = K$, and thus the polyhedron of K is an identification space of the prescribed type. ∎

5.3 The Topological Classification of Closed Surfaces

We wish to classify up to homeomorphism the identification spaces obtained from polygons by identifying edges in pairs.

Suppose that we are given a polygon with its edges identified in pairs. Choose an orientation (i.e., a 'direction') on each edge of the polygon in such a way that, for each pair of identified edges, the orientations on those edges correspond under the identification of these edges. Denote each pair of identified edges by some letter $a, b, c, \ldots$. Suppose that we travel round the boundary of the polygon in the anticlockwise direction, starting at some chosen vertex. We obtain a *surface symbol* consisting of a sequence of symbols taken from $a, a^{-1}, b, b^{-1}, c, c^{-1}$, ordered so as to represent the order in which the corresponding edges of the polygon are traversed (on travelling round the polygon in the anticlockwise direction), and where an edge represented by some letter x occurs in the surface symbol as 'x' if the chosen orientation on the edge agrees with the anticlockise orientation, or as 'x^{-1}' if the chosen orientation on the edge is opposite to the anticlockwise orientation. For example, the surface symbol $xyx^{-1}y^{-1}$ represents the torus (see Figure 3), and the surface symbol $xy^{-1}xy^{-1}$ represents the real projective plane (see Figure 4).

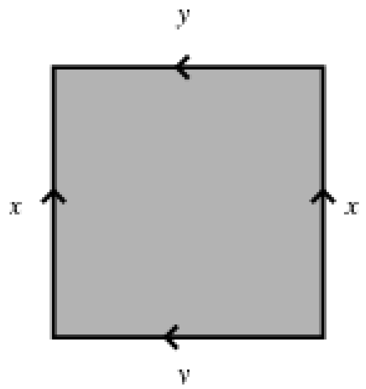

Figure 3: Torus

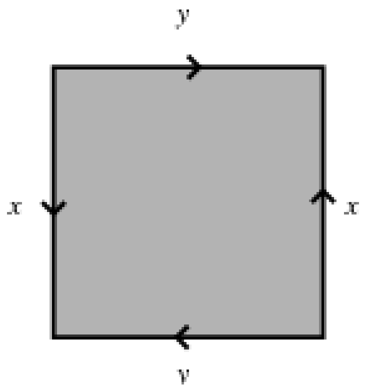

Figure 4: Real projective plane

Lemma 5.5 shows that any connected triangulated closed surface can be described by such a *surface symbol*. This is a finite sequence of symbols of the form $a, a^{-1}, b, b^{-1}, c, c^{-1}, \ldots,$ $a, b, c, \ldots$ representing some suitable list of 'letters' that label the pairs of identified edges of the polygon representing the surface. Each 'letter' x present occurs exactly twice in the surface symbol, either as 'x' or as 'x^{-1}'. Conversely any surface symbol of this form determines a scheme for identifying in pairs the edges of a suitable polygon to obtain a closed surface. We wish to determine necessary and sufficient conditions for determining whether or not the surfaces obtained in this way from two such surface symbols are homeomorphic.

Let us use capital letters $A, B, C, \ldots$ to denote (possibly empty) sequences of symbols taken from the list $a, a^{-1}, b, b^{-1}, c, c^{-1}, \ldots$. Also if A is such a sequence of symbols, given by $A = a_1 a_2 \cdots a_n$, then we write A^{-1} for the sequence $a_n^{-1} a_{n-1}^{-1} \cdots a_1^{-1}$ (where $(x^{-1})^{-1} \equiv x$ for any letter x.) Using these conventions, we now state three rules which enable one to transform one surface symbol into another in such a way that the two surface symbols represent surfaces that are homeomorphic.

- **Rule 1.** cyclically permute the symbols occurring in the surface symbol,
- **Rule 2.** Replace $ABxCDxE$ by $AyDB^{-1}yC^{-1}E$ (where y represents some letter not occurring in the first surface symbol).
- **Rule 3.** Replace $ABxCDx^{-1}E$ by $AyDCy^{-1}BE$.
- **Rule 4.** Replace $Axx^{-1}B$ or $Ax^{-1}xB$ by AB, provided that AB contains at least two letters (each occurring twice).

