AN INVITATION TO TOPOLOGY

Lecture notes by Răzvan Gelca

Contents

Part I

General Topology

Chapter 1

Topological Spaces and Continuous Functions

Topology studies properties that are invariant under continuous transformations (homeomorphisms). As such, it can be thought of as rubber-sheet geometry. It is interested in how things are connected, but not in shape and size. The fundamental objects of topology are topological spaces and continuous functions.

1.1 The topology of the real line

The Weierstrass $\epsilon - \delta$ definition for the continuity of a function on the real axis

Definition. A function $f : \mathbb{R} \to \mathbb{R}$ is continuous if and only if for every $x_0 \in \mathbb{R}$ and every $\epsilon > 0$ there is $\delta > 0$ such that for all $x \in \mathbb{R}$ with $|x - x_0| < \delta$, one has $|f(x) - f(x_0)| < \epsilon$.

can be rephrased by the more elegant

Definition. A function $f : \mathbb{R} \to \mathbb{R}$ is continuous if and only if the preimage of each open interval is a union of open intervals

or even by the most elegant

Definition. A functions $f : \mathbb{R} \to \mathbb{R}$ is *continuous* if and only if the preimage of each union of open intervals is a union of open intervals.

For simplicity, a union of open intervals will be called an *open set*. And because the complement of an open interval consists of one or two closed intervals, we will call the complements of open sets *closed sets*. Our *topological space* is $\mathbb{R}$, and the *topology* on $\mathbb{R}$ is defined by the open sets.

Let us examine the properties of open sets. First, notice that the union of an arbitrary family of open sets is open. This is not true for the intersection though, since for example the intersection of all open sets centered at 0 is just $\{0\}$. However the intersection of finitely many open sets is open, provided that the sets intersect nontrivially. Add the empty set to the topology so that the intersection of finitely many open sets is always open. Notice also that $\mathbb{R}$ is open since it is the union of all its open subintervals.

Open intervals are the building blocks of the topology. For that reason, they are said to form a *basis*. If we just restrict ourselves to bounded open intervals, they form a basis as well. Each bounded open interval is of the form $(x_0 - \delta, x_0 + \delta)$, and as such it consists of all points that are at distance less than δ from x_0. So the distance function (metric) on $\mathbb{R}$ can be used for defining a topology.

1.2 The definitions of topological spaces and continuous maps

We will define the notions of topological space and continuous maps to cover $\mathbb{R}^n$ with continuous functions on it (real analysis), spaces of functions with continuous functionals on them (functional analysis, differential equations, mathematical physics), manifolds with continuous maps, algebraic sets (zeros of polynomials) and regular (polynomial) maps (algebraic geometry).

Definition. A *topology* on a set X is a collection $\mathcal{T}$ of subsets of X with the following properties

(1) $\emptyset$ and X are in $\mathcal{T}$,

(2) The union of arbitrarily many sets from $\mathcal{T}$ is in $\mathcal{T}$,

(3) The intersection of finitely many sets from $\mathcal{T}$ is in $\mathcal{T}$.

The sets in $\mathcal{T}$ are called *open*, their complements are called *closed*. Let us point out that closed sets have the following properties: (1) X and $\emptyset$ are closed, (2) the union of finitely many closed sets is closed, (3) the intersection of an arbitrary number of closed sets is closed.

Example 1. On $\mathbb{R}^n$ we define the open sets to consist of the whole space, the empty set and the unions of open balls $B_{x_0,\delta} = \{x \in \mathbb{R}^n \,|\, \mathrm{dist}(x, x_0) < \delta\}$. This is the standard topology on $\mathbb{R}^n$.

Example 2. Let $C[a, b]$ be the set of real-valued continuous functions on the interval $[a, b]$ endowed with the distance function $\mathrm{dist}(f, g) = \sup_x |f(x) - g(x)|$. Then $C[a, b]$ is a topological space with the open sets being the unions of "open balls" of the form $B_{f,\delta} = \{g \in C[a, b] \,|\; \mathrm{dist}(f, g) < \delta\}$.

Example 3. The Lebesgue space $L^2(\mathbb{R})$ of integrable functions f on $\mathbb{R}$ such that $\int |f(x)|^2 dx < \infty$, with open sets being the unions of "open balls" of the form $B_{f,\delta} = \{g \in L^2(\mathbb{R}) \,|\; \int |f(x)-g(x)|^2 dx < \delta\}$.

Example 4. In $\mathbb{C}^n$, let the closed sets be intersections of zeros of polynomials. That is, closed sets are of the form

$$V = \{z \in \mathbb{C}^n \,|\, f(z) = 0 \text{ for } f \in S\}$$

where S is a set of n-variable polynomials. The open sets are their complements. This is called the *Zariski topology.*

A particular case is that of $n = 1$. In that case every polynomial has finitely many zeros (maybe no zeros at all for constant polynomials), except for the zero polynomial whose zeros are the entire complex plane. Moreover, any finite set is the set of zeros of some polynomial. So the closed sets are the finite sets together with $\mathbb{C}$ and $\emptyset$. The open sets are $\mathbb{C}$, $\emptyset$, and the complements of finite sets.

Example 5. Inspired by the Zariski topology on $\mathbb{C}$, given an arbitrary infinite set X we can let $\mathcal{T}_c$ be the collection of all subsets U of X such that $X \backslash U$ is either countable or it is all of X.

Example 6. We can cook up examples of exotic topologies, such as $X = \{1, 2, 3, 4\}$, $\mathcal{T} = \{\emptyset, X, \{1\}, \{2, 3\}, \{1, 2, 3\}, \{2, 3, 4\}\}$.

Example 7. There are two silly examples of topologies of a set X. One is the *discrete* topology, in which every subset of X is open and the other is the *trivial* topology, whose only open sets are $\emptyset$ and X.

Example 8. Here is a fascinating topological proof given in 1955 by H. Fürstenberg to Euclid's theorem.

Theorem 1.2.1. (Euclid) There are infinitely many prime numbers.

Proof. Introduce a topology on $\mathbb{Z}$, namely the smallest topology in which any set consisting of all terms of a nonconstant arithmetic progression is open. As an example, in this topology both the set of odd integers and the set of even integers are open. Because the intersection of two arithmetic progressions is an arithmetic progression, the open sets of $\mathcal{T}$ are precisely the unions of arithmetic progressions. In particular, any open set is either infinite or void.

If we denote

$$A_{a,d} = \{\ldots, a-2d, a-d, a, a+d, a+2d, \ldots\}, \quad a \in \mathbb{Z},\ d > 0,$$

then $A_{a,d}$ is open by hypothesis, but it is also closed because it is the complement of the open set $A_{a+1,d} \cup A_{a+2,d} \cup \ldots \cup A_{a+d-1,d}$. Hence $\mathbb{Z}\backslash A_{a,d}$ is open.

Now let us assume that only finitely many primes exist, say $p_1, p_2, \ldots, p_n$. Then

$$A_{0,p_1} \cup A_{0,p_2} \cup \ldots \cup A_{0,p_n} = \mathbb{Z}\backslash\{-1, 1\}.$$

This union is the complement of the open set

$$(\mathbb{Z}\backslash A_{0,p_1}) \cap (\mathbb{Z}\backslash A_{0,p_2}) \cap \cdots \cap (\mathbb{Z}\backslash A_{0,p_n}),$$

hence it is closed. The complement of this closed set, which is the set $\{-1, 1\}$, must therefore be open. We reached a contradiction because this set is neither empty nor infinite. Hence our assumption was false, and so there are infinitely many primes. $\square$

Given two topologies $\mathcal{T}$ and $\mathcal{T}'$ such that $\mathcal{T}' \subset \mathcal{T}$, one says that $\mathcal{T}$ is *finer* than $\mathcal{T}'$, or that $\mathcal{T}'$ is *coarser* then $\mathcal{T}$.

Definition. Given a point x, if a set V contains an open set U such that $x \in U$ then V is called a *neighborhood* of x.

Let X and Y be topological spaces.

Definition. A map $f : X \to Y$ is continuous if for every open set $U \in Y$, the set $f^{-1}(U)$ is open is X.

Example 1. This definition covers the case of continuous maps $f : \mathbb{R}^m \to \mathbb{R}^n$ encountered in multivariable calculus.

Example 2. Let $X = C[a, b]$, the topological space of continuous functions from Example 2 above, and let $Y = \mathbb{R}$. The functional $\phi : C[a, b] \to \mathbb{R}$, $\phi(f) = \int_a^b f(x)dx$ is continuous.

Example 3. Let $X = L^p(\mathbb{R})$, $Y = \mathbb{R}$ and $\phi : X \to Y$, $\phi(f) = (\int_p |f(x)|^p dx)^{1/p}$.

Remark 1.2.1. An alternative way of phrasing the defintion is to say that for every neighborhood W of $f(x)$ there is a neighborhood V of x such that $f(V) \subset W$.

Proposition 1.2.1. The composition of continuous maps is continuous.

Proof. Let $f : X \to Y$ and $g : Y \to Z$ be continuous, and let us show that $g \circ f$ is continuous. If $U \subset$ is open, then $g^{-1}(U)$ is open, so $f^{-1}(g^{-1}(U))$ is open. Done. $\square$

Definition. If $f : X \to Y$ is a one-to-one and onto map between topological spaces such that both f and f^{-1} are continuous, then f is called a homeomorphism.

If there is a homeomorphism between the topological spaces X and Y then from the topological point of view they are indistinguishable.

1.3 Procedures for constructing topological spaces

1.3.1 Basis for a topology

Rather than specifying all open sets, we can exhibit a family of open sets from which all others can be recovered. In general, basis elements mimic the role of open intervals in the topology of the real line.

Definition. Given a set X, a *basis* for a topology on X is a collection $\mathcal{B}$ of subsets of X such that

(1) For each $x \in X$, there is at least one basis element B containing x,

(2) If $x \in B_1 \cap B_2$ with B_1, B_2 basis elements, then there is a basis element B_3 such that $x \in B_3 \subset B_1 \cap B_2$.

Proposition 1.3.1. Let $\mathcal{T}$ be the collection of all subsets U of X with the property that for every $x \in U$, there is $B_x \in \mathcal{B}$ such that $x \in B_x \subset U$. Then $\mathcal{T}$ is a topology.

Proof. (1) X and $\emptyset$ are in $\mathcal{T}$ trivially.

(2) If $U_\alpha \in \mathcal{T}$ for all α, let us show that $U = \cup_\alpha U_\alpha \in \mathcal{T}$. Given $x \in U$, there is U_α such that $x \in U_\alpha$. By hypothesis there is $B_\alpha \in \mathcal{B}$ such that $x \in B_\alpha \subset U_\alpha$, and hence $x \in B_\alpha \subset U$.

(3) Let us show that the intersection of two elements U_1 and U_2 from $\mathcal{T}$ is in $\mathcal{T}$. For $x \in U_1 \cup U_2$ there are basis elements B_1, B_2 such that $x \in B_i \subset U_i$, $i = 1, 2$. Then there is a basis element B_3 such that $x \in B_3 \subset B_1 \cap B_2 \subset U_1 \cap U_2$, and so $U_1 \cap U_2 \in \mathcal{T}$. The general case of the intersection of n sets follows by induction. □

Now suppose that we are already given the topology $\mathcal{T}$. How do we recognize if a basis $\mathcal{B}$ is indeed a basis for this topology.

Proposition 1.3.2. Let X be a topological space with topology $\mathcal{T}$. Then a family of subsets of X, $\mathcal{B}$, is a basis for $\mathcal{T}$, if and only if

(1) Every element of $\mathcal{B}$ is open and if $U \in \mathcal{T}$ and $x \in U$, then there is $B \in \mathcal{B}$ with $x \in B \subset U$.

(2) If $x \in B_1 \cap B_2$ with B_1, B_2 basis elements, then there is a basis element B_3 such that $x \in B_3 \subset B_1 \cap B_2$.

In this case $\mathcal{T}$ equals the collection of all unions of elements in $\mathcal{B}$.

Proof. Condition (2) is required by the definition of basis. Also the fact that $\mathcal{T}$ consists of unions of elements of $\mathcal{B}$ implies that $\mathcal{B}$ consists of open sets. Finally, since by definition, the elements of $\mathcal{T}$ are unions of elements in $\mathcal{B}$, we get (1). □

Example 1. The collection of all disks in the plane is a basis for the standard topology of the plane.

Example 2. The collection of all rectangular regions in the plane that have sides parallel to the axes of coordinates is a basis for the standard topology.

Example 3. The basis consisting of all intervals of the form $(a, b]$ with $a < b$ and $a, b \in \mathbb{R}$ generates a topology called the upper limit topology. This topology is different from the standard topology, since for example $(a, b]$ is not open in the standard topology. Since $(a, b) = \cup_n (a, b - 1/n]$, we see that the standard topology is coarser than the upper limit topology.

Similarly, the sets $[a, b)$ with $a < b$ and $a, b \in \mathbb{R}$ form a basis for the lower limit topology.

Taking into account both unions and finite intersections, one can simplify further the generating family for a topology. A *subbasis* $\mathcal{S}$ for a topology on X is a collection of subsets of X whose union equals X.

Proposition 1.3.3. The set $\mathcal{T}$ consisting of all unions of finite intersections of elements of $\mathcal{S}$ and the empty set is a topology on X.

Proof. (1) $\emptyset, X \in \mathcal{T}$ by hypothesis.

(2) The union of unions of finite intersections of elements in $\mathcal{S}$ is a union of finite intersections of elements in $\mathcal{S}$.

(3) It suffices to show that the set $\mathcal{B}$ of all finite intersections of elements in S is a basis for a topology. And indeed, if $B_1 = S_1 \cap S_2 \cap \cdots \cap S_m$ and $B_2 = S_1' \cap S_2' \cap \cdots \cap S_n'$, then $B_1 \cap B_2 = S_1 \cap S_2 \cap \cdots \cap S_1' \cap S_2' \cap \cdot \cap S_n'$ which is again in $\mathcal{B}$. Done. $\square$

Here is a criterion that allows us to recognize at first glance bases for topologies.

Proposition 1.3.4. Let X be a topological space with topology $\mathcal{T}$. Suppose that $\mathcal{C}$ is a collection of open sets of X such that for each open set $U \subset X$ and each $x \in U$, there is $C \in \mathcal{C}$ such that $x \in C \subset U$. Then $\mathcal{C}$ is a basis for $\mathcal{T}$.

Proof. First, we show that $\mathcal{C}$ is a basis. Since for every $x \in X$, there is $C \in \mathcal{C}$ such that $x \in C \subset X$, it follows that X is the union of the elements of $\mathcal{C}$. For the second condition, let $C_1, C_2 \in \mathcal{C}$, and $x \in C_1 \cap C_2$. Since $C_1 \cap C_2$ is open (both C_1 and C_2 are), there is $C_3 \in \mathcal{C}$ such that $x \in C_3 \subset C_1 \cap C_2$.

Let us show now that $\mathcal{C}$ is a basis for the topology $\mathcal{T}$. First, given $U \in \mathcal{T}$, for each $x \in U$, there is $C_x \in \mathcal{C}$ such that $x \in C_x \subset U$. Then $U = \cup_{x \in U} C_x$. Thus all open sets belong to the topology generated by $\mathcal{C}$. On the other hand, every union of elements of $\mathcal{C}$ is a union of open sets in $\mathcal{T}$, thus is in $\mathcal{T}$. Hence the conclusion. $\square$

Working with a basis simplifies the task of comparing topologies.

Proposition 1.3.5. Let $\mathcal{B}$ and $\mathcal{B}'$ be bases for the topologies $\mathcal{T}$ respectively $\mathcal{T}'$ on X. Then $\mathcal{T}'$ is finer than $\mathcal{T}$ if and only if for each $x \in X$ and each $B \in \mathcal{B}$ that contains x, there is $B' \in \mathcal{B}'$ such that $x \in B' \subset B$.

Proof. If $\mathcal{T}'$ is finer than $\mathcal{T}$, then every $B \in \mathcal{B}$ is in $\mathcal{T}'$. Hence for every $x \in B$, there is $B' \in \mathcal{B}'$ such that $x \in B' \subset B$.

For the converse, let us show that every $U \in \mathcal{T}$ is also in $\mathcal{T}'$. For every $x \in U$, there is $B_x \in \mathcal{B}$ such that $x \in B_x \subset U$, and hence there is $B_x' \in \mathcal{B}'$ such that $x \in B_x' \subset B_x \subset U$. Then $U = \cup_{x \in U} B_x'$, showing that $U \in \mathcal{T}'$. $\square$

Example. The collection of all disks in the plane and the collection of all squares in the plane generate the same topology. Indeed, for every disk, and every point in the disk there is a square centered at that point included in the disk, and for every square and every point in the square there is a disk centered at the point included in the square.

Using a basis makes it easier to check continuity.

Proposition 1.3.6. Let X and Y be topological spaces. Than $f : X \to Y$ is continuous if and only if for every basis element of the topology on Y, $f^{-1}(B)$ is open in X.

1.3.2 Subspaces of a topological space

One studies continuous functions on subsets of the real axis, as well, such as continuous functions on open and closed intervals. Continuity is then rephrased by restricting open intervals to the domain of the function, that is by intersecting open sets with the domain.

Definition. Let X be a topological space with topology $\mathcal{T}$. If Y is a subset of X, then Y itself is a topological space with the *subspace topology*

$$\mathcal{T}_Y = \{Y \cap U \,|\, U \in \mathcal{T}\}.$$

Proposition 1.3.7. The set $\mathcal{T}_Y$ is a topology on Y. If $\mathcal{B}$ is a basis of $\mathcal{T}$, then

$$\mathcal{B}_Y = \{B \cap Y \,|\, B \in \mathcal{B}\}$$

is a basis for $\mathcal{T}_Y$.

Proof. (1) $Y = X \cap Y$ and $\emptyset = \emptyset \cap Y$ are in $\mathcal{T}_Y$.
(2) and (3) follow from

$$(U_1 \cap Y) \cap \cdots \cap (U_n \cap Y) = (U_1 \cap U_2 \cdots \cap U_n) \cap Y$$
$$\cup_\alpha (U_\alpha \cap Y) = (\cup_\alpha U_\alpha) \cap Y.$$

For the second part, let U be open in X and $y \in U \cap Y$. Choose $B \in \mathcal{B}$ such that $y \in B \subset U$. Then $y \in B \cap Y \subset U \cap Y$, and the conclusion follows. □

Example 1. For $[0,1] \subset \mathbb{R}$, then a basis for the subspace topology consists of all the sets of the form (a,b), $[0,b)$, $(a,1]$ with $a,b \in (0,1)$.