Lemma 5.6 *The application of Rules 1–4 to a surface symbol gives a new surface symbol such that the surfaces determined by the two surface symbols are homeomorphic.*

Proof Rule 1 corresponds to traversing the boundary of the polygon starting from a different vertex. Rules 2–4 are justified by the simple 'cut and paste' operations depicted in Figures 5, 6 and 7. ∎

Rules 1–4 allow the reduction of surface symbols to certain standard forms. Now each letter x in a surface symbol occurs exactly twice; if the letter x either occurs both times as 'x' or else occurs both times as 'x^{-1}', then we call the occurrence of the letter x in the surface symbol a *similar pair*; otherwise we call the occurrence of this letter a *reversed pair*. Two reversed pairs in some given surface symbol are said to *interlock* if they occur in the order $\cdots y \cdots z \cdots y^{-1} \cdots z^{-1} \cdots$ (after interchanging y and y^{-1}, or z and z^{-1}, if necessary).

Proposition 5.7 *Any surface symbol can be reduced, by suitable applications of the transformations described in Rules 1–4 above and their inverses, to one of the following canonical forms:—*

$$x_1 y_1 x_1^{-1} y_1^{-1} \cdots x_g y_g x_g^{-1} y_g^{-1} \qquad (g \geq 1),$$

$$x_1 x_1 x_2 x_2 \cdots x_h x_h \qquad (h \geq 2),$$

$$x x^{-1} y y^{-1}, \qquad x x y y^{-1}.$$

Proof First we note that if C is any sequence of the form $x_1 x_1 x_2 x_2 \ldots$, then a sequence of transformations

$$CDxExF \rightarrow CyD^{-1}yE^{-1}F \rightarrow CzzDE^{-1}F$$

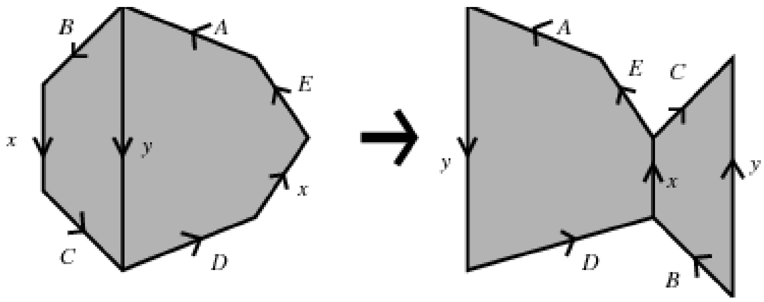

Figure 5: Rule 2

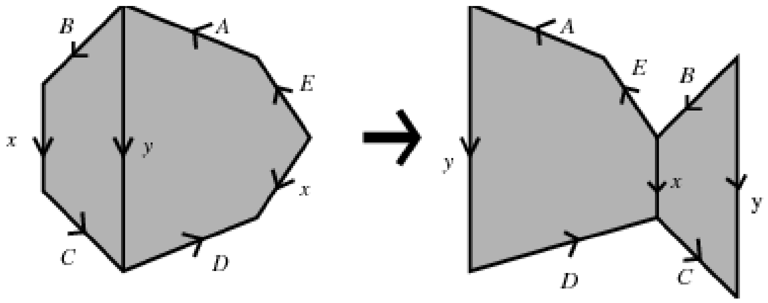

Figure 6: Rule 3

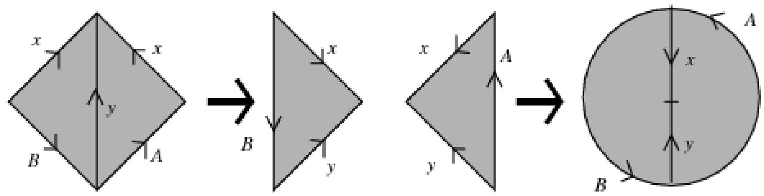

Figure 7: Rule 4

of the type specified by Rule 2 will transform any surface symbol $CDxExF$ with a similar pair to one of the form $CzzDE^{-1}$. Repeated applications of this procedure reduce any surface symbol to one of the form AB, where A is of the form form $x_1x_1x_2x_2\cdots x_rx_r$ and B contains only reversed pairs (where either A or B may be empty).