Example 2. For $\mathbb{Z} \subset \mathbb{R}$, then the subset topology is the discrete topology.

Example 3. For $(0,1) \cup \{2\}$, then the open sets of the subset topology are all sets of either the form U or $U \cup \{2\}$, where U is a union of open intervals in $(0,1)$.
Example 4. the n-dimensional sphere

$$S^n = \{(x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1} \,|\, x_0^2 + x_1^2 + \cdots + x_n^2 = 1\}$$

is a subspace of $\mathbb{R}^{n+1}$.

Proposition 1.3.8. If $f : X \to Z$ is a continuous map between topological spaces and if $Y \subset X$ is a topological subspace, then the restriction $f|_Y : Y \to Z$ is a continuous map.

Proof. Let $U \subset Z$ be open. Then $f^{-1}(U)$ is open in X. But $f|_Y^{-1}(U) = f^{-1}(U) \cap Y$, which is open in Y. Hence f is continuous. □

1.3.3 The product of two topological spaces

By examining how the standard topology on $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ compares to the one on $\mathbb{R}$, we can make the following generalization

Definition. Let X and Y be topological spaces. The *product topology* on $X \times Y$ is the topology having as basis the collection $\mathcal{B}$ of all the sets of the form $U \times V$, where U is an open set of X and V is an open set of Y.

Of course, for this to work we need the following

Proposition 1.3.9. The collection $\mathcal{B}$ defined this way is a basis.

Proof. The first condition for the basis just states that $X \times Y$ is in $\mathcal{B}$, which is obvious. For the second condition, note that if $U_1 \times V_1$ and $U_2 \times V_2$ are basis elements, then

$$(U_1 \times V_1) \cap (U_2 \times V_2) = (U_1 \cap U_2) \times (V_1 \cap V_2),$$

and the latter is a basis element because $U_1 \cap U_2$ and $V_1 \cap V_2$ are open. □

Proposition 1.3.10. If $\mathcal{B}_X$ is a basis for the topology on X and $\mathcal{B}_Y$ is a basis for the topology on Y, then

$$\mathcal{B} = \{B_1 \times B_2 \mid B_1 \in \mathcal{B}_X, B_2 \in \mathcal{B}_Y\}$$

is a basis for the topology of $X \times Y$.

Proof. We will apply the criterion from Proposition 1.3.4. Given an open set $W \subset X \times Y$ and $(x, y) \in W$, by the definition of the product topology there is a basis element of the form $U \times V$ such that $(x, y) \in U \times V \subset W$. Then, there are $B_1 \in \mathcal{B}_X$ such that $x \in B_1 \subset U$ and $B_2 \in \mathcal{B}_Y$, such that $y \in B_2 \subset V$. Then $(x, y) \subset B_1 \times B_2 \subset U \times V$. It follows that $\mathcal{B}$ meets the requirements of the criterion, so $\mathcal{B}$ is a basis for $X \times Y$. □

Using an inductive construction we can extend the definition of product topology to a cartesian product of finitely many topological spaces.

1.3.4 The product of an arbitrary number of topological spaces

There are two ways in which the definition of product topology can be extended to an infinite product of topological spaces, the box topology and what we will call the product topology. Let X_α, $\alpha \in A$ be a family of topological spaces.

Definition. The *box topology* is the topology on $\prod_\alpha X_\alpha$ with basis all sets of the form $\prod_\alpha U_\alpha$ with U_α open in X_α, for all $\alpha \in A$.

Definition. The *product topology* is the topology on $\prod_\alpha X_\alpha$ with basis all sets of the form $\prod_\alpha U_\alpha$, with U_α open in X_α and $U_\alpha = X_\alpha$ for all but finitely many $\alpha \in A$.

Notice that the second topology is coarser than the first. At first glance, the box topology seems to be the right choice, but unfortunately it is to fine to be of any use in applications. In the case of normed spaces, the second topology becomes the weak topology, which is quite useful (e.g. in the theory of differential equations). In fact, the next result is a good reason for picking this as the right topology on the product space.

Proposition 1.3.11. Let X_α, $\alpha \in A$ and Y be topological spaces. Then $f : Y \to \prod_\alpha X_\alpha$ is continuous if and only if the coordinate functions $f_\alpha : Y \to X_\alpha$ are all continuous.

Proof. Assume first that for each α, f_α is continuous. Let B be a basis element for the topology of X, say $B = \prod_{\alpha \in A_0} U_\alpha \times \prod_{\alpha \notin A_0} X_\alpha$, where A_0 is finite and U_α are open. Then $f^{-1}(B) = \cap_\alpha f_\alpha^{-1}(U_\alpha)$. But there are finitely many U_α's! It follows that

$$f^{-1}(B) = \cap_\alpha f_\alpha^{-1}(U_\alpha)$$

which is open, being an intersection of finitely many open sets.

For the converse, notice that the projection maps $\pi_\alpha : \prod_\alpha X_\alpha$ are continuous because of the way the topology was defined, and that $f_\alpha = \pi_\alpha \circ f$. By Proposition 1.2.1, f_α is continuous. QED. □

Lemma 1.3.1. The addition, subtraction, and multiplication operations are continuous functions from $\mathbb{R} \times \mathbb{R}$ into $\mathbb{R}$; and the quotient operation is continuous from $\mathbb{R} \times (\mathbb{R}\backslash\{0\})$ into $\mathbb{R}$.

Proposition 1.3.12. If X is a topological space and $f, g : X \to \mathbb{R}$ are continuous functions, then $f + g$, $f - g$ and $f \cdot g$ are continuous. If $g(x) \neq 0$ for all x, then f/g is continuous.

Proof. Let $\mu : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be one of the (continuous) operations from Lemma 1.3.1. The function $\phi : X \to \mathbb{R} \times \mathbb{R}$, $\phi(x) = (f(x), g(x))$ is continuous by Proposition 1.3.11. The conclusion follows by taking the composition $\mu \circ \phi$. □

1.3.5 The disjoint union of two topological spaces.

Definition. Given a family X_α of topological spaces, $\alpha \in A$, the topological space $\amalg_\alpha X_\alpha$ is the disjoint union of the spaces X_α endowed with the topology in which U is open if and only if $U \cap X_\alpha$ is open for all α.

Example 1. If X is any topological space, then $\amalg_{x \in X}\{x\}$ equals X as a set, but it is now endowed with the discrete topology.

Proposition 1.3.13. If X_α, $\alpha \in A$, Y are topological spaces then $f : \amalg_\alpha X_\alpha \to Y$ is continuous if and only if $f|X_\alpha$ is continuous for each α.

1.3.6 Metric spaces as topological spaces

Metric spaces are examples of topological spaces that are widely used in areas such as geometry, real analysis, or functional analysis.

Definition. A *metric* (distance) on a set X is a function

$$d : X \times X \to \mathbb{R}$$

satisfying the following properties

(1) $d(x, y) \geq 0$ for all $x, y \in X$, with equality if and only if $x = y$.

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

(3) $d(x, y) + d(y, z) \geq d(x, z)$ for all $x, y, z \in X$.

For $\epsilon > 0$, set

$$B(x, \epsilon) = \{y \,|\, d(x, y) < \epsilon\}.$$

This is called the *ϵ-ball centered at x.*

Proposition 1.3.14. If d is a metric on a set X, then the collection of all balls $B(x, \epsilon)$ for $x \in X$ and $\epsilon > 0$ is a basis for a topology on X.

Proof. The first condition for a basis is trivial, since each point lies in a ball centered at that point. For the second condition, let $B(x_1, \epsilon_1)$ and $B(x_2, \epsilon_2)$ be balls that intersect, and let x be a point in their intersection. Choose

$$\epsilon < \min(\epsilon_1 - d(x, x_1), \epsilon_2 - d(x, x_2)).$$

Then the triangle inequality implies that if $y \in B(x, \epsilon)$, then

$$d(y, x_i) < d(y, x) + d(x, x_i) < \epsilon_i - d(x, x_i) + d(x, x_i) < \epsilon_i, \quad i = 1, 2.$$

Hence y lies in both balls. This shows that $B(x, \epsilon) \subset B(x_1, \epsilon_1) \cap B(x_2, \epsilon_2)$, and the condition is satisfied. □

Definition. The topology with basis all balls in X is called the *metric topology.*

Remark 1.3.1. Every open set U is of the form $\cup_{x \in U} B(x, \epsilon_x)$.

Example 1. If X is a metric space with distance function d and $A \subset X$, then A is a metric space with the same distance.

Example 2. The standard topology of $\mathbb{R}^n$ induced by the Euclidean metric.

Example 3. Given a set X, define

$$d(x, y) = 1, \quad \text{if } x \neq y$$
$$d(x, y) = 0, \quad \text{if } x = y.$$

Then d is a metric which induces the discrete topology.

Example 4. On $\mathbb{R}^n$ define the metric

$$\rho(\mathbf{x}, \mathbf{y}) = \max(|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_n - y_n|)$$

Then this is a metric that induces the standard topology on $\mathbb{R}^n$.

The fact that ρ is a metric is easy to check. Just the triangle inequality poses some difficulty, and here is the proof:

$$|x_i - z_i| \leq |x_i - y_i| + |y_i - z_i|, \quad \text{for all } i.$$

Thus

$$|x_i - z_i| \leq \rho(\mathbf{x}, \mathbf{y}) + \rho(\mathbf{y}, \mathbf{z}).$$

Taking the maximum over all i on the left yields the triangle inequality.

The fact that the metric ρ defined above induces the same metric is a corollary of the following result.

Lemma 1.3.2. Let d and d' be two metrics on X inducing the topologies $\mathcal{T}$ respectively $\mathcal{T}'$. Then $\mathcal{T}'$ is finer than $\mathcal{T}$ if and only if for each $x \in X$ and each $\epsilon > 0$ there is $\delta > 0$ such that

$$B_{d'}(x, \delta) \subset B_d(x, \epsilon).$$

Proof. Indeed, if $\mathcal{T}'$ is finer than $\mathcal{T}$, then any ball in $\mathcal{T}$ is the union of balls in $\mathcal{T}'$, and, by eventually shrinking the radius, we can make sure that such a ball is centered at any desired point.

Conversely, suppose the $\epsilon - \delta$ condition holds. Let U be open in $\mathcal{T}$ and $x \in U$. Choose $B_d(x, \epsilon) \subset U$. Then there is $B_{d'}(x, \delta) \subset B_d(x, \epsilon) \subset U$. This shows that U is open in $\mathcal{T}'$, as desired. □

Example 5. Let A be an index set and consider $X = \prod_{a \in A} \mathbb{R}$. Define the metric

$$\rho(\mathbf{x}, \mathbf{y}) = \sup_{\alpha \in A}(|x_\alpha, y_\alpha|).$$

This is called the uniform metric on X. Note that X is in fact the set of all functions on A. If $A = [a, b]$, then $C[a, b]$, the space of all continuous functions on $[a, b]$, is a subset of the set of all functions, hence it is a metric space with the uniform metric.

Definition. Let X be a metric space with metric d. A subset A of X is said to be *bounded* if there is some $x \in X$ and $M > 0$ such that $A \subset B(x, M)$.

An equivalent way of saying this is that the distances between points in A are bounded.

Proposition 1.3.15. Let X be a metric space with metric d. Define $\bar{d} : X \times X \to \mathbb{R}$ by

$$\bar{d}(x, y) = \min(d(x, y), 1).$$

Then $\bar{d}$ is a metric that induces the same topology as d.

Proof. The first two conditions for a metric are trivially satisfied. For the triangle inequality,

$$\bar{d}(x, z) \leq \bar{d}(x, y) + \bar{d}(y, z),$$

note that if any of the distances on the right are 1 the inequality is obvious since $\bar{d}(x, y) \leq 1$. If all three distances are less than 1, then the inequality follows from that for d. If only the distance on the left is 1, then we have

$$\bar{d}(x, z) \leq d(x, z) \leq d(x, y) + d(y, z) = \bar{d}(x, y) + \bar{d}(y, z).$$

To show that the two metrics generate the same topology, note that open sets can be defined using only small balls, namely balls of radius less than 1. □

Theorem 1.3.1. Let X and Y be metric spaces with metrics d_X and d_Y. Then $f : X \to Y$ is continuous if and only if for every $x_0 \in X$ and every $\epsilon > 0$ there is $\delta > 0$ such that $d_X(x_0, x) < \delta$ implies $d_Y(f(x_0), f(x)) < \epsilon$.

Proof. An open set in Y is a unions of balls $B(y, \epsilon)$ over all $y \in Y$. The condition from the statement is equivalent to the fact that the preimage of any open set is a union of balls in X, which is the same as saying that the preimage of any open set is open. □

For metric spaces there is a stronger notion of continuity.

Definition. Given the metric spaces X and Y, a function $f : X \to Y$ is *uniformly continuous* if for every $\epsilon > 0$ there is $\delta > 0$ such that if $x_1, x_2 \in X$ with $d_X(x_1, x_2) < \delta$ then $d_Y(f(x_1), f(x_2)) < \epsilon$.

1.3.7 Quotient spaces

Definition. Let X be a topological space and $p : X \to Y$ be a surjective map. The *quotient topology* on Y is defined by the condition that U in Y is open if and only if $p^{-1}(U)$ is open in X.

Proposition 1.3.16. The above definition gives rise to a topology on Y.

Proof. (1) $\emptyset$ and Y are clearly open.

(2) If U_α are open sets in Y, then

$$p^{-1}(\cup U_\alpha) = \cup p^{-1}(U_\alpha)$$

which is open in X.

(3) If $U_1, U_2, \ldots, U_n$ are open in Y then

$$p^{-1}(U_1 \cap U_2 \cap \ldots \cap U_n) = p^{-1}(U_1) \cap p^{-1}(U_2) \cap \ldots \cap p^{-1}(U_n)$$

which is open in X. □

Definition. Let X be a topological space, and let X^* be a partition of X into disjoint subsets whose union is X. Let $p : X \to X^*$ be the surjective map that carries each of the points of X to the element of X^* containing it. In the quotient topology induced by p, the space X^* is called the *quotient space* of X.

Example 1. The circle.

Let $f : \mathbb{R} \to \mathbb{C}$, $f(x) = \exp(2\pi i x)$. The image of f is the circle

$$S^1 = \{z \in \mathbb{C} \,|\, |z| = 1\}.$$

The quotient topology makes S^1 into a topological space.

Example 2. The 2-dimensional torus.

Consider the square $[0,1] \times [0,1]$ with the subspace topology, and define on it the equivalence relation

$$(x_1, 0) \sim (x_1, 1)$$
$$(0, x_2) \sim (1, x_2).$$

The quotient space is the 2-dimensional torus. This space is homeomorphic to $S^1 \times S^1$.

Example 3. The 2-dimensional projective plane.

In projective geometry there is a viewpoint O in the space and all planes not passing through O are identified by the rays that pass through O. In coordinates,

$$\mathbb{R}P^2 = (\mathbb{R}^3 \backslash \{0\}) / \sim$$

where $\mathbf{x} \sim \mathbf{y}$ if there is $\lambda \neq 0$ such that $\mathbf{x} = \lambda \mathbf{y}$.

Equivalently, $\mathbb{R}P^2$ is the quotient of the sphere

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 + y^2 + z^2 = 1\}$$

obtained by identifying antipodes ($(x, y, z) \sim (-x, -y, -z)$). Even simpler, it is the quotient of the upper hemisphere

$$S^2_+ = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 + y^2 + z^2 = 1, z \geq 0\}$$

obtained by identifying diametrically opposite points on the circle $z = 0$ (this circle is the line at infinity).

Example 4. On $[0,1] \cup [2,3]$ introduce the equivalence relation $0 \sim 1 \sim 2 \sim 3$. The quotient space is the figure eight.

Example 5. Let X be a topological space. The suspension ΣX is defined as the quotient $X \times [-1,1]/\sim$, where the equivalence relation is the following

- for $\lambda \neq -1, 1$, $(x, \lambda) \sim (y, \mu)$ if and only if $x = y$, $\lambda = \mu$;
- $(x, 1) \sim (y, 1)$ for all x, y;
- $(x, -1) \sim (y, -1)$ for all x, y.

1.3.8 Manifolds

The first three examples from the previous section are particular cases of manifolds. Manifolds are a special type of quotient spaces, obtained by patching together open sets in $\mathbb{R}^n$, for some positive integer n.

Definition. A topological space M is an n-dimensional real manifold if there is a family of subsets U_α, $\alpha \in A$, of $\mathbb{R}^n$ and a quotient map $f : \amalg_\alpha U_\alpha \to M$ such that $f|U_\alpha$ is a homeomorphism onto the image for all α.

The n-dimensional manifolds over complex numbers are defined in the same way by replacing $\mathbb{R}^n$ by $\mathbb{C}^n$. It is customary to denote the maps $f|U_\alpha$ by f_α. The maps $f_\alpha : U_\alpha \to M$ are called coordinate charts. By requiring the maps $f_\beta^{-1} \circ f_\alpha$ (where they are defined) to be smooth or analytical, one obtains the notions of smooth manifolds or of analytical manifolds. If the maps are complex analytical (i.e. holomorphic) then the manifold is called complex.

Example 1. The circle.

Let $U_1 = (0, 2\pi)$, $U_2 = (-\pi, \pi)$, $U_1, U_2 \subset \mathbb{R}$. The quotient map

$$f : U_1 \amalg U_2 \to S^1,$$

$f(x) = \exp(ix)$ determines a 1-dimensional real manifold structure on S^1.

Example 2. The 2-dimensional torus.

Consider the family of $(a, a+1) \times (b, b+1)$, $a, b \in \frac{1}{2}\mathbb{Z}$. The map

$$f : \cup_{a,b}(a, a+1) \times (b, b+1) \to S^1 \times S^1,$$

$f(x_1, x_2) = (\exp(ix_1), \exp(ix_2))$ induces a manifold structure on the torus.

Example 3. The real projective space

$$\mathbb{R}P^n = \mathbb{R}^{n+1}/\sim$$

where $\mathbf{x} \sim \mathbf{y}$ if there is a real number $\lambda \neq 0$ such that $\mathbf{x} = \lambda \mathbf{y}$. What are the coordinate charts?

Example 4. The complex projective space

$$\mathbb{C}P^n = \mathbb{C}^{n+1}/\sim$$

where $\mathbf{z} \sim \mathbf{w}$ if there is a complex number $\lambda \neq 0$ such that $\mathbf{z} = \lambda \mathbf{w}$.

Example 5. If M_1 and M_2 are manifolds of dimension n_1 and n_2, then $M_1 \times M_2$ is a manifold of dimension $n_1 + n_2$. If $f_1 : \amalg_\alpha U_\alpha \to M_1$ and $f_2 : \amalg_\beta V_\beta \to M_2$ are the maps that define M_1 respectively M_2, then $f : \amalg_\alpha U_\alpha \times \amalg_\beta U_\beta \to M_1 \times M_2$, $f(x, y) = (f_1(x), f_2(y))$ is the map that defines the manifold structure on the product.

As such, the n-dimensional torus $(S^1)^n$ is an n-dimensional manifold.

Example 6. The figure eight is not a manifold. This is not easy to prove, the proof requires examining the number of connected components obtained by removing the "crossing point" from a small open set containing it.

Chapter 2

Closed sets, connected and compact spaces

2.1 Closed sets and related notions

2.1.1 Closed sets

The natural generalization of a closed interval is that of a closed set.

Definition. A subset A of a topological space X is said to be closed if the set $X \backslash A$ is open.

Example 1. In the standard topology on $\mathbb{R}$, each singleton $\{x\}$, $x \in \mathbb{R}$ is closed.

Example 2. The Cantor set.

$$C = [0,1] \backslash \cup_{n=1}^{\infty} \cup_{k=0}^{3^{n-1}-1} \left(\frac{3k+1}{3^n}, \frac{3k+2}{3^n} \right).$$

Alternatively, the Cantor set consists of all numbers in $[0,1]$ that allow a ternary expansion with only the digits 0 and 2 (note that $1 = .2222...$, so it is in the Cantor set.)

Example 3. The Sierpinski triangle (Figure 2.1). It is obtained by starting with the set T consisting of an equilateral triangle together with its interior. Divide T into four congruent triangles, then remove the interior of the triangle in the middle. Repeat this operation with each of the three other equilateral triangle, and then continue forever.

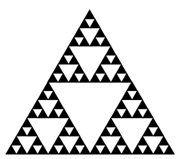

Figure 2.1:

Example 4. In the discrete topology every set is both closed and open.

Example 5. In the topology on $\mathbb{Q}$ induced by the standard topology on $\mathbb{R}$, every set of the form $(a,b) \cap \mathbb{Q}$, with a, b irrational is both open and closed.

Example 6. In the standard topology on $\mathbb{R}^n$, each set of the form

$$\overline{B}(\mathbf{x},\epsilon) = \{\mathbf{y} \in \mathbb{R} \,|\, d(\mathbf{x},\mathbf{y}) \leq \epsilon\}$$

is closed.

Example 7. In the Zariski topology the closed sets are the algebraic sets (called by some algebraic varieties), which are the sets of solutions of a system of polynomial equations.

As a corollary of de Morgan's laws, we obtain the following result.

Proposition 2.1.1. In a topological space X, the following are true:

(1) X and $\emptyset$ are closed.

(2) Arbitrary intersections of closed sets are closed.

(3) Finite unions of closed sets are closed.

The notion of a closed set is well behaved with respect to taking subspaces and products of topological spaces.

Proposition 2.1.2. (1) If Y is a subspace of X then $A \subset Y$ is closed if and only if $A = B \cap Y$ with B a closed subset of X.
(2) Let Y be a subspace of X. If A is closed in Y and Y is closed in X, then A is closed in X.
(3) If A is closed in X and B is closed in Y, then $A \times B$ is closed in $X \times Y$.
(4) If A_α, $\alpha \in \mathcal{A}$ are closed, then $\prod_\alpha A_\alpha$ is closed in the product topology.

Proof. (1) If B is closed in X, then $X\backslash B$ is open. Thus $A = B \cap Y$ is the complement of the open set $(X\backslash B) \cap Y$, and hence is closed.

For the converse, if A is closed then $Y\backslash A$ is open, thus there is an open set U in X such that $U \cap Y = Y\backslash A$. Then $B = X\backslash U$ is closed and $A = B \cap Y$, as desired.

(2) If A is closed in Y, then $Y\backslash A$ is open in Y, so there is an open $U \subset X$ such that $Y\backslash A = Y \cap U$. Then

$$X\backslash A = U \cup (X\backslash Y)$$

which is a union of open sets, so it is open. Consequently A is closed in X.

(3) This follows from

$$(X \times Y)\backslash(A \times B) = X \times (Y\backslash B) \cup (X\backslash A) \times Y.$$

(4) We have

$$\prod_\alpha A_\alpha = \cap_\alpha A_\alpha \times \prod_{\beta\neq\alpha} A_\beta.$$

By (3) each of the sets $A_\alpha \cap \prod_{\beta\neq\alpha} A_\beta$ is closed, so their intersection is also closed. □

Also, we have the following "alternative definition" of continuity.

Proposition 2.1.3. Let X and Y be topological spaces. Then $f : X \to Y$ is continuous if and only if the preimage of every closed set is closed.

Proof. Since

$$f^{-1}(Y\backslash A) = X\backslash f^{-1}(A)$$

the condition from the statement is equivalent to the fact that the preimage of every open set is open. □

2.1.2 Closure and interior of a set

Definition. Given a subset A of a topological space X, the *interior* of A, denoted by $\mathrm{Int}(A)$, is the union of all open sets contained in A and the *closure* of A, denoted by $\overline{A}$, is the intersection of all closed sets containing A.

Because arbitrary unions of open sets are open, the interior of a set is open; it is the largest open set contained in the set. Also, because arbitrary intersections of closed sets are closed, the closure of a set is closed; it is the smallest closed set containing the given set. We have

$$\mathrm{Int}(A) \subset A \subset \overline{A}.$$

Note also that A is closed if and only if $\overline{A} = A$ and A is open if and only if $\mathrm{Int}(A) = A$.

Lemma 2.1.1. For every set $A \subset X$,

$$\overline{X \backslash A} = X \backslash \mathrm{Int}(A).$$

Proof. We have $X\backslash A \subset X\backslash \mathrm{Int}(A)$, so $\overline{X\backslash A} \subset X\backslash \mathrm{Int}(A)$. For the converse inclusion, note that $X\backslash\overline{X\backslash A} \subset A$, and because it is open, we have $X\backslash\overline{X\backslash A} \subset \mathrm{Int}(A)$. Hence $X\backslash\mathrm{Int}(A) \subset \overline{X\backslash A}$. $\square$

Example 1. For $\mathbb{Q} \subset \mathbb{R}$ with the subset topology we have $\mathrm{Int}(\mathbb{Q}) = \emptyset$ and $\overline{\mathbb{Q}} = \mathbb{R}$.

Example 2. The closure of an open ball in $\mathbb{R}^n$ is the closed ball with the same center and radius. The interior of a closed ball in $\mathbb{R}^n$ is the open ball with the same center and radius.

Definition. A subset A of a topological space X is called *dense* if $\overline{A} = X$.

Example. The set of polynomials is dense in the space of continuous functions with the sup norm (this is the content of the Stone-Weierstrass Theorem).

Theorem 2.1.1. Let A be a subset of a topological space X. Then x is in $\overline{A}$ if and only if every open set U containing x intersects A. Moreover, it suffices for the condition to be verified only for basis elements containing x.

Proof. Note that indeed, the two conditions are equivalent because for every open set U containing x, there is a basis element B such that $x \in B \subset U$.

For the direct implication, we use Lemma 2.1.1 for $X\backslash A$:

$$\overline{A} = X\backslash \mathrm{Int}(X\backslash A).$$

Let $x \in X$ and assume there is $U \subset X\backslash A$ open, such that $x \in U$. Then $U \subset \mathrm{Int}(X\backslash A)$, which shows that $x \in \mathrm{Int}(X\backslash A)$. This implies that $x \notin \overline{X\backslash(X\backslash A)} = \overline{A}$.

Conversely, assume that every open set that contains x intersects A. Then $\mathrm{Int}(X\backslash A)$ does not contain x, so $x \in X\backslash \mathrm{Int}(X\backslash A) = \overline{A}$. $\square$

So x is in $\overline{A}$ if and only if every neighborhood of x intersects A. Let us see now how the closure behaves under passing to a subspace and under products.

Proposition 2.1.4. (1) Let Y be a subspace of X and A a subset of Y. Let $\overline{A}_X$ denote the closure of A in X. Then the closure of A in Y equals $\overline{A}_X \cap Y$.
(2) Let Y be a closed subspace of X, and A a subset of Y. Then the closure of A in X and Y is the same.
(3) Let (X_α), $\alpha \in \mathcal{A}$, be a family of topological spaces, and let $A_\alpha \subset X_\alpha$, $\alpha \in \mathcal{A}$. If we endow $\prod X_\alpha$ with either the product or the box topology, then

$$\prod \overline{A_\alpha} = \overline{\prod A_\alpha}.$$

Proof. (1) Let $\overline{A}_Y$ be the closure of A in Y. The set $\overline{A}_X$ is closed in X, so $\overline{A}_X \cap Y$ is closed in Y. This means that $\overline{A}_X \cap Y$ contains $\overline{A}_Y$. On the other hand, by Theorem 2.1.1, every point $x \in \overline{A}_X \cap Y$ has the property that every open set $U \subset X$ intersects A. It follows that $U \cap Y$ intersects A as well, so $x \in \overline{A}_Y$ by Theorem 2.1.1.

(2) As seen above, $\overline{A}_Y \subset \overline{A}_X$. Also, $\overline{A}_Y$ is closed in X by Proposition 2.1.2. Hence $\overline{A}_Y \supset \overline{A}_X$. Consequently $\overline{A}_X = \overline{A}_Y$.

(3) We prove the equality by double inclusion. Let $\mathbf{x} = (x_\alpha)$ be a point in $\prod \overline{A_\alpha}$. Let $U = \prod U_\alpha$ be a basis element in either topology that contains $\mathbf{x}$. Then $U_\alpha \cap A_\alpha$ is nonempty (when we have the product topology all but finitely many of the U_α's coincide with X_α's. If y_α, $\alpha \in \mathcal{A}$ are points in the intersections, then $U \cap \prod \overline{A_\alpha}$ contains (y_α). By Theorem 2.1.1, $\mathbf{x} \in \overline{\prod A_\alpha}$.

Conversely, let $\mathbf{x} = (x_\alpha)$ be a point in $\overline{\prod A_\alpha}$. For a given A_{α_0}, and an open set U_{α_0} containing x_α, the set

$$U = U_{\alpha_0} \times \prod_{\alpha \neq \alpha_0} X_\alpha$$

intersects $\prod A_\alpha$. Then U_{α_0} must intersect A_{α_0}, so $x_{\alpha_0} \in \overline{A_{\alpha_0}}$. This proves the other inclusion. □

Regarding the properties of the interior, it is not true that if Y is a subspace of X and $A \subset Y$ then the interior of A in Y is the intersection with Y of the interior of A in X; the interior of A in Y might be larger. Nor is it true that, for infinitely many spaces, the product of the interiors is the interior of the product in the product topology. We only have

Proposition 2.1.5. If X_α, $\alpha \in \mathcal{A}$, are topological spaces and $A_\alpha \subset X_\alpha$, then $\prod \mathrm{Int}(A_\alpha)$ equals the interior of $\prod A_\alpha$ in the box topology.

Proof. Since $\prod \mathrm{Int}(A_\alpha)$ is open in the box topology, it is included in $\mathrm{Int}(\prod_\alpha A_\alpha)$. If $\mathbf{x} \notin \prod \mathrm{Int}(A_\alpha)$, for each α there is U_α such that $x_\alpha \in U_\alpha$ and $U_\alpha \cap (X_\alpha \backslash A_\alpha) \neq \emptyset$. Consequently, the open set $\prod U_\alpha$ contains $\mathbf{x}$, and so by Theorem 2.1.1, $\mathbf{x} \in \overline{\prod X_\alpha \backslash A_\alpha}$. Hence $x \notin \mathrm{Int}(\prod A_\alpha)$. □

As a corollary, for *finitely* many spaces, the product of the interiors is the interior of the *product* in the product topology.

There is a characterization of continuity using closures of sets.

Proposition 2.1.6. Let X, Y be topological spaces. Then $f : X \to Y$ is continuous if and only if for every subset A of X, one has

$$f(\overline{A}) \subset \overline{f(A)}.$$

Proof. Assume that f is continuous and let A be a subset of X. Let also $x \in \overline{A}$. For an open set U in Y containing $f(x)$, $f^{-1}(U)$ is open in X, so by Theorem 2.1.1 it intersects A. Hence U intersects $f(A)$, showing that $f(x) \in \overline{f(A)}$.

Conversely, let us assume that $f(\overline{A}) \subset \overline{f(A)}$ for all subsets A of X, and show that f is continuous. We will use Proposition 2.1.3. Let B be closed in Y and $A = f^{-1}(B)$. We wish to prove that A is closed in X, namely that $A = \overline{A}$. We have

$$f(\overline{A}) \subset \overline{f(A)} = \overline{B} = B = f(A).$$

Hence the conclusion. □

2.1.3 Limit points

Definition. Let X be a topological space, A a subset, and $x \in X$. Then x is said to be a *limit point* (or accumulation point) of A if every open set containing x intersects A in some point other than x itself.

This means that x is a limit point of A if and only if every neighborhood of x contains a point in A which is not x. Said differently, x is a limit point of A if it belongs to the closure of $A\backslash\{x\}$. The set of all limit points of a set A is denoted by A'.

Example 1. If $A = \{1/n \,|\, n = 1, 2, 3, \ldots\}$, then $A' = \{0\}$.

Example 2. If $A = (0, 1) \subset \mathbb{R}$, in the standard topology, then $A' = [0, 1]$.

Example 3. If C is the Cantor set (see §2.1.1) then $C' = C$ (prove it).

Example 4. For $\mathbb{Z} \subset \mathbb{R}$, $\mathbb{Z}' = \emptyset$.

Proposition 2.1.7. Let A be a subset of a topological space X. Then

$$\overline{A} = A \cup A'.$$

Proof. A point x is in $\overline{A}$ if and only if every open set U containing x intersects A. If for some x that intersection is x itself, then $x \in A$. Otherwise $x \in A'$ by definition. □

Corollary 2.1.1. A subset of a topological space is closed if and only if it contains all its limit points.

For metric spaces, limit points can be characterized using convergent sequences.

Definition. In an arbitrary topological space, one says that a sequence $(x_n)_n$ of points in X *converges* to a point $x \in X$ provided that, corresponding to each neighborhood V of x, there is a positive integer N such that $x_n \in V$ for all $n \geq N$. The point x is called the *limit* of x_n.

The notion of convergence can be badly behaved in arbitrary topological spaces, for example in the trivial topology any sequence converges to all points in the space. In the Zariski topology on $\mathbb{C}$, all sequences that do not contain constant subsequences converge to all points in $\mathbb{C}$. In metric spaces however, we have the following result.

Proposition 2.1.8. Given a metric space X with metric d, if a sequence $(x_n)_n$ converges, then its limit is unique.

Proof. Assume that $(x_n)_n$ converges to both x and y, $x \neq y$. Then for every ϵ, all terms of the sequence but finitely many lie in both $B(x, \epsilon)$ and $B(y, \epsilon)$. But for $\epsilon < d(x, y)/2$, this is impossible, since the balls do not intersect. Hence $(x_n)_n$ can have at most one limit. □

In metric spaces the closure and the limit points of a set can be described in terms of convergent sequences.

Lemma 2.1.2. (The sequence lemma) Let X be a metric space and A a subset of X.

(1) A point x is in $\overline{A}$ if and only if there is a sequence of points in A that converges to x.

(2) A point x is in A' if and only if there is a sequence of points in A converging to x that does not eventually become constant.

Proof. Using Proposition 2.1.7 we see that (2) implies (1) since if $x \in A$ we can use the constant sequence $x_n = x$, $n \geq 1$.

To prove (2), assume first that $x \in A'$. Then for every ϵ, there is a point $y \neq x$ in A such that $y \in B(x, \epsilon)$. Start with $\epsilon = 1$ and let x_1 be such a point. Consider the ball $B(x, d(x, x_1)/2)$ and let $x_2 \neq x$ be a point of A that lies in this ball. Choose $x_3 \in B(x, d(x, x_2)/2)$ in the same fashion, and repeat to obtain the sequence $x_1, x_2, \ldots, x_n, \ldots$, whose terms are all distinct.

Because $d(x, x_n) \to 0$, and because for every neighborhood V of x there is an ϵ such that $B(x, \epsilon) \subset V$, it follows that all but finitely many terms of the sequence are in V. Hence $(x_n)_n$ is a sequence of points in A converging to x that does not eventually become constant.

Conversely, assume that there is a sequence $(x_n)_n$ of points in A convering to x that does not eventually become constant. Given an arbitrary neighborhood V of x, there are infinitely many terms of the sequence in that neighborhood, and infinitely many of those must be different from x. So $x \in A'$ by definition. □

In fact one of the implications in (1) is true in topological spaces, namely if there is a sequence $(x_n)_n$ of points in A that converges to x then $x \in \overline{A}$. Indeed, by the definition of convergence, every neighborhood of x contains infinitely many points of the sequence, hence it contains points in A. By Theorem 2.1.1, $x \in \overline{A}$.

For metric spaces continuity can also be characterized in terms of convergent sequences.

Theorem 2.1.2. Let X be a metric space and Y a topological space. Then $f : X \to Y$ is continuous if and only if for every $x \in X$ and every sequence $(x_n)_n$ in X that converges to x, $f(x_n)$ converges to $f(x)$.

Proof. Assume that f is continuous and that $x_n \to x$. If V is a neighborhood of $f(x)$, then $f^{-1}(V)$ is a neighborhood of x, which contains therefore all but finitely many terms of the sequence. Hence all but finitely many terms of $(f(x_n))_n$ are in V. This proves that $f(x_n) \to f(x)$.

For the converse we will use Proposition 2.1.6. Let A be a subset of X and $x \in \overline{A}$. Then by Lemma 2.1.2, there is a sequence $(x_n)_n$ of points in A such that $x_n \to x$. Then $f(x_n) \to f(x)$ so by the same lemma, $f(x) \in \overline{f(A)}$. It follows that $f(\overline{A}) \subset \overline{f(A)}$, which proves that f is continuous. □

2.2 Hausdorff spaces

Topologies in which sequences converge to more than one point are are counterintuitive and they seldom show up in other branches of mathematics, the Zariski topology being a rare example. We will therefore introduce a large class of "nice" topological spaces in which this bizarre phenomenon does not occur.

Definition. A topological space X is called a *Hausdorff space* if for each pair x_1, x_2 of distinct points of X, there exist neighborhoods U_1 and U_2 of x_1 respectively x_2 that are disjoint.

Example 1. Every metric space is a Hausdorff space.

Example 2. The product space $\prod_{n=1}^{\infty} \mathbb{R}$ is Hausdorff but is not a metric space. To see that it is Hausdorff, choose two points $\mathbf{x} \neq \mathbf{y}$. Then there is some n such that $x_n \neq y_n$. Choose neighborhood U and V of x_n and y_n in $\mathbb{R}$ such that $U \cap V = \emptyset$. Then

$$\prod_{i=1}^{n-1} \mathbb{R} \times U \times \prod_{i=n+1}^{\infty} \mathbb{R} \text{ and } \prod_{i=1}^{n-1} \mathbb{R} \times V \times \prod_{i=n+1}^{\infty} \mathbb{R}$$

are disjoint neighborhoods of $\mathbf{x}$ and $\mathbf{y}$ in $\prod_{n=1}^{\infty} \mathbb{R}$.

Example 3. $\mathbb{C}^n$ endowed with the Zariski topology is *not* Hausdorff.

Proposition 2.2.1. If X is a Hausdorff space and $x \in X$, then $\{x\}$ is a closed set.

Proof. For $y \in X \backslash \{x\}$ there is an open neighborhood V of y such that $x \notin V$. Hence $V \subset X \backslash \{x\}$, so $X \backslash \{x\}$ is open. Hence $\{x\}$ is closed. □

As a corollary, finite subsets of a Hausdorff space are closed.

Proposition 2.2.2. (1) A subspace of a Hausdorff space is Hausdorff.
(2) The product of Hausdorff spaces is a Hausdorff space in both the product and the box topology.

Proof. (1)Let $Y \subset X$, and let $x, y \in Y$. Then there are disjoint open sets U, V in X such that $x \in U$, $y \in V$. Then the sets $U \cap Y$ and $V \cap Y$ are open in Y, still disjoint, and the first contains x, the second contains Y.

(2) Let X_α, $\alpha \in \mathcal{A}$ be Hausdorff. If $(x_\alpha)_\alpha$ and $(y_\alpha)_\alpha$ are distinct, then there is α_0 such that $x_{\alpha_0} \neq y_{\alpha_0}$. There are disjoint open sets U_{α_0} and V_{α_0} such that $x_{\alpha_0} \in U_{\alpha_0}$ and $y_{\alpha_0} \in V_{\alpha_0}$. We conclude that the open sets

$$U_{\alpha_0} \times \prod_{\alpha \neq \alpha_0} X_\alpha \text{ and } V_{\alpha_0} \times \prod_{\alpha \neq \alpha_0} X_\alpha$$

separate $(x_\alpha)_\alpha$ from $(y_\alpha)_\alpha$. □

Remark 2.2.1. In a Hausdorff space a convergent sequence has exactly one limit. Indeed, if $x \neq y$ were limits of the sequence, and U and V are disjoint neighborhoods of x respectively y, then both U and V should contain all but finitely many terms of the sequence, which is impossible.

Example. The space $\mathbb{C}^n$ with the standard topology and the space $\mathbb{C}^n$ with the Zariski topology are not homeomorphic because one is Hausdorff and one is not.

2.3 Connected spaces

2.3.1 The definition of a connected space and properties

Definition. Let X be a topological space. Then X is called *connected* if there are no disjoint nonempty open sets U and V such that $X = U \cup V$.

If such U and V exist then they are said to form a separation of X. Thus X is not connected if it has a separation. Another way of formulating the definition is to say that the only subspaces of X that are both open and closed are X and the empty set.

Connectedness is difficult to verify. It is much easier to disprove it.

Example 1. The real line is connected. (We will prove this later).

Example 2. The set of rational numbers $\mathbb{Q}$ with the topology induced by the standard topology on $\mathbb{R}$ is not connected. Indeed, the open sets $(-\infty, \sqrt{2}) \cap \mathbb{Q}$ and $(\sqrt{2}, \infty) \cap \mathbb{Q}$ are a separation of $\mathbb{Q}$. In fact for every two points a and b of $\mathbb{Q}$, there is a separation $\mathbb{Q} = U \cup V$ with $a \in U$ and $b \in V$. We say that $\mathbb{Q}$ is *totally disconnected*.

Proposition 2.3.1. (1) If A and B are two disjoint nonempty subsets of a topological space X such that $X = A \cup B$ and neither of the two subsets contains a limit point of the other, then A and B form a separation of X.
(2) If U and V form a separation of X and if Y is a connected subspace of X, then Y lies entirely within either U or V.

Proof. (1) Since $\overline{A} \subset X \backslash B$, it follows that $\overline{A} = A$. Similarly, $\overline{B} = B$. So A and B are closed, which means that their complements, which are again A and B, are open. So A and B form a separation of X.

(2) If this were not true, then $Y \cap U$ and $Y \cap V$ were a separation of Y. □

Theorem 2.3.1. The image under a continuous map of a connected space is connected.

Proof. This is a powerful result with a trivial proof. If $f : X \to Y$ is continuous and $f(X)$ is not connected, and if U and V are a separation of $f(X)$, then $f^{-1}(U)$ and $f^{-1}(V)$ are a separation of X. □

Proposition 2.3.2. (1) The union of a collection of connected spaces that have a common point is connected.
(2) Let A be a connected dense subspace of a topological space X. Then X is connected.
(3) The product of connected spaces is connected in the product topology.

Proof. (1) Let $X = \cup_\alpha X_\alpha$ and a be a common point of the X_α's. Assume that $U \cup V$ is a separation of X. Then by Proposition 2.3.1 (1), each X_α is included in either U or V. In fact, each is included in that of the two sets which contains a, say U. But then V is empty, a contradiction. The conclusion follows.

(2) Assume by contrary that X is not connected, and let $X = U \cup V$ be a separation of X. Then A lies entirely in one of the sets U or V, say U. But since $U = \overline{U}$, $\overline{A} = X \subset U$, a contradiction. Hence X is connected.

(3) Let us prove first that the product of two connected spaces X_1 and X_2 is connected. Fix $x_i \in X_i$, $i = 1, 2$. By part (1),

$$(\{x_1\} \times X_2) \cup (X_1 \times \{x_2\})$$

is connected being the union of two connected sets that share (x_1, x_2). Now vary x_2 and take the union of all such sets. This union is the entire space $X_1 \times X_2$, and each of the spaces contains $\{x_1\} \times X_2$. Again from (1) it follows that $X_1 \times X_2$ is connected.

An inductive argument shows that the product of finitely many connected sets is connected.

Now let us consider a product $X = \prod_\alpha X_\alpha$ $\alpha \in \mathcal{A}$ of connected spaces endowed with the product topology. For each α, fix a point $a_\alpha \in X_\alpha$. Then each set of the form

$$A_{\alpha_1,\alpha_2,\ldots,\alpha_n} = X_{\alpha_1} \times X_{\alpha_2} \times \cdots \times X_{\alpha_n} \times \prod_{\alpha \neq \alpha_i} \{a_\alpha\}$$

are connected, being finite products of connected spaces, and hence their union is also connected because these sets have the common point (a_α). Let us show that

$$A = \cup_{n=1}^{\infty} \cup_{\alpha_1,\alpha_2,\ldots,\alpha_n \in \mathcal{A}} A_{\alpha_1,\alpha_2,\ldots,\alpha_n}$$

is dense in X. Indeed, if $(x_\alpha) \in X$ and

$$B = U_{\alpha_1} \times U_{\alpha_2} \times \cdots \times U_{\alpha_n} \times \prod_{\alpha \neq \alpha_i} X_\alpha$$

is a basis element containing x, then

$$\{x_{\alpha_1}\} \times \{x_{\alpha_2}\} \times \cdots \times \{x_{\alpha_n}\} \times \prod_{\alpha \neq \alpha_i} \{a_\alpha\} \in B \cap A_{\alpha_1, \alpha_2, \ldots, \alpha_n}.$$

This shows that $\overline{A} = X$, hence X is connected. □

Remark 2.3.1. The product of infinitely many connected spaces in the box topology is not necessarily connected. For example a separation of $\mathbb{R}^{\mathbb{N}}$ in the box topology consists of the set of all bounded sequences and the set of all unbounded sequences.

Corollary 2.3.1. If $A \subset B \subset \overline{A}$, then B is also connected. This follows from (2) by letting $B = X$.

Definition. A maximal connected subset of a topological space is called a *connected component*.

Theorem 2.3.2. Every topological space can be partitioned into connected components.

Proof. Each singleton $\{x\}$ of a topological space X is connected. The union of all connected sets that contain x is connected by Proposition 2.3.2 (1). This union is a maximal connected set that contains x, hence it is a connected component. Varying x we partition the set into connected componets. □

If X and Y are homeomorphic, then there is a bijective correspondence between their connected components.

Definition. A space X is said to be *locally connected* if for every neighborhood U of x there is a connected neighborhood V of x such that $V \subset U$.

Proposition 2.3.3. A space is locally connected if and only if the connected components of any open set are open.

Proof. Let us assume that the topological space X is locally connected, and let U be an open set. If x is a point in U, then there is a connected open neighborhood of x, V, which is contained in U. But then V must lie in a connected component of U (Proposition 2.3.1 (2)). So the connected components of U are unions of open sets, so they are open.

Conversely, suppose that the connected components of open sets are open. Then the neighborhood V from the definition can be taken to be just one such connected component. □

Example 3. The comb space defined as

$$(\{0\} \times [0,1]) \cup ([0,1] \times \{0\}) \cup \cup_{n=1}^{\infty} \left(\left\{ \frac{1}{n} \right\} \times [0,1] \right)$$

is connected but not locally connected.

2.3.2 Connected sets in $\mathbb{R}$ and applications

Theorem 2.3.3. The only connected subsets of the real line in the standard topology are the intervals and $\mathbb{R}$.

Proof. Let A be a subset of $\mathbb{R}$. If there are $a, b \in A$ $a < b$ such that $[a, b]$ is not a subset of A, that is there is c, $a < c < b$ and $c \notin A$, then $(-\infty, c) \cap A$ and $(c, \infty) \cap A$ form a separation of A. So in this case A is not connected. Hence if $\alpha = \inf A$ and $\beta = \sup A$, $\alpha, \beta \in \mathbb{R} \cup \{\pm\infty\}$, then $(\alpha, \beta) \subset A \subset [\alpha, \beta]$, which shows that A is an interval or the whole space.

Conversely, let us show that $\mathbb{R}$ and all intervals are connected. If $U \cup V$ is a separation of an interval I (or of $\mathbb{R}$), let $a, b \in I$ with $a \in U$ and $b \in V$, and without loss of generality let us assume that $a < b$. Consider $c = \sup\{x \mid x < b, x \in U\}$. Then $c \in \overline{U}$ on the one hand, and because $c = \inf\{x \mid x \in V\}$, $c \in \overline{V}$. But this is impossible. It follows that I (and for the same reason $\mathbb{R}$) does not admit a separation. $\square$

As a corollary of Proposition 2.3.2 we obtain the following examples of connected spaces:

Example 1. The product $[0, 1]^a$, where a can be a positive integer or can be infinite is connected.

Example 2. Every polygonal line in the plane is connected.

Example 3. The comb is connected.

Theorem 2.3.1 becomes the well known

Theorem 2.3.4. (The intermediate value theorem) Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then f maps intervals to intervals.

Here intervals can consist of just one point (for example when f is constant). Let us see some applications.

Theorem 2.3.5. Let $f : [a, b] \to [a, b]$ be a continuous map. Then f has a fixed point, meaning that there is $x \in [a, b]$ such that $f(x) = x$.

Proof. Assume f has no fixed points. Consider the function $g : [a, b] \to \mathbb{R}$, $g(x) = f(x) - x$. Then $g([a, b])$ is an interval. We have $f(a) > a$ and $f(b) < b$ (because f has no fixed points), so $g([a, b])$ is an interval that contains both positive and negative numbers. It must therefore contain 0. This is a contradiction, which proves that f has a fixed point. $\square$

Theorem 2.3.6. (The one-dimensional Borsuk-Ulam theorem) Given a continuous map of a circle into a line, there is a pair of diametrically opposite points that are mapped to the same point.

Proof. First let us notice that S^1 is connected, because it is the image of $\mathbb{R}$ through the continuous map $f(x) = e^{ix}$. Let $f : S^1 \to \mathbb{R}$ be the continuous map. For a point $z \in S^1$, the diametrically opposite point is $-z$. Define $g : S^1 \to \mathbb{R}$, $g(z) = f(z) - f(-z)$. If for some z, $g(z) = 0$, then z has the desired property. If for some z, $g(z) > 0$, then $g(-z) < 0$, and because the set $g(S^1)$ is connected, this set must contain 0. The conclusion follows. $\square$

Theorem 2.3.7. Let A and B be two polygonal regions in the plane. Then there is a line that divides each of the regions in two (not necessarily connected) parts of equal areas.

Proof. For each given line ℓ there is one and only one line parallel to ℓ that divides A into two regions of equal areas, and one and only one line parallel to ℓ that divides B into two parts of equal areas.

Now fix a line ℓ_0 in the plane, then fix on ℓ_0 a point 0, as well as a positive direction and a unit of length. Consider the lines perpendicular to ℓ_0 that cut A respectively B into equal areas, and let x_A and x_B be the coordinates of their intersections with ℓ_0. Now rotate ℓ_0 keeping 0 fixed, and let $x_A(\theta)$ and $x_B(\theta)$ be now the same coordinates on ℓ_0 depending on the angle of rotation.

Define $g : [0, 2\pi] \to \mathbb{R}$, $g(\theta) = x_A(\theta) - x_B(\theta)$. Then $g(0) = -g(\pi)$ (in fact $g(x) = -g(x+\pi)$ for all x). The function g is continuous, and $g([0, 2\pi])$ must be an interval. This interval must contain both nonpositive and nonnegative numbers, hence it contains 0. Thus there is an angle θ such that $x_A(\theta) = x_B(\theta)$. In this case the two lines perpendicular to ℓ_0 coincide, they form a line that cuts both A and B in parts of equal area. □

2.3.3 Path connected spaces

There is a property that is much easier to verify in particular applications, and which guarantees that a space is connected. This is the property of being path connected.

Definition. Given a topological space X and $x, y \in X$, a *path* from x to y is a continuous map $\phi : [0, 1] \to X$ such that $f(0) = x$ and $f(1) = y$.

In fact any continuous map $\phi : [a, b] \to X$, $\phi(a) = x$, $\phi(b) = y$ defines a path, since we can rescale it to $\psi(t) = \phi((b-a)t + a)$.

Proposition 2.3.4. The relation on X defined by $x \sim y$ if there is a path from x to y is an equivalence relation.

Proof. Clearly $x \sim x$ by using the constant path. Also, if $\phi : [0, 1] \to X$ is a path from x to y, then $\psi(t) = \phi(1-t)$ is a path from y to x. Hence if $x \sim y$ then $y \sim x$.

Finally, if $x \sim y$ and $y \sim z$, that is if there are paths $\phi_1, \phi_2 : [0, 1] \to X$ from x to y and from y to z, then

$$\psi(t) = \begin{cases} \phi_1(2t) & \text{if } 0 \leq t \leq 1/2 \\ \phi_2(2t-1) & \text{if } 1/2 \leq t \leq 1 \end{cases}$$

is a path from x to z. □

Definition. The equivalence classes of $\sim$ are called the *path components* of X.

Note that $\sim$, being an equivalence relation, partitions X into its path components.

Definition. If the space X consists of only one path component, it is called *path connected.*

Proposition 2.3.5. Each path component of a topological space X is included in a connected component. Consequently, a path connected space is connected.

Proof. Each path is connected, being the image of a connected set through a continuous map, so by Proposition 2.3.1 its image is included in a connected component of X. This means that if $x \sim y$, then x and y belong to the same connected component of X. Hence the conclusion. □

Proposition 2.3.6. (1) The union of a collection of connected spaces that have a common point is connected.
(3) The product of path connected spaces is path connected.

The property of a space to be path connected is well-behaved under continuous maps.

Theorem 2.3.8. Let $f : X \to Y$ be a continous map from the path-connected topological space X to the topological space Y. Then $f(X)$ is path connected.

Proof. If ϕ is a path from x to y, then $f \circ \phi$ is a path from $f(x)$ to $f(y)$. □

Corollary 2.3.2. If X and Y are homeomorphic, then there is a bijective correspondence between their path components.

Example 1. Every convex set in an $\mathbb{R}$-vector space is path connected. Indeed, a set A is convex if for every $x, y \in A$, the segment $\{tx + (1-t)y \,|\, t \in [0,1]\}$ is in A. This segment *is* the path.

In particular every $\mathbb{R}$-vector space, such as $\mathbb{R}^n$, $C[a,b]$, $L^p(\mathbb{R})$, is path connected.

Example 2. Every curve in $\mathbb{R}^n$ is path connected, being the continuous image of an interval.

Example 3. If $n \geq 2$ and $\mathbf{x} \in \mathbb{R}^n$, then $\mathbb{R}^n \backslash \{\mathbf{x}\}$ is path connected.

Indeed, given $\mathbf{y}$ and $\mathbf{z}$ in $\mathbb{R}^n \backslash \{\mathbf{x}\}$, consider a circle of diameter $\mathbf{yz}$. Then one of the semicircles does not contain $\mathbf{x}$, and a parametrization of this semicircle defines a path from $\mathbf{y}$ to $\mathbf{z}$.

Example 4. If $x \in \mathbb{R}$, the space $\mathbb{R} \backslash \{x\}$ has two path components, which are also its connected components, namely $(-\infty, x)$ and (x, ∞).

Example 5. The n-dimensional sphere and the n-dimensional projective space are path connected. The n-dimensional sphere is the image through the continuous map $f(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ of the connected space $\mathbb{R}^{n+1} \backslash \{0\}$. The projective space is the image of the sphere through the continuous map that identifies the antipodes.

Example 6. The n-dimensional torus $(S^1)^n$ is path connected. This follows from Proposition 2.3.6.

Here are some applications.

Theorem 2.3.9. If $n \geq 2$, the spaces $\mathbb{R}$ and $\mathbb{R}^n$ are not homeomorphic.

Proof. Arguing by contradiction, let us assume that there is a homeomorphism $f : \mathbb{R}^n \to \mathbb{R}$. Choose $\mathbf{x} \in \mathbb{R}^n$. Then $f : \mathbb{R}^n \backslash \{\mathbf{x}\} \to \mathbb{R} \backslash \{f(x)\}$ is still a homeomorphism (it is one-to-one and onto, the preimage of each open set is open, and the image of each open set is open). But $\mathbb{R}^n \backslash \{\mathbf{x}\}$ is path connected, while its image through the continuous map f is not. This is a contradiction, which proves that the two spaces are not homeomorphic. $\square$

Example 4. The figure eight from §1.3.7 is not a manifold.

To prove this, recall that the figure eight is obtained by factoring $[0,1] \cup [2,3]$ by $0 \sim 1 \sim 2 \sim 3$. Let us denote this space by X. Let $\hat{0}$ be the equivalence class of 0. If X were an n-dimensional manifold, then there would be a neighborhood U of $\hat{0}$ homeomorphic to an open disk $D \subset \mathbb{R}^n$; let f be this homeomorphism. This neighborhood can be chosen small enough as to be included in $[0,1/3) \cup (2/3,1] \cup [2,7/3) \cup (8/3,3]$. Then $U \backslash \{\hat{0}\}$ is homeomorphic with $D \backslash \{f(\hat{0}\}$. But $U \backslash \{\hat{0}\}$ has at least four path components, while $D \backslash \{f(\hat{0})\}$ has either one path component, if $n \geq 2$, or two path components, if $n = 1$. This is a contradiction. Hence the figure eight is not a manifold.

Note that a similar argument shows that a 1-dimensional manifold cannot be an n-dimensional manifold for $n \geq 2$.

Definition. A topological space X is called *locally path connected* if for every $x \in X$, and open set U containing x, there is a path connected neighborhood V of x such that $V \subset U$.

Example 1. If we remove from $\mathbb{R}^2$ a finite set of lines, the remaining set (with the induced topology) is locally path connected, but not connected.

Example 2. Every manifold is locally path connected, since every point has a neighbourhood homeomorphic to a ball.

Proposition 2.3.7. (1) A topological space X is locally path connected if and only if for every open set U of X, each path component of U is open in X.
(2) If X is locally path connected, then the components and the path components are the same.

Proof. The proof of (1) is the same as for Proposition 2.3.3.

For (2), note that the path components are open, hence they form a partition of X into open sets. This means that they must also be the connected components of X (recall that the path components are connected). $\square$

And now some pathological examples.

Example 1. The topologists sine curve

$$T = \left\{ \left(x, \sin \frac{1}{x} \right) \mid x \in (0,1] \right\} \cup \{(0,0)\}$$

with the topology induced by the standard topology of $\mathbb{R}^2$.

This space is connected. Indeed, the graph of $\sin \frac{1}{x}$ is connected, because it is the image in the plane of the connected interval $(0,1]$ through the continuous map $h(x) = (x, \sin \frac{1}{x})$. Thus any separation of T must separate the origin from this graph. But any neighborhood of the origin contains a part of this graph. This proves connectivity.

The topologists sine curve is not locally connected, because any neighborhood of $(0,0)$ contained in $B((0,0), 1/2) \cap T$ is not connected (it consists of the origin and several disjoint arcs).

The topologists sine curve is not path connected. This is equivalent to the fact that $f(x) = \sin \frac{1}{x}$ cannot be extended continuously to $[0,1]$. Any path $\phi : [0,1] \to T$ would have as limit points when $t \to 0$ the entire interval $\{0\} \times [-1,1]$, and so it could not be continuous.

Example 2. The comb space defined in §2.3.1 is path connected by not locally path connected.
Example 3. The deleted comb, which is a subspace of the comb defined as

$$D = \{(0,1)\} \cup \cup_{n=1}^{\infty} \left(\left\{ \frac{1}{n} \right\} \times [0,1] \right) \cup ([0,1] \times \{0\}).$$

This space is connected but not path connected, since there is no path from $(0,1)$ to $(1,0)$ (Prove it!).

2.4 Compact spaces

2.4.1 The definition of compact spaces and examples

Definition. A collection $\mathcal{U}$ is called an *open cover* of a topological space X if the elements of $\mathcal{U}$ are open subsets of X and the union of all elements in $\mathcal{U}$ is X.

Remark 2.4.1. In general, if A is a subset of a topological space X, an open cover of A is a collection of open sets in X whose intersections with A is an open cover of A in the subspace topology.

Definition. A space X is said to be *compact* if every open cover of X contains a finite subcover (i.e. a finite family that also covers X).

Remark 2.4.2. Some mathematicians are unhappy with this very general definition, and require the space to be Hausdorff, too.

Example 1. Any topological space that has finitely many points is compact.

Example 2. $\mathbb{R}$ with the standard topology is not compact because the family $\mathcal{U} = \{(-n, n) \,|, n \geq 1\}$ is an open cover that does not have a finite subcover.

As you can see, it is much easier to prove that a space is not compact, then to prove that it is compact.

The next result will show that there are many (nontrivial) compact spaces.

Theorem 2.4.1. (The Heine-Borel Theorem) A subspace of $\mathbb{R}^n$ is compact if and only if it is closed and bounded (in the Euclidean metric).

Proof. Let us first prove that a compact set $K \subset \mathbb{R}^n$ is closed and bounded. If K were not bounded, then the collection of open balls

$$B(\mathbf{0}, k) = \{\mathbf{x} \in \mathbb{R}^n \,|\, d(\mathbf{x}, \mathbf{0}) < k\}, \quad k = 1, 2, 3, \ldots,$$

would be an open cover of K that does not have a finite subcover. If K were not closed, and $\mathbf{x} \in K' \backslash K$, then the open sets which are complements of the closed balls

$$\overline{B}(\mathbf{x}, 1/k) = \{\mathbf{y} \in \mathbb{R}^n \,|\, d(\mathbf{x}, \mathbf{y}) \leq 1/k\}, \quad k = 1, 2, 3, \ldots,$$

would be an open cover of K with no finite subcover.

For the converse, let us assume that K is closed and bounded in $\mathbb{R}^n$ but has an open cover $\mathcal{U}$ with no finite subcover. Add to $\mathcal{U}$ the complement of K, so that now we have a cover of the whole space.

Place K in an n-dimensional cube, which by a translation and rescaling, can be made $[0, 1]^n$. Cut the cube into 2^n equal cubes. Each of these cubes is covered by some sets in $\mathcal{U}$, and because the open cover of K does not have a finite subcover, there is some cube which is covered by infinitely many elements in $\mathcal{U}$, and which furthermore cannot be covered by finitely many elements in $\mathcal{U}$. Cut this cube into 2^n equal cubes, and again there would be one that cannot be covered by just finitely many open sets in $\mathcal{U}$. And this would go on forever.

Note that at kth step, the choice of the cube specifies the kth digits of the binary expansions of the coordinates of the points inside that cube. Repeating the process for all n and choosing the corresponding kth digits in the binary expansion, we define a point $\mathbf{x} \in \mathbb{R}^n$ which belongs to all cubes that were chosen. This point is in the closure of K (just because every of the cubes must contain points of K or else it is covered by the complement of K). And $\mathbf{x}$ must belong to some open set U in $\mathcal{U}$.

Because U is open, there is some open ball $B(\mathbf{x}, \epsilon)$ contained in U. Note that the kth cube in the process has diameter equal to $\sqrt{n}/2^k$, so, for k sufficiently large, it will be contained in $B(\mathbf{x}, \epsilon)$ and hence in U. This is a contradiction because that cube does not have a finite subcover. It follows that our assumption was false, and consequently K is compact. □

2.4.2 Properties of compact spaces

Proposition 2.4.1. A topological space X is compact if and only if for any collection $\mathcal{C}$ of closed subsets of X, with the property that the intersection any finitely many of them is nonempty, the intersection of all elements of $\mathcal{C}$ is nonempty.

Proof. By looking at the complements of the elements in $\mathcal{C}$ and applying de Morgan's law the condition from the statement turns into the definition of compactness. □

Definition. The collection $\mathcal{C}$ as in the statement of this result is said to have *the finite intersection property.*

Proposition 2.4.2. (1) Given a subspace Y of X, Y is compact if and only if every cover by open sets in X has a finite subcollection that covers Y.
(2) Every closed subspace of a compact space is compact.
(3) Every compact subspace of a Hausdorff space is closed.
(4) If Y is a compact subspace of a Hausdorff space X and if x is not in Y, then there are disjoint open sets U and V of X such that $Y \subset U$ and $x \in V$.
(5) (The tube lemma) Consider the product space $X \times Y$, where Y is compact. If x_0 is in X and N is an open set of $X \times Y$ containing $\{x_0\} \times Y$, then N contains a set of the form $W \times Y$ with W a neighborhood of x_0 in X.

Proof. (1) This follows from the fact that the open sets in Y are those of the form $Y \cap U$ with U open in X.

(2) Given $Y \subset X$ with X compact and Y closed, any open cover $\mathcal{U}$ of Y by open sets of X can be extended to an open cover of X by adding the open set $X\backslash Y$. This will have a finite subcover of X, which is a finite cover of Y as well. We can remove the set $X\backslash Y$ from this collection and still have a finite subcover of Y. Using (1) we conclude that Y is compact.

(3) If Y is a compact subspace of the Hausdorff space X, then for every $y \in Y$ and $x \in X\backslash Y$, there are disjoint open sets $U_{x,y}$ and $V_{x,y}$ in X such that $x \in U_{x,y}$ and $y \in V_{x,y}$. Fix x. The sets $V_{x,y}$ form an open cover of Y, from which we can extract a finite subcover $V_{x,y_1}, V_{x,y_2}, \ldots, V_{x,y_n}$. The open set $U_{x,y_1} \cap U_{x,y_2} \cap \cdots \cap U_{x,y_n}$ contains x and is disjoint from Y. It follows that $X\backslash Y$ is open so Y is closed.

(4) This is just a corollary of the proof of (3).

(5) Choose an open cover of this set by basis elements of the form $U \times V$ that are included in N. Since $\{x_0\} \times Y$ is compact, there is a subcover $U_1 \times V_1, U_2 \times V_2, \ldots, U_n \times V_n$. If we set $W = U_1 \cap U_2 \cap \ldots \cap U_n$, then $\{x_0\} \times Y \subset W \times Y \subset N$ and we are done. □

The most important property of compact spaces is the following result:

Theorem 2.4.2. The image of a compact space through a continuous function is compact.

Proof. Let $f : X \to Y$ be continuous with X compact. Let $\mathcal{U}$ be an open cover of $f(X)$. The collection of open sets

$$\{f^{-1}(U) \,|\, U \in \mathcal{U}\}$$

is an open cover of X, which has a finite subcover because X is compact. The image through f of that subcover is a finite subcover of $f(X)$. □

We list two useful corollaries of this theorem.

Theorem 2.4.3. Let $f : X \to Y$ be a bijective continuous function. If X is compact and Y is Hausdorff, then f is a homeomorphism.

Proof. The only thing that we have to show is that f^{-1} is continuous. We use the definition of continuity based on closed sets from Proposition 2.1.3 (2). Let C be closed in X. By Proposition 2.4.2, C is compact in X, so $(f^{-1})^{-1}(C) = f(C)$ is compact in Y. The space Y being Hausdorff, $f(C)$ is closed by Proposition 2.1.3 (3). Hence the preimage under f^{-1} of every closed set $C \subset X$ is a closed subset of Y. Therefore f^{-1} is continuous. □

Theorem 2.4.4. If $f : X \to \mathbb{R}$ is continuous and X is compact, then f has an absolute maximum and minimum.

Proof. The set $f(X)$ is compact in $\mathbb{R}$, so by the Heine-Borel Theorem it is closed and bounded. The maximum and minimum of this set are the maximum and the minimum of f. $\square$

This theorem is very useful, and we list below several applications.

Theorem 2.4.5. (The Lebesgue number theorem) Let $\mathcal{U}$ be an open covering of the compact metric space X. Then there is $\delta > 0$, called the *Lebesgue number*, such that for each subset of X having diameter less than δ, there is an element of $\mathcal{U}$ containing it.

Proof. If X belongs to $\mathcal{U}$, we are done. Otherwise, choose a finite subcover $U_1, U_2, \ldots, U_n$, and consider the closed sets $C_i = X \backslash U_i$, $i = 1, 2, \ldots, n$.

Recall that the diameter of a set A is

$$\operatorname{diam}(A) = \sup\{d(a_1, a_2) \mid a_1, a_2 \in A\},$$

where d is the distance function on X. Additionally, for a point $x \in X$ and a set $A \subset X$, define

$$d(x, A) = \inf\{d(x, a) \mid a \in A\}.$$

Define $f : X \to \mathbb{R}$,

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} d(x, C_i).$$

Let us show that f is continuous, which amounts to showing that $d(x, C_i)$ is a continuous function in x.

Lemma 2.4.1. If A is a subset of the metric space X, then $d(x, A)$ is a continuous function of x.

Proof. We will show that if $x_n \to x$, then $d(x_n, A) \to d(x, A)$. Indeed, if $d(x_n, x) \leq \epsilon$, then by the triangle inequality $|d(x_n, a) - d(x, a)| \leq d(x, x_n) < \epsilon$. If we choose a such that $d(x, a) - d(x, A) < \epsilon$, the $d(x_n, a) < d(x, A) + 2\epsilon$. Hence $d(x_n, A) < d(x, A) + 2\epsilon$. Also, if we choose a such that $d(x_n, a) - d(x_n, A) < \epsilon$, then $d(x, a) < d(x_n, A) + 2\epsilon$, and so $d(x, A) < d(x_n, A) + 2\epsilon$. Thus $|d(x, A) - d(x_n, A)| < 2\epsilon$, and the lemma is proved. $\square$

By Theorem 2.4.4, f has a minimum. If $x \in X$ is a minimum of f and U_i contains x, then there is a ball $B(x, \epsilon)$ contained in U_i. So $d(x, C_i) > 0$ and consequently the minimal value of f is strictly positive.

Let us show that we can choose the Lebesgue number δ to be the minimum of f. If A is a set of diameter let than δ, and $x_0 \in A$, then $A \subset B(x_0, \delta)$. Let us show that $B(x_0, \delta)$ lies in one of the sets U_i. If this is not the case, then $d(x_0, C_i) < \delta$, and hence

$$f(x_0) = \frac{1}{n}\sum_{i=1}^{n} d(x_0, C_i) < \frac{1}{n} n\delta = \delta.$$

This is impossible, since $f(x_0)$ has to be at least δ. Therefore $B(x_0, \delta)$ lies in some U_i and we are done. $\square$

Corollary 2.4.1. Let $f : X \to Y$ be a continuous function from the compact metric space X to the metric space Y. Then f is uniformly continuous.

Proof. Let $\epsilon > 0$. Then for every $x_0 \in X$ there is $\delta_{x_0} > 0$ such that if $d_X(x, x_0) < \delta_{x_0}$, then $d_Y(f(x), f(x_0)) < \epsilon/2$.

The balls $B(x_0, \delta_{x_0})$, $x_0 \in X$ cover X. If δ is their Lebesgue number, then every set of diameter δ lies inside some $B(x_0, \delta_{x_0})$. If x_1, x_2 have the property that $d_X(x_1, x_2) < \delta$, and if x_0 is such that $x_1, x_2 \in B(x_0, \delta_{x_0})$, then

$$d_Y(f(x_1), f(x_2)) \leq d_Y(f(x_1), f(x_0)) + d(f(x_0), f(x_2)) < \epsilon/2 + \epsilon/2 = \epsilon.$$

This proves that f is uniformly continuous. □

Theorem 2.4.6. Among all n-gons that are inscribed in the unit circle, the regular n-gon has the largest area.

Proof. Let us define $f : (S^1)^n \to [0, \infty)$ by $f(z_1, z_2, \ldots, z_n)$ equal to the area of the convex n-gon with vertices at $z_1, z_2, \ldots, z_n$ (in some order). The function f is continuous, and the set $(S^1)^n$ is closed and bounded in $\mathbb{R}^{2n}$. Hence f has a maximum.

Let us show that an arbitrary n-gon which is not regular does not maximize f. Indeed, such an n-gon would have two adjacent sides AB and BC such that $AB < BC$. If we choose B' the midpoint of the arc $\overset{\frown}{ABC}$, then the area of ABC is strictly smaller than the area of $AB'C$ (because the altitude from B' is greater than the altitude from B). Hence if we replace vertex B by B', then we obtain a polygon with strictly larger area. This proves our claim. It follows that the regular n-gon is the (unique) maximum. □

Example 1. (1984 Balkan Mathematical Olympiad) Let $\alpha_1, \alpha_2, \cdots, \alpha_n$ be positive real numbers, $n \geq 2$, such that $\alpha_1 + \alpha_2 + \cdots + \alpha_n = 1$. Prove that

$$\frac{\alpha_1}{1 + \alpha_2 + \cdots + \alpha_n} + \frac{\alpha_2}{1 + \alpha_1 + \cdots + \alpha_n} + \cdots + \frac{\alpha_n}{1 + \alpha_1 + \cdots + \alpha_{n-1}} \geq \frac{n}{2n-1}.$$

Solution: Rewrite the inequality as

$$\frac{\alpha_1}{2 - \alpha_1} + \frac{\alpha_2}{2 - \alpha_2} + \cdots + \frac{\alpha_n}{2 - \alpha_n} \geq \frac{n}{2n-1},$$

then define the function

$$f(\alpha_1, \alpha_2, \ldots, \alpha_n) = \frac{\alpha_1}{2 - \alpha_1} + \frac{\alpha_2}{2 - \alpha_2} + \cdots + \frac{\alpha_n}{2 - \alpha_n}.$$

As said in the statement, this function is defined on the subset of $\mathbb{R}^n$ consisting of points whose coordinates are positive and add up to 1. We would like to show that on this set f is greater than or equal to $\frac{n}{2n-1}$.

Does f have a minimum? The domain of f is bounded but is not closed, being the interior of a tetrahedron. We can enlarge it, though, adding the boundary, to the set

$$M = \{(\alpha_1, \alpha_2, \ldots, \alpha_n) \,|\, \alpha_1 + \alpha_2 + \cdots + \alpha_n = 1,\, \alpha_i \geq 0, i = 1, 2, \ldots, n\}.$$

We now know that f has a minimum on M.

A look at the original inequality suggests that the minimum is attained when all α_i's are equal. So let us choose a point $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ for which $\alpha_i \neq \alpha_j$ for some indices i, j. Assume that $\alpha_i < \alpha_j$

and let us see what happens if we substitute $\alpha_i + x$ for α_i and $\alpha_j - x$ for α_j, with $0 < x < \alpha_j - \alpha_i$. In the defining expression of f, only the ith and jth terms change. Moreover

$$\frac{\alpha_i}{2-\alpha_i} + \frac{\alpha_j}{2-\alpha_j} - \frac{\alpha_i + x}{2-\alpha_i - x} - \frac{\alpha_j - x}{2-\alpha_j + x}$$
$$= \frac{2x(\alpha_j - \alpha_i - x)(4 - \alpha_i - \alpha_j)}{(2-\alpha_i)(2-\alpha_j)(2-\alpha_i - x)(2 - \alpha_j - x)} > 0,$$

so when moving the numbers closer, the value of f decreases. It follows that the point that we picked was not a minimum. Hence the only possible minimum is $\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ in which case the value of f is $\frac{n}{2n-1}$. This proves the inequality.

Example 2. The proof of the arithmetic mean - geometric mean inequality (AM-GM):

$$x_1 + x_2 + \cdots + x_n \geq n \sqrt[n]{x_1 x_2 \cdots x_n},$$

which holds for any nonnegative numbers $x_1, x_2, \ldots, x_n$.

For the proof, notice that the inequality is homogeneous, meaning that it does not change if we multiply each of the numbers $x_1, x_2, \ldots, x_n$ by the same positive constant. Hence it suffices to prove the inequality for the case where $x_1 + x_2 + \cdots + x_n = 1$. Consider $f(x_1, x_2, \ldots, x_n) = \sqrt[n]{x_1 x_2 \cdots x_n}$ defined on the thetrahedron $x_1 + x_2 + \cdots + x_n = 1$, which is a compact set being closed and bounded. Then f has a maximum. If the x_i are not all equal, say $x_i < x_j$ for some i and j, choose $\epsilon < x_j - x_i$ and replace x_i by $x_i + \epsilon$ and x_j by $x_j - \epsilon$. We are still in the domain of f, and because $(x_i + \epsilon)(x_j - \epsilon) = x_i x_j + \epsilon(x_j - x_i - \epsilon) > x_i x_j$. Hence the maximum is not attained at points where the x_i are not all equal. Consequently, the maximum is attained at $x_1 = x_2 = \cdots = x_n = \frac{1}{n}$, when f equals $\frac{1}{n}$. In this case we have equality in AM-GM, and the inequality is proved.

2.4.3 Compactness of product spaces

Theorem 2.4.7. The product of finitely many compact spaces.

Proof. It suffices to check that the product of two compact spaces is compact. Let X and Y be these spaces. Let $\mathcal{U}$ be an open covering of $X \times Y$. Given $x \in X$, the slice $\{x\} \times Y$ is compact, and can be covered by finitely many elements $U_1, U_2, \ldots, U_n$ of $\mathcal{U}$. Their union is an open set N_x of $X \times Y$, which by Proposition 2.4.2 (5) contains a tube $W_x \times Y$. The open sets W_x, $x \in X$ cover X, and since X is compact, there are $x_1, x_2, \ldots, x_n$ such that $W_{x_1}, W_{x_2}, \ldots, W_{x_n}$ cover X. It follows that $N_{x_1}, N_{x_2}, \ldots, N_{x_n}$ cover $X \times Y$, and consequently the open sets in $\mathcal{U}$ that comprise them form a finite cover of $X \times Y$. □

Theorem 2.4.8. (The Tychonoff theorem) The product of an arbitrary number of compact spaces is compact in the product topology.

Proof. This case is significantly harder, and makes use of the Axiom of Choice in the guise of Zorn's Lemma.

Lemma 2.4.2. If $\mathcal{C}$ is a collection of subsets of a set A having the finite intersection property, then there exists a maximal collection $\mathcal{M}$ of subsets of A that has the finite intersection property and contains $\mathcal{C}$.

Proof. We will apply Zorn's lemma to the set S whose elements are collections of subsets of A with the finite intersection property that contain $\mathcal{C}$.

We will show that if S' is a subset of S that is (totally) ordered under inclusion then S' has a maximal element. We will show that the union $\mathcal{U}$ of all collections in S' is an element of S. If $C_1, C_2, \ldots, C_n$ are finitely many elements of $\mathcal{U}$, then each of them belongs to some $\mathcal{S}_i \in S'$, $i = 1, 2, \ldots, n$. Hence all of them belong the the largest of the $\mathcal{S}_i$, and because this one has the finite intersection property, the sets $C_1, C_2, \ldots, C_n$ have nonempty intersection. Consequently $\mathcal{U}$ has the finite intersection property. This proves the lemma. □

Lemma 2.4.3. Let $\mathcal{M}$ be a collection of subsets of a set A that is maximal with respect to the finite intersection property. Then
(1) Any finite intersection of elements in $\mathcal{M}$ is in $\mathcal{M}$.
(2) If B is a subset of A that intersects every element of $\mathcal{M}$, then B is an element of $\mathcal{M}$.

Proof. (1) If we add to $\mathcal{M}$ the finite intersections of elements in $\mathcal{M}$ we still get a collection of sets that has the finite intersection property. Because of maximality, this collection is $\mathcal{M}$.

(2) Let $\mathcal{M}' = \mathcal{M} \cup \{B\}$. Then $\mathcal{M}'$ still has the finite intersection property. Indeed, if $M_1, M_2, \ldots, M_n$ are in $\mathcal{M}'$, then either none of them is B, in which case their intersection is nonempty, or one of them, say M_n is B. Then

$$M_1 \cap M_2 \cap \cdots \cap M_n = (M_1 \cap M_2 \cap \cdots \cap M_{n-1}) \cap B.$$

By part (a), $M_1 \cap M_2 \cap \cdots \cap M_{n-1} \in \mathcal{M}$, and so by the hypothesis, its intersection with B is not empty. This shows that $M_1, M_2, \ldots, M_n$ have nonempty intersection, and consequently that $\mathcal{M}'$ has the finite intersection property. Because of maximality, $\mathcal{M}' = \mathcal{M}$. Done. □

Let us now proceed with the proof of the theorem. Let

$$X = \prod_{\alpha \in \mathcal{A}} X_\alpha,$$

where each X_α is compact. Let $\mathcal{C}$ be a collection of closed subsets with the finite intersection property. We will show that

$$\cap_{C \in \mathcal{C}} C$$

is nonempty, which will imply that X is compact by Proposition 2.4.1. Choose $\mathcal{M}$ be a maximal family that contains $\mathcal{C}$ and has the finite intersection property. It suffices to show that

$$\cap_{M \in \mathcal{M}} \overline{M}$$

is nonempty. Let $\pi_\alpha : X \to X_\alpha$, $\alpha \in \mathcal{A}$ be the projection. Consider the collections

$$\mathcal{M}_\alpha = \{\pi_\alpha(M) \,|\, M \in \mathcal{M}\}$$

of subsets of X_α. This collection has finite intersection property because $\mathcal{M}$ does. By the compactness of X_α and Proposition 2.4.1, the intersection of the sets $\overline{\pi_\alpha(M)}$, $M \in \mathcal{M}$ is nonempty. Let x_α be a point in the intersection. We let $\mathbf{x} = (x_\alpha)_{\alpha \in \mathcal{A}}$ and show that $\mathbf{x} \in \cap_{M \in \mathcal{M}} \overline{M}$.

To this end, we show that every open set containing x intersects the closure of every set in $\mathcal{M}$. First we verify this for an open set of the form $\pi_\alpha^{-1}(U_\alpha)$, where U_α is open in X_α. Indeed, U_α intersects $\overline{\pi_\alpha(M)}$ and $\pi_\beta(\pi_\alpha^{-1}(U_\alpha)) = X_\beta$ for $\alpha \neq \beta$. Hence this is true for such a set. In particular, by Lemma 2.4.3 (2), each set of the form $\pi_\alpha^{-1}(U_\alpha)$ is in $\mathcal{M}$. By Lemma 2.4.3 (1), intersections of such sets are in $\mathcal{M}$ as well, and these intersections form the basis for the product topology of X. The finite intersection property of $\mathcal{M}$ implies that every basis element that contains $\mathbf{x}$ intersects all $\overline{M}$, $M \in \mathcal{M}$. Hence $\mathbf{x} \in \cap_{M \in \mathcal{M}} M$, which shows that this intersection is non-empty. The theorem is proved. □

2.4.4 Compactness in metric spaces and limit point compactness

There are two other definitions for compactness, which historically precede the one given above.

Definition. A space is said to be *limit point compact* if every infinite subset has a limit point.

Definition. A space is said to be *sequentially compact* if every sequence has a convergent subsequence.

Proposition 2.4.3. Every compact space is limit point compact.

Proof. We argue by contradiction. Let X be a compact subspace that has an infinite subspace Y such that $Y' = \emptyset$. Because $\overline{Y} = Y \cup Y'$, it follows that Y is closed, and consequently compact by Proposition 2.4.2 (2). The fact that $Y' = \emptyset$ implies that every point in Y is contained in some open set that does not contain other points of Y. These open sets form an open cover of Y with no finite subcover (because Y is infinite). This is a contradiction which proves that the original assumption was false. Hence X is limit point compact. □

Example. Let $X = \{0, 1\} \times \mathbb{Z}$, where $\{0, 1\}$ is given the trivial topology and $\mathbb{Z}$ the discrete topology. In X, every set has a limit point, since in every neighborhood of a point $\{t, n\}$ you can find the point $\{1 - t, n\}$. On the other hand, the cover of the space by the open sets $\{0, 1\} \times \{n\}$, $n \in \mathbb{Z}$ has no finite subcover.

The three notions of compactness coincide for metric spaces.

Theorem 2.4.9. Let X be a metric space. The following are equivalent:
(1) X is compact.
(2) X is limit point compact.
(3) X is sequentially compact.

Proof. We have shown above that (1) implies (2).

To see why (2) implies (3), let $(x_n)_{n\geq 1}$ be a sequence and let $A = \{x_n \,|\, n \geq 1\}$. If A is finite, we are done, because there is a constant subsequence. If A is infinite, then it has a limit point, and there is a subsequence that converges to that point.

Let us prove that (3) implies (1). For this we will use the notion of an ϵ-net. By definition, and ϵ-net is a set N_ϵ of points in X such that any point in X is at distance less than ϵ from N_ϵ.

Lemma 2.4.4. If X is sequentially compact then for every ϵ there is a finite ϵ-net.

Proof. Assume that for some ϵ, X does not have a finite ϵ-net. If we start with a point x_1 then there must exist $x_2 \in X \backslash B(x_1, \epsilon)$, $x_3 \in X \backslash (B(x_1, \epsilon) \cup B(x_2, \epsilon))$, etc. The sequence $(x_n)_{n\geq 1}$ does not have a convergent subsequence, because the distance between any two terms is at least ϵ. This contradicts the hypothesis, absurd. Hence the conclusion. □

Returning to the theorem, we mimic the proof of the Heine-Borel theorem in a more general setting. Assume that X is not compact and let $\mathcal{U}$ be an open cover of X with no finite subcover. Consider a finite 1/2-net, $N_{1/2}$. The balls $B(x, 1/2)$ cover X, so one of them, say $B(x_1, 1/2)$ is not covered by finitely many elements of $\mathcal{U}$. Now consider a finite 1/4-net, $N_{1/4}$. Among the balls $B(x, 1/4)$ that intersect $B(x_1, 1/2)$ one must not admit a finite cover by elements of $\mathcal{U}$, or else $B(x_1, 1/2)$ would have a finite cover. Let $B(x_2, 1/4)$ be this ball. Continue the construction to obtain a sequence $x_1, x_2, \ldots, x_n, \ldots$ such that $B(x_n, 1/2^n)$ intersects $B(x_{n+1}, 1/2^{n+1})$ and such that none of the balls $B(x_n, 1/2^n)$ is covered by finitely many elements of $\mathcal{U}$. This sequence has a convergent subsequence; let l be its limit. Because

$$d(x_n, x_{n+1}) < \frac{1}{2^n} + \frac{1}{2^{n+1}} = \frac{3}{2^{n+1}},$$

and therefore

$$d(x_n, x_{n+k}) < 3\left(\frac{1}{2^{n+1}} + \frac{1}{2^{n+2}} + \cdots + \frac{1}{2^{n+k}}\right) < \frac{3}{2^n},$$

it follows that $d(x_n, x_m) \to 0$ as $m, n \to \infty$. This implies that $x_n \to l$, as $n \to \infty$. Note that by passing to the limit as $k \to \infty$ in the inequality $d(x_n, x_{n+k}) < 3/2^n$ we obtain $d(x_n, l) \leq 3/2^n$.

Let U be an element of $\mathcal{U}$ that contains l. Then there is a ball $B(l, \epsilon) \subset U$. If we choose n such that $1/2^{n-2} < \epsilon$, then $B(x_n, 1/2^n) \subset B(l, \epsilon)$. Indeed, if $x \in B(x_n, 1/2^n)$ then by the triangle inequality

$$d(x, l) \leq d(x, x_n) + d(x_n, l) < \frac{1}{2^n} + \frac{3}{2^n} = \frac{1}{2^{n-2}} < \epsilon.$$

Hence $B(x_n, 1/2^n)$ is covered by just one element of $\mathcal{U}$, namely U, a contradiction. We conclude that our initial assumption was false, and therefore X is compact. □

2.4.5 Alexandroff compactification

A compactification of a space X is a compact space Y in which X is dense. The Alexandroff compactification is obtained by adding one point to X, and can be performed on locally compact Hausdorff spaces.

Definition. A space X is said to be *locally compact* if every point has a compact neighborhood.

Example 1. $\mathbb{R}^n$ is locally compact, but clearly not compact.

Theorem 2.4.10. Let X be a locally compact Hausdorff space. Then there exists a compact space Y containing X such that $Y \backslash X$ consists of one point. Moreover, if Y' is a space with the same properties, then Y and Y' are homeomorphic.

Proof. We can let $Y = X \cup \{\infty\}$. We let the open sets of Y be the open sets of X together with the sets of the form $\{\infty\} \cup (X \backslash K) = Y \backslash K$, where K is compact in X. Let us check that this is indeed a topology on Y.

First, Y and $\emptyset$ are clearly open sets (the emptyset in X is compact). Let us check that an arbitrary union of open sets in X and sets of the form $Y \backslash K$ is open in Y. Let U be the union of the open sets in X. Because the intersection of compact sets is compact, the union of the sets that are complements of compact sets is the complement of a compact set. Hence the set we are supposed to check is open is of the form $U \in X$ (if no complements of compact sets are taken), or $Y \backslash K$ (if no open sets in X are taken), or $U \cup\ Y \backslash K$. The first two types are clearly open. Let us check that the third type is open.

Indeed,

$$U \cup (Y \backslash K)) = Y \backslash (K \backslash U)$$

which is of the desired form since $K \backslash U$ is closed in K and therefore compact.

For the intersection of two open sets in Y we have the following possibilities

$$\begin{aligned}
&U_1 \cap U_2 \\
&(Y\backslash K_1) \cap (Y\backslash K_2) = Y\backslash(K_1 \cup K_2) \\
&U \cap (Y\backslash K) = U \cap (X\backslash K).
\end{aligned}$$

All these are open. By induction, the intersection of finitely many open sets is open.

We need to check that Y is compact. Consider an open cover of Y. One of the open sets covers $\{\infty\}$, hence is of the form $Y\backslash K$. The other open sets must cover K, and because K is compact, there is a finite subcover. Add to this the open set that covers $\{\infty\}$ to obtain a finite subcover of Y.

Let us show that X is a subspace of Y (in the sense that the topology on X is induced by the topology on Y). Indeed, for every open subset U of X, $U \cap X = U$ is open in X. Also, if K is compact in X, then $(Y\backslash K) \cap X = X\backslash K$, and since K is closed, this is open in X.

Let us prove that Y is Hausdorff. We only have to check that any point $x \in X$ can be separated from ∞. Choose a compact neighborhood K of X. Then $V_1 = \mathrm{Int}(K)$ and $V_2 = X\backslash K$ are disjoint open sets such that $x \in V_1$ and $\infty \in V_2$.

Finally, we check uniqueness. If $Y' = X \cup \{\infty'\}$ is another space with the same property, define $h : Y \to Y'$, $h(x) = x$ if $x \in X$ and $h(\infty) = \infty'$.

Then h is bijective. By Theorem 2.4.3, it suffices to check that h is continuous. Let U be an open set in Y'. If $U \subset X$, then $h^{-1}(U) = U$, open. If $\infty' \in U$, then $U = U_0 \cup \{\infty'\}$ where U_0 is open in X. Note that because Y' is compact, the complement of U_0 in X, which is the same as the complement of U in Y, is compact. We thus have $h^{-1}(U) = U_0 \cup \{\infty\}$, which is of the form $Y\backslash K$, with K compact in X. This set is open, and hence h is continuous. The theorem is proved. $\square$

Definition. The space Y defined by this theorem is called the *Alexandroff (or one-point) compactification* of X.

It is important to note that $\overline{X} = Y$ only when X is not already compact.

Example 2. The one point compactification of the plane $\mathbb{R}^2$ is the sphere S^2. We stress out that there are other possible compactifications (compacs spaces in which the original space is dense), for example the projective plane is a compactification of $\mathbb{R}^2$ as well.

Chapter 3

Separation Axioms

3.1 The countability axioms

Definition. A space is called *separable* if it contains a countable subset that is dense.

Example 1. $\mathbb{R}$ is separable because $\mathbb{Q}$ is dense in $\mathbb{R}$.

Definition. A space is called *first-countable* if each point has a countable system of neighborhoods.

Definition. A space is called *second-countable* if its topology has a countable basis.

Example 2. $\mathbb{R}$ is second-countable because (a, b), $a, b \in \mathbb{Q}$ is a countable basis for the standard topology.

Example 3. If we endow $\mathbb{R}$ with the topology defined by the basis $[a, b)$, $a, b \in \mathbb{R}$, then it is not second-countable. Indeed, if we write $[a, b)$ as a union of basis elements, then one of these basis elements has its minimum equal to a. Thus there are at least as many basis elements as there are real numbers (the function $B \to \inf B$ is onto), and hence there is no countable basis.

Example 4. The disjoint union of an uncountable family of copies of $\mathbb{R}$ with the standard topology is first-countable but not second-countable.

Proposition 3.1.1. If X is second-countable then it is separable and first-countable.

Proof. Choose one point in each basis element (different basis elements may have the same chosen point). The set containing these points is countable and dense.

The system of neighborhoods consisting of all basis elements that contain the given point is countable. □

3.2 Regular spaces

Definition. A topological space X is said to be *regular* if for every point x and every closed set C there are disjoint open set U and V such that $x \in U$ and $C \subset V$.

Proposition 3.2.1. (i) A space X is regular if and only if given a point x and a neighborhood U of x, there is a neighborhood V of x such that $\overline{V} \subset U$.
(ii) A subspace of a regular space is regular.
(iii) A product of regular spaces is regular.

Proof. (i) If the space is regular, let x be the point and C the closed set which is the complement of U. Then there are open sets V and W such that $x \in V$, $C \subset W$. Then $x \in V \subset \overline{V} \subset U$, as desired.

Conversely, let x be a point and C a closed set that does not contain x. Then for $U = X \backslash C$ there is an open neighborhood V of x such that $\overline{V} \subset U$. The sets V and $W = X \backslash \overline{V}$ satisfy the condition from the definition of a regular space.

(ii) Let Y be a subspace of a regular space X, $x \in Y$ and C a closed subset of Y which does not contain x. We know that $C = C' \cap Y$, where C' is closed in X. Clearly C' does not contain x. So there are disjoint open subsets of X, U and V, such that $x \in U$, $C' \subset V$. The open subsets of Y $U \cap Y$ and $V \cap Y$ separate x from C.

(iii) Let X_α, $\alpha \in \mathcal{A}$ be a family of regular spaces. We use (i) to check the regularity of $\prod_\alpha X_\alpha$ in the product topology. Let $\mathbf{x} = (x_\alpha)_{\alpha \in \mathcal{A}}$ and U an open neighborhood of this point. Without loss of generality we may assume $U = \prod_\alpha U_\alpha$ (or else we pass to a smaller neighborhood). Here $U_\alpha = X_\alpha$ for all but finitely many α.

For every α choose an open neighborhood V_α of x_α such that $\overline{V_\alpha} \subset U_\alpha$, with the condition that we choose $V_\alpha = X_\alpha$ whenever $U_\alpha = X_\alpha$. Then $V = \prod_\alpha V_\alpha$ is a neighborhood of $\mathbf{x}$. Since by Proposition 2.1.4 (3) $\overline{V} = \prod_\alpha \overline{V_\alpha}$, it follows that $\overline{V} \subset U$, and we are done. □

3.3 Normal spaces

3.3.1 Properties of normal spaces

Definition. A topological space X is said to be normal if for every disjoint closed sets C_1 and C_2 there are disjoint open sets U_1 and U_2 such that $C_1 \subset U_1$ and $C_2 \subset U_2$.

Proposition 3.3.1. (i) A space X is normal if and only if given a closed set C and an open set U containing C, there is an open set V containing C such that $\overline{V} \subset U$.
(ii) A compact Hausdorff space is normal.

Proof. (i) Suppose X is normal. Let $C' = X \backslash U$, which is a closed set. By hypothesis there are disjoint open sets V and V' containing C respectively C'. Then $\overline{V}$ is disjoint from C' (since C' lies in an open set disjoint from V) and therefore $\overline{V} \subset U = X \backslash C'$, as desired.

For the converse, let C_1 and C_2 be disjoint closed sets. Let $U = X \backslash C_1$, and consider the open set V such that $C_1 \subset V \subset \overline{V} \subset U$. The disjoint open sets V and $X \backslash \overline{V}$ separate C_1 and C_2.

(ii) Let C_1 and C_2 be disjoint closed sets in X, which are therefore compact. For every pair of points x and y in C_1 respectively C_2, there are disjoint open sets $U_{x,y}$ respectively $V_{x,y}$ such that $x \in U_{x,y}$ and $y \in V_{x,y}$. For a fixed y, the open sets $U_{x,y}$, $x \in C_1$ cover C_1, and because C_1 is compact there is a finite cover $U_{x_1,y}, U_{x_2,y}, \ldots, U_{x_n,y}$. Define $U_y = U_{x_1,y} \cup U_{x_2,y} \cup \cdots U_{x_n,y}$ and $V_y = V_{x_1,y} \cap V_{x_2,y} \cap \cdots V_{x_n,y}$. Then the open sets U_y and V_y are disjoint and separate C_1 from y. If we vary y, the sets V_y cover C_2. There is a therefore a finite cover $V_{y_1}, V_{y_2}, \ldots V_{y_n}$. Define $U = U_{y_1} \cap U_{y_2} \cap \cdots U_{y_n}$ and $V = V_{y_1} \cup V_{y_2} \cup \cdots \cup V_{y_n}$. Then U and V are disjoint and separate C_1 and C_2. This completes the proof. □

Proposition 3.3.2. Every regular space with countable basis is normal.

Proof. Let X be the space and let C_1 and C_2 be two disjoint closed sets. By regularity, each $x \in C_1$ has a neighborhood U that does not intersect C_2. Choose an open neighborhood V of x such that $\overline{V} \subset U$ (which exists by Proposition 3.2.1). There exists a basis element B such that $B \subset V$ and hence $\overline{B} \subset U$ and in particular $\overline{B}$ does not intersect C_2.

Since the basis is countable, there basis elements $B_1, B_2, B_3, \ldots$ that cover C_1 and whose closures do not intersect C_2. Similarly choose basis elements $B_1', B_2', B_3', \ldots$ that cover C_2 and whose closures do not intersect C_1. Naively we can try $U = \cup B_n$ and $U' = \cup B_n'$, but these need not be disjoint.

Instead let

$$V_n = B_n \setminus \cup_{i=1}^n \overline{B_i'}, \quad V_n' = B_n' \setminus \cup_{i=1}^n \overline{B_i}.$$

Note that V_m and V_n' are disjoint for every m and n. Hence $V = \cup V_n$ and $V' = \cup V'n$ are disjoint as well. These two sets separate C_1 from C_2, and we are done. □

To rephrase, every second-countable regular space is normal. Normal spaces avoid many of the pathologies of the general topological spaces, and are the most common spaces that we encounter in mathematics.

Example 1. Every metric space is normal. Indeed, let A and B be two disjoint closed subsets of X. The sets

$$U = \{x \mid d(x, A) < d(x, B)\} \text{ and } V = \{x \mid d(x, A) > d(x, B)\}$$

are open because $x \to d(x, A)$ is continuous (see proof of Lebesque number theorem). They are disjoint, and $A \subset U$, $B \subset V$.

Example 2. Every second-countable manifold is normal. In view of Proposition 3.3.2, we only have to check that it is regular. But Proposition 3.2.1 (i) shows that regularity is a local property, and clearly $\mathbb{R}^n$ with the standard topology is regular, even normal, as seen in Example 1.

We now prove two fundamental properties of normal spaces.

3.3.2 Urysohn's lemma

Theorem 3.3.1. (Urysohn's lemma) Given a normal space X and disjoint closed subsets C_0 and C_1, there is a continuous function $f : X \to [0, 1]$ such that $f(C_0) = \{0\}$ and $f(C_1) = \{1\}$.

Proof. Here is one way of writing the proof: For every dyadic number $r \in [0, 1]$ (meaning that r is of the form $m/2^n$ with $m, n \in \mathbb{Z}_+$) we will construct an open set U_r such that
(i) U_r contains C_0 and is disjoint from C_1,
(ii) if $r < s$ then $\overline{U_r} \subset U_s$.

Define the function $f : X \to [0, 1]$, $f(x) = \inf\{r \mid x \in U_r\}$. To prove that it is continuous, it suffices to show that for every $a, b \in (0, 1)$, $f^{-1}([0, b))$ and $f^{-1}((a, 1])$ are open, because these sets form a subbasis of the standard topology.

We have

$$f^{-1}([0, b)) = \{x \mid x \in U_r \text{ for some } r < b\} = \cup_{r<b} U_r$$

and this is open. Also,

$$f^{-1}((a, 1]) = \{x \mid \text{there is } s > a \text{ such that } x \notin U_s\}.$$

Because numbers of the form $m/2^n$ are dense, and because of the way the sets U_s were constructed, this is the same as

$$\{x \mid \text{there is } r > a \text{ such that } x \notin \overline{U_r}\} = \cup_{r>a}(X \setminus \overline{U_r})$$

which is open, being the union of open sets. This proves continuity.

To show the existence of the sets U_r, we prove a stronger fact, namely that for every r there are disjoint open sets U_r and V_r that satisfy the following nesting conditions:
(a) $C_0 \subset U_r$, $C_1 \subset V_r$;
(b) for $r < s$ the complement of V_r is contained in U_s.

We proceed by induction on n. The disjoint open sets $U_{1/2}$ and $V_{1/2}$ such that $C_1 \in U_{1/2}$, $C_2 \in V_{1/2}$ exist because X is normal. Now assume that the sets $U_{m/2^k}$ and $V_{m/2^k}$ were chosen for $k \leq n$ and $m = 1, \ldots, 2^k - 1$. We want to construct $U_{m/2^{n+1}}$ and $V_{m/2^{n+1}}$ for $m = 0, 1, \ldots, 2^{n+1} - 1$. This was already done for m even, since $2l/2^{n+1} = l/2^n$. Thus we have to do it for m odd, say $m = 2l + 1$.

If $l \neq 0$ or $l \neq 2^n - 1$, then, because the space is normal, we can find disjoint open sets U and V that separate the complement of $V_{l/2^n}$ from the complement of $U_{(l+1)/2^n}$. Let these be $U_{(2l+1)/2^{n+1}}$ and $V_{(2l+1)/2^{n+1}}$. Then $V_{l/2^n} \subset U_{(2l+1)/2^{n+1}}$ and $V_{(2l+1)/2^{n+1}} \subset U_{(l+1)/2^n}$, which shows that the new sets satisfy the required nesting conditions with their neighbours. Because all other sets satisfy the nesting conditions (b) and (c) they will satisfy these with the newly constructed sets.

If $l = 0$, we replace in this construction the complement of $V_{l/2^n}$ by C_0, and if $l = 2^n - 1$, we replace in this construction $U_{2l+1}/2^n$ by C_1. Again, the nesting conditions are satisfied. The sets U_r constructed this way satisfy the desired properties, and the result is proved.

Here is a second way of writing the proof: Order the rational numbers in the interval $(0, 1]$ as $r_0 = 0, r_1 = 1, r_2, r_3, r_4, \ldots$. Using Proposition 3.3.1, we will construct a family of open sets U_r, $r \in \mathbb{Q} \cap (0, 1]$ such that for $r < s$ $C_0 \subset U_r \subset \overline{U_r} \subset U_s \subset X \backslash C_1$. Choose U_0 and U_1 to separate C_0 from C_1. Assume that $U_{r_1}, U_{r_2}, \ldots, U_{r_k}$ have been chosen and let us construct $U_{r_{k+1}}$. Suppose $r_m = \max\{r_l \,|\, r_l < r_{k+1}, 1 \leq l \leq k\}$, $r_n = \max\{r_l \,|\, r_l < r_{k+1}, 1 \leq l \leq k\}$. Let $U_{r_{k+1}}$ to be an open set such that

$$\overline{U_{r_m}} \subset U_{r_{k+1}} \subset \overline{U_{r_{k+1}}} \subset U_{r_n}.$$

Continue inductively and you obtain the desired result.

From here continue like in the previous proof: Define $f : X \to [0, 1]$, $f(x) = \inf\{r \,|\, x \in U_r\}$, etc. $\square$

3.3.3 The Tietze extension theorem

As a corollary of Urysohn's lemma we have the following result.

Theorem 3.3.2. (Tietze extension theorem) Let C be a closed subspace of a normal space X. Then any continuous map $f : C \to [-1, 1]$ can be extended to a continuous map $\tilde{f} : X \to [-1, 1]$.

Proof. We start by proving the following

Lemma 3.3.1. If X is normal and C is closed in X, then for any continuous function $f : C \to [-1, 1]$, there is a continuous function $g : X \to [-1/3, 1/3]$ such that $|f(x) - g(x)| \leq 2/3$ for all $x \in C$.

Proof. The sets $C_1 = f^{-1}(-\infty, -1/3])$ and $C_2 = f^{-1}([1/3, \infty))$ are disjoint and closed in A. By Urysohn's lemma, there is a continuous function $g : X \to [-1/3, 1/3]$ such that $g|C_1 = -1/3$ and $g|C_2 = 1/3$. So on $C \cap (C_1 \cup C_2)$, $|f(x) - g(x)| \leq 1 - 1/3 = 2/3$, and on $C \backslash (C_1 \cup C_2)$, $|f(x) - g(x)| \leq |f(x)| + |g(x)| \leq 1/3 + 1/3 = 2/3$. $\square$

Let us return to the proof of the theorem. We will construct a sequence of continuous functions that approximate f on C. To this end we use the lemma to construct $g_0 : X \to [-1/3, 1/3]$, so that

$|f(x)-g_0(x)| \leq 2/3$ for $x \in C$. Now we apply the lemma to the function $(f-g_0) : C \to [-2/3, 2/3]$. By rescaling, we conclude that there is $g_1 : C \to [-2/9, 2/9]$ such that $|f(x) - g_0(x) - g_1(x)| \leq 4/9$. Repeating we obtain a sequence of continuous functions $g_0, g_1, g_2, \ldots$, with the property that $|g_n(x)| \leq 2^n/3^{n+1}$ and

$$|f(x) - g_0(x) - g_1(x) - \cdots g_n(x)| \leq 2^{n+1}/3^{n+1} \text{ for } x \in C.$$

The series of functions $\sum_n g_n$ converges absolutely and uniformly to a function $\tilde{f} : X \to [-1, 1]$, and this function has the property that $\tilde{f}(x) = f(x)$ for all $x \in C$. The theorem is proved. □

Here is an application of Tietze's theorem.

Theorem 3.3.3. (Peano) There is a continuous map $f : [0, 1] \to [0, 1] \times [0, 1]$ that is onto.

Proof. Let C be the Cantor set defined in §2.1.1. Define the function $f : C \to [0, 1] \times [0, 1]$, as follows. A number $x \in C$ has a unique ternary representation using only 0 and 2 as $0.a_1a_2a_3\ldots$. Let $b_n = a_n/2$; as such $b_n = 0$ or 1. We let $f(x) = (0.b_1b_3b_5\ldots, 0.b_2b_4b_6\ldots)$, where the expansions are considered in base 2. It is easy to check that convergence in C implies convergence digit-by-digit, so f is continuous. The function f is also onto, because we can construct the number x from the digits $b_1, b_2, b_3, \ldots$ by doubling them and then reading in base 3. By Tietze's theorem f can be extended to the entire interval $[0, 1]$ and we are done. □

Part II

Algebraic topology

Chapter 4

Homotopy theory

4.1 Basic notions in category theory

To explain the main idea behind algebraic topology we need to introduce a new language.

Definition. A category $\mathcal{C}$ consists of
(a) a class of objects $\mathrm{Ob}(\mathcal{C})$ and
(b) a class of morphisms $\mathrm{Hom}(\mathcal{C})$, where each morphism has a source object A and a target object B, and is usually written as $f : A \to B$.
If the source of one morphism is the target of another, the two can be composed, that is if $f : A \to B$ and $g : B \to C$, then there is a morphism $g \circ f : A \to C$. The following axioms should hold:
(associativity) if $f : A \to B, g : B \to C, h : C \to D$, then $h \circ (g \circ f) = (h \circ G) \circ f$.
(identity) for every object X, there is a morphism $1_X : X \to X$, called the identity morphism, such that for every morphisms $f : A \to X$ and $g : X \to B$, $1_X \circ f = f$ and $g \circ 1_X = g$.

Examples. The cateory of sets (objects=sets, morphisms=functions), the category of groups (objects=groups, morphisms=group homomorphisms), the category of abelian groups, the cateory of rings, the category of vector spaces (objects=vector spaces, morphisms=linear transformations), the category of topological spaces (objects=topological spaces, morphisms=continuous maps), the category of differentiable manifolds (objects=manifolds, morphisms=differentiable maps).

Definition. Let $\mathcal{C}_1$, and $\mathcal{C}_2$ be two categories. A functor F from $\mathcal{C}$ to $\mathcal{D}$ associates to
(a) each object X in $\mathcal{C}_1$ and object $F(X)$ in $\mathcal{C}_2$, (b) each morphism $f : X \to Y$ in $\mathcal{C}_1$ a morphism $F(f) : F(X) \to F(Y)$ in $\mathcal{C}_2$ such that
(i) $F(1_X) = 1_{F(X)}$
(ii) $F(g \circ f) = F(g) \circ F(f)$ if the functor is covariant, or $F(g \circ f) = F(f) \circ F(g)$ if the functor is contravariant.

Example. The forgetful functor from the category of groups to the category of sets which replaces each group with the underlying set and views group homomorphisms simply as functions.

The main idea of algebraic topology is to associate to construct functors from the category of topological spaces to various categories that arise in algebra: the category of groups, the category of rings, the category of vector spaces.

4.2 Homotopy and the fundamental group

4.2.1 The notion of homotopy

The idea of homotopy is to model more complicated types of connectedness, and to find ways of distinuishing topological spaces. topological spaces. Throughout this chapter we will work with locally path-connected spaces, and most of the time with path-connected spaces.

Examples. Domains in $\mathbb{R}^n$.

Recall the way path components are defined. We say that $x \sim y$ in X if there is a path $f : [0,1] \to X$ such that $f(0) = x$ and $f(1) = y$.

Here is a different way to state this, which is generalizable to other types of "connectivity". The point x can be thought of as a function $f_x : \{0\} \to X$. The fact that $x \sim y$ (i.e. x and y are in the same path component) means that there is a continuous function $F : \{0\} \times [0,1] \to X$ such that $F|(0,0) = f_x$ and $F|(0,1) = f_y$. The set of path components is usually denoted by $\pi_0(X)$.

Now let us assume that X is *path-connected.* We want to look at other types of connectivity, which record holes in the space. Here are some examples that explain why this is useful.

Example 1. Consider the differential equation

$$udx + vdy = 0.$$

To solve this equation by integration, the differential form on the left should satisfy

$$\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x}.$$

If this condition holds, then we can find a function f such that $df = udx + vdy$, and the equation has the implicit solution f =constant. Well, not quite. The function f exists if the domain of definition of the form has no "holes", but otherwise f might not exist.

For example, for

$$-\frac{y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy = 0,$$

we can ideed find $f = \arctan \frac{y}{x}$, but... there is a problem when we cross the y-axis. There, f becomes discontinuous.

Example 2. If $f(z)$ is a holomorphic function on a domain D in the plane, and the loop Γ can be shrunk to a point (in the domain), then

$$\int_\Gamma f(z)dz = 0.$$

However, this is not necessarily true if Γ cannot be shrunk to a point. For example

$$\int_{S^1} \frac{1}{z} dz = 2\pi i$$

for the function $1/z$ which is defined in $\mathbb{C}\backslash\{0\}$.

Definition. Two continuous maps $f, g : X \to Y$ are called *homotopic* if there is a continuous map $H : X \times [0,1] \to Y$ such that $H|X \times \{0\} = f$ and $H|X \times \{1\} = g$.

Notation: $f \sim g$.

In other words, two continuous maps $f_0, f_1 : X \to Y$ are homotopic if there is a (continuous) path of continuous maps $f_t : X \to Y$ that starts at f_0 and ends at f_1. If X is just one point, then the homotopy is just a path, as we explained before.

Homotopy comes with a natural notion of equivalence between topological spaces.

Definition. Two topological spaces X and Y are called homotopically equivalent if there are maps $f : X \to Y$ and $g : Y \to X$ such that $g \circ f$ is homotopic to id_X and $f \circ g$ is homotopic to id_Y.

Notice that being homotopically equivalent is a weaker condition than being homeomorphic.

Definition. Let A be a subset of the topological space X and let $f, g : X \to Y$ be two continuous maps such that $f = g$ on A. We say that f and g are homotopic relative to A if there is a continuous map $H : X \times [0,1] \to Y$ such that $H|X \times \{0\} = f$, $H|X \times \{1\} = g$ and $H|A \times [0,1] = f|A = g|A$.

Notation: $f \sim_A g$. If $X = [0,1]$, that is if f and g are paths, we denote the homotopic equivalence of f and g relative to the endpoints by $f \sim_p g$.

This notion of relative equivalence tells us to consider continuous maps between pairs of topological spaces in the following sense: if $A \subset X$ and $B \subset Y$ are topological spaces, then a continuous map $f : (X, A) \to (Y, B)$ is a continuous map $f : X \to Y$ such that $f(A) \subset B$.

Proposition 4.2.1. The relations $\sim$ and $\sim_p$ are equivalence relations.

Proof. Let us show that $\sim$ is an equivalence relation. We have $f \sim f$ since $H(x,s) = f(x)$ is a homotopy between f and f. Also, if $H(x,s)$ is a homotopy between f and g, then $H'(x,s) = H(x, 1-s)$ is a homotopy between g and f. Hence $f \sim g$ implies $g \sim f$. Finally, if $f \sim g$ and $g \sim h$, with homotopies $H(x,s)$ and $H'(x,s)$, then

$$H''(x,s) = \begin{cases} H(x,2s), & s \in [0,1/2] \\ H'(x,2s-1) & s \in [1/2,1] \end{cases}$$

is a homotopy between f and h, so $f \sim h$. □

Notation: We denote by $[f]$ the equivalence class of f.

Definition. Let f be a path from x_0 to x_1 and g a path from x_1 to x_2. Then the *product* of f and g is

$$f * g = \begin{cases} f(2t), & t \in [0,1/2] \\ g(2t-1), & t \in [1/2,1]. \end{cases}$$

Proposition 4.2.2. The product of paths factors to a product of homotopy equivalence classes of paths relative to the endpoints.

Proof. It suffices to notice that if $f \sim_p f'$ and $g \sim_p g'$, with homotopies relative to the endpoints $H(t,s)$ respectively $H'(t,s)$, then

$$H''(t,s) = \begin{cases} H(2t,s), & t \in [0,1/2] \\ H'(2t-1,s) & t \in [1/2,1] \end{cases}$$

is a homotopy between $f * g$ and $f' * g'$. □

Definition. A groupoid is a category whose objects form a set and all of whose morphisms are invertible.

Theorem 4.2.1. The set of equivalence classes of paths relative to the endpoints is a groupoid, meaning that the operation $*$ has the following properties:
(1) (associativity) if $f(1) = g(0)$ and $g(1) = h(0)$ then

$$[f] * ([g] * [h]) = ([f] * [g]) * [h].$$

(2) (left and right identities) Given $x \in X$, define $e_x : I \to X$, $e_x(t) = x$. Then

$$[f] * [e_{f(1)}] = [f] = [e_{f(0)}] * [f].$$

(3) Given $[f]$ there is a path $[\bar{f}]$ such that

$$[f] * [\bar{f}] = [e_{f(0)}], \quad \text{and } [\bar{f}] * [f] = [e_{f(1)}].$$

Proof. (1) A homotopy between $f * (g * h)$ and $(f * g) * h$ is given by

$$H(t,s) = \begin{cases} f\left((2s+2)t\right) & t \in [0, \frac{1}{2} - \frac{s}{4}] \\ g\left(4t - 2 + s\right) & t \in [\frac{1}{2} - \frac{s}{4}, \frac{3}{4} - \frac{s}{4}] \\ h\left((4-2s)t - 3 + 2s\right) & t \in [\frac{3}{4} - \frac{s}{4}, 1] \end{cases}$$

(2) A homotopy between $e_{f(0)} * f$ and f is given by

$$H(t,s) = \begin{cases} f(0) & t \in [0, \frac{1}{2} - \frac{1}{2}s] \\ f\left(\frac{2}{2-s}t - \frac{s}{2-s}\right) & t \in [\frac{1}{2} - \frac{1}{2}s, 1] \end{cases}$$

(3) A homotopy between $e_{f(0)}$ and $f * \bar{f}$ is given by

$$H(t,s) = \begin{cases} f(2st) & t \in [0, \frac{1}{2}] \\ f(2s(1-t)) & t \in [\frac{1}{2}, 1] \end{cases}$$

These three constructions can be schematically represented as in Figure 4.1, with $x_0 = f(0)$. □

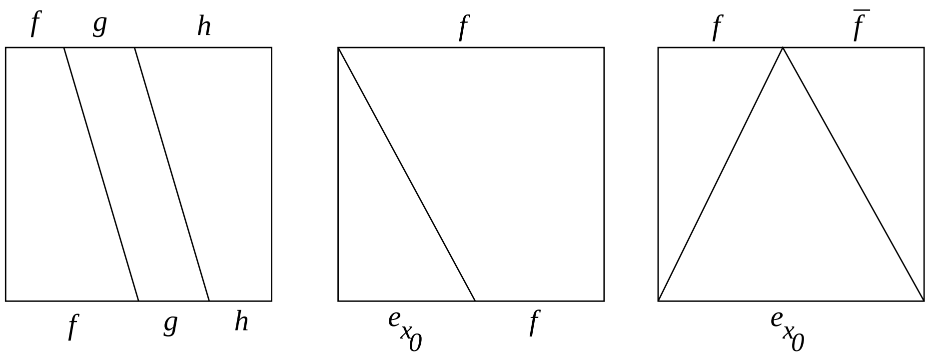

Figure 4.1:

4.2.2 The definition and properties of the fundamental group

The central point of this chapter is the construction a functor from the category of path connected topological spaces with one specified point to the category of groups.

Definition. A path whose endpoints coincide is called a *loop*. More precisely, if $f : [0,1] \to X$ has the property that $f(0) = f(1) = x_0$, then f is called a loop based at x_0.

Alternatively, a loop based at x_0 can be identified with a map $f : S^1 \to X$, with $f(1) = x_0$. As such it is a continuous map $(S^1, 1) \to (X, x_0)$. In literature the set of loops is denoted by $\Omega(X, x_0)$. The point x_0 is refered to as the *base point*.

From this moment on we fix the base point x_0 and look at all loops based at x_0. Recall the equivalence relation of loops relative to the base point: two loops are called homotopic (relative to the base point) if there is a continuous family of loops (with the same basepoint) that interpolate between the two. As a corollary of Theorem 4.2.1 we obtain that the set of homotopy equivalence classes of loops based at x_0, denoted by $\pi_1(X, x_0)$, is a group.

Definition. The group $\pi_1(X, x_0)$ is called the *fundamental group* (or the first homotopy group) of X at x_0.

Remark 4.2.1. In the same way one can define the higher homotopy groups. One considers again homotopy equivalence classes of continuous maps $f : (S^n, p) \to (X, x_0)$, where $p = (1, 0, 0, \ldots, 0)$. Then one defines the composition of f and g to be obtained by pinching the sphere along the equator and thus obtaining two spheres, putting f on the upper sphere and g on the lower sphere. This is an abelan group denoted by $\pi_n(X, x_0)$ (the *nth homotopy group*). It is important to know that the computation of all homotopy groups of spheres is one of the hardest problems in today's mathematics.

Theorem 4.2.2. Let X be a path connected space and x_0 and x_1 two points in X. Then $\pi_1(X, x_0)$ is isomorphic to $\pi_1(X, x_1)$.

Proof. Let γ be a path from x_0 to x_1. Then the isomorphism between $\pi_1(X, x_0)$ and $\pi_1(X, x_1)$ is given by

$$[f] \mapsto [\bar{\gamma}] * [f] * [\gamma].$$

The fact that this is a group homomorphism follows from

$$\bar{\gamma} * f * \gamma * \bar{\gamma} * g * \gamma \sim_p \bar{\gamma} * f * e_{x_0} * g * \gamma \sim_p \bar{\gamma} * f * g * \gamma.$$

It is invertible, with the inverse given by $[f] \mapsto [\gamma] * [f] * [\bar{\gamma}]$. $\square$

Definition. A path connected space is called *simply connected* if its fundamental group is trivial.

Definition. The loops in the equivalence class of e_{x_0} are called *null-homotopic*.

So a space is simply connected if and only if all loops are null-homotopic.

Example 1. Any convex set in $\mathbb{R}^n$ is simply connected. In particular $\mathbb{R}^n$ is simply connected. Indeed, let $f : S^1 \to \mathbb{R}^n$ is a loop in a convex set C, and let x_0 be the base point. Then $H(t, s) = (1 - s)f(t) + sx_0$ is a homotopy between f and the trivial loop based at x_0.

Theorem 4.2.3. S^n is simply connected for $n \geq 2$.

Proof. We will do the proof only in the case $n = 2$, where it is easier to phrase. The other cases are analogous.

Let $f : (S^1, 1) \to (S^2, x_0)$ be a loop based at x_0. We want to shrink the loop to a point by pushing it away from a point that does not belong to it. In view of Theorem 3.3.3, f could pass through every single point of S^2. To fix this problem, we approximate it by a piece-wise linear loop homotopic to it.

To construct this piece-wise linear loop, we use the fact that S^1 is compact, which implies, by Corollary 2.4.1, that f is uniformly continuous. So for every ϵ there is a positive integer n such that the image of any interval of length $1/n$ lies inside a ball of radius ϵ. Choose $\epsilon < 1/4$, divide $[0, 1]$ into n intervals of length $1/n$ and define a loop $g : [0, 1] \to S^2$ such that $g|[\frac{k}{n}, \frac{k+1}{n}]$ is the (shortest) arc of a great circle connecting $f(\frac{k}{n})$ with $f(\frac{k+1}{n})$. Because ϵ was chosen small enough, for each t there is a unique arc of a great circle from $f(t)$ to $g(t)$. Parametrize this arc by its arc-length, then rescale, to make it $\phi_t : [0, 1] \to S^2$ (with the variable of ϕ_t being s).

Define $H : [0, 1] \times [0, 1] \to S^2$, $H(t, s) = \phi_t(s)$. It is not hard to check that H is continuous, because close-by points yield close-by arcs of great circles. Then H is a homotopy from f to g. So the loop g is homotopic to f relative to the endpoint.

Now we will show that g is null homotopic. Let $N \in S^2$ be a point that does not belong to the image of g (such a point exists because the image of g is a finite union of arcs, hence has area zero on the sphere). Then $g : (S^1, 1) \to (S^2\backslash N, x_0)$. We know that $S^2\backslash N$ is homeomorphic to $\mathbb{R}^2$; let h be the homeomorphism. The loop $h \circ g : (S^1, 1) \to (\mathbb{R}^2, h(x_0))$ is null homotopic because $\mathbb{R}$ is simply connected. Let H' be the homotopy relative to the endpoints from $h \circ g$ and $e_{h(x_0)}$. Then $H''(t, s) = h^{-1}(H'(t, s))$ is a homotopy relative to the endpoints from g to e_{x_0}. Hence g is null homotopic, and so is f. $\square$

4.2.3 The behavior of the fundamental group under continuous transformations

Up to this point we have explained how to associate to each path-connected topological space X with base-point x_0 a group $\pi_1(X, x_0)$. To complete the definition of the functor π_1, we have to show how it associates to continuous maps group homomorphisms.

Definition. Let $g : (X, x_0) \to (Y, y_0)$ be a continuous map. Define $g_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ by $g_*([f]) = [g \circ f]$.

Proposition 4.2.3. (a) The map g_* is well defined.
(b) If $g : (X, x_0) \to (Y, y_0)$ and $h : (Y, y_0) \to (Z, z_0)$ are continuous, then $(h \circ g)_* = h_* \circ g_*$.
(c) If 1_X is the identity map of X, then $(1_X)_*$ is the identity group homomorphism.

Proof. The fact that g_* is well defined follows from the fact that if H is a homotopy relative to the base point between the loops f and f' based at x_0, then $g \circ H$ is a homotopy relative to the basepoint between the loops $g \circ f$ and $g \circ f'$.

For (b), we have $(h \circ g) \circ f = h \circ (g \circ f)$ so

$$(h \circ g)_*([f]) = h_*([g \circ f]) = (h_* \circ g_*)([f]).$$

Finally, (c) is obvious. $\square$

Now we know indeed that we have defined a functor

- (X, x_0) topological space with base-point $\mapsto \pi_1(X, x_0)$ group;
- $f : (X, x_0) \to (Y, y_0)$ continuous map $\mapsto f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ group homomorphism.

Proposition 4.2.4. If $g, g' : (X, x_0) \to (Y, y_0)$ are homotopic (relative to x_0), then $g_* = g'_*$.

Proof. This follows from the fact that for every loop f based at x_0, $g \circ f$ and $g' \circ f$ are homotopic relative to the basepoint. □

Theorem 4.2.4. If X and Y are homotopy equivalent, then $\pi_1(X)$ is isomorphic to $\pi_1(Y)$.

Proof. Let $g : X \to Y$ and $h : Y \to X$ be such that $h \circ g \sim 1_X$ and $g \circ h \sim 1_Y$. We will show that g_* is an isomorphism. We will apply the following

Lemma 4.2.1. Let $g : (X, x_0) \to (X, x_1)$ be a map homotopic to the identity. Then $g_* : \pi_1(X, x_0) \to (X, x_1)$ is a group isomorphism.

Proof. Let $H : X \times [0, 1] \to X$ be the homotopy between 1_X and g. Then for a loop f based at x_0 $H'(t, s) = H(f(t), s)$ is a homotopy between f and $g \circ f$, but they are not homotopic relative to the base point (they don't even have the same base point). To fix this problem, let $\gamma(s) = H(x_0, s) = H'(0, s)$ be the path from x_0 to x_1 traveled by x_0 during the homotopy. We will show that $g \circ f \sim_p \bar{\gamma} * f * \gamma$.

The homotopy is

$$H''(t,s) = \begin{cases} \bar{\gamma}(3st) & t \in \left[0, \frac{1}{3}\right] \\ H'(3t-1, 1-s) & t \in \left[\frac{1}{3}, \frac{2}{3}\right] \\ \gamma\left(3st + (1-3s)\right) & t \in \left[\frac{2}{3}, 1\right]. \end{cases}$$

□

Using the lemma we obtain that $g_* \circ h_*$ and $h_* \circ g_*$ are group isomorphisms. Hence g_* is both one-to-one and onto, which implies that it is an isomorphism. This proves the theorem. □

As a consequence of the theorem, the fundamental group defines a functor from the category of homotopy equivalence classes of path-connected topological spaces to the category of groups. So from the perspective of the fundamental group, spaces that are homotopy equivalent are indistinguishable. This is not so good because there are homotopy equivalent spaces that are not homeomorphic (for example an annulus is homotopy equivalent to a circle).

Corollary 4.2.1. The spaces $\mathbb{R}^n \backslash \{\mathbf{0}\}$, $n \geq 3$ are simply connected.

Proof. We will show that $\mathbb{R}^n \backslash \{\mathbf{0}\}$ is homotopy equivalent to S^{n-1} for all $n \geq 1$.

To this end, consider the inclusion of S^{n-1} in $\mathbb{R}^n \backslash \{\mathbf{0}\}$, and the map $r : \mathbb{R}^n \backslash \{\mathbf{0}\} \to S^{n-1}$, $r(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$. Then $r \circ i = 1_{S^{n-1}}$ and $i \circ r \sim 1_{\mathbb{R}^n \backslash \{\mathbf{0}\}}$, with homotopy $H(\mathbf{x}, s) = \mathbf{x}/\|\mathbf{x}\|^s$. □

In view of the proof of this result, we make the following definition.

Definition. If $A \subset X$, a *retraction* of X onto A is a continuous map $r : X \to A$ such that $r|A = 1_A$. In this case A is called a *retract* of X.

Proposition 4.2.5. If A is a retract of X, and $i : A \to X$ is the inclusion, then $i_* : \pi_1(A) \to \pi_1(X)$ is a monomorphism and $r_* : \pi_1(X) \to \pi_1(A)$ is an epimorphism.

Proof. We have $r \circ i = 1_A$, thus $r_* \circ i_* = 1_{\pi_1(A)}$. This implies that i_* is one-to-one, and that r_* is onto. □

There is a related notion.

Definition. Let A be a subspace of X. We say that A is a *deformation retract* of X if 1_X is homotopic to a map that carries all X into A, such that each point of A remains fixed during the homotopy. The homotopy is called *deformation retraction*.

Example 1. The circle is a deformation retract of $\mathbb{C}\backslash\{0\}$. The deformation retraction is $H(z,s) = z/|z|^s$.

Example 2. The figure eight is a deformation retract of the plane without two points, $\mathbb{C}\backslash\{-\frac{1}{2},\frac{1}{2}\}$. Another deformation retract of the plane without two points, $\mathbb{C}\backslash\{-\frac{i}{2},\frac{i}{2}\}$, is $S^1 \cup (\{0\} \times [-1,1])$.

Proposition 4.2.6. Let A be a deformation retract of X. Then the inclusion map $i : (A, x_0) \hookrightarrow (X, x_0)$ induces an isomorphism at the level of fundamental groups.

Proof. Let $H : X \times [0,1] \to X$ be the deformation retraction. Define $r : X \to A$, $r(x) = H(x,1)$. Then $r \circ i = 1_A$. On the other hand $i \circ r$ is homotopic to 1_X, the homotopy being H itself. The conclusion