One can now use Rule 3 in order to reduce a surface symbol of the form AB to one of the form ACD, where C is of the form

$$y_1z_1y_1^{-1}z_1^{-1}\cdots y_sz_sy_s^{-1}z_s^{-1},$$

and D contains only non-interlocking reversed pairs. Indeed if E is any surface symbol of the required form, then successive applications of Rule 3 show that any surface symbol of the form $EFaGbHa^{-1}Ib^{-1}J$ can be transformed to $Eefe^{-1}f^{-1}FIHGJ$ by the following sequence of transformations:

$$\begin{aligned} EFaGbHa^{-1}Ib^{-1}J &\rightarrow EcGbHc^{-1}FIb^{-1}J \\ &\rightarrow EcGdFIHc^{-1}d^{-1}J \\ &\rightarrow EeFIHGde^{-1}d^{-1}J \\ &\rightarrow Eefe^{-1}f^{-1}FIHGJ. \end{aligned}$$

The stated reduction of AB to ACD now follows by induction on the number of interlocking reversed pairs.

If A is non-empty then one can reduce a surface symbol of the form ACD (where A, C and D are as above) to one of the form ED, where E is of the form $x_1x_1x_2x_2\cdots$ and D contains only non-interlocking reversed pairs. This follows from successive applications of the following sequence of transformations (which are inverses of transformations of the type specified by Rule 2):

$$Fxxaba^{-1}b^{-1}G \rightarrow Fyb^{-1}a^{-1}ya^{-1}b^{-1}G \rightarrow Fyay^{-1}accG \leftarrow FyyddccG.$$

Now consider D, which consists only of non-interlocking reversed pairs. Let $\cdots x\cdots x^{-1}\cdots$ be the closest reversed pair occurring in D. Then x and x^{-1} must be adjacent (since otherwise D would contain two interlocking reversed pairs). We can therefore 'cancel' xx^{-1}, by Rule 4, provided that the resultant symbol always contains at least two letters. It follows that any surface symbol with more than two letters can be reduced to one or other of the first two canonical forms specified (with $g \geq 1$ or $h \geq 2$).

Finally consider surface symbols with two letters occurring. The procedures described above reduce such a surface symbol to one of the following forms: $xyx^{-1}y^{-1}$, $xxyy$, $xxyy^{-1}$, $xx^{-1}yy^{-1}$, $xyy^{-1}x^{-1}$. The first four of these are included in the list of canonical forms. The symbol $xyy^{-1}x^{-1}$ reduces to $zz^{-1}yy^{-1}$ on cyclically permuting the symbol (according to Rule 1) and replacing x^{-1} and x by z and z^{-1} respectively, as required. ∎

Let M_g ($g \geq 1$) be the space obtained from a regular $4g$-sided polygon by identifying the edges according to the sequence $x_1 y_1 x_1^{-1} y_1^{-1} \cdots x_g y_g x_g^{-1} y_g^{-1}$ (see Figure 8), and let N_h ($h \geq 2$) be defined similarly using $x_1 x_1 \cdots x_h x_h$ (see Figure 9). Also let M_0 and N_1 be surfaces whose surface symbols are $xx^{-1}yy^{-1}$ and $xxyy^{-1}$ respectively. We have so far proved that the polyhedron of any connected triangulated closed surface is homeomorphic to one of the spaces M_g ($g \geq 0$) or N_h ($h \geq 1$).

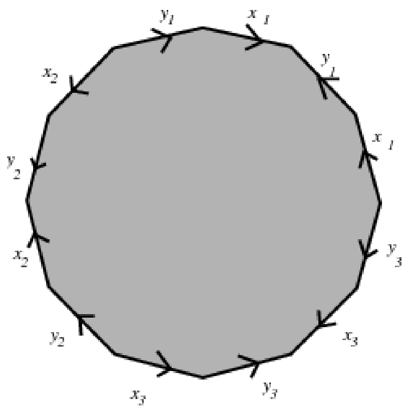

Figure 8: The surface M_g ($g = 3$)

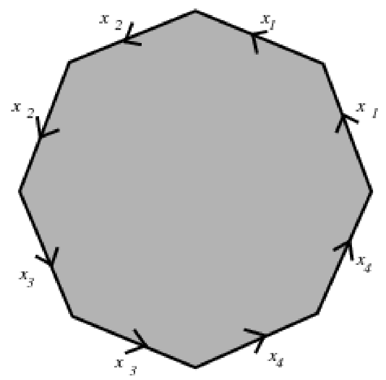

Figure 9: The surface N_h ($h = 4$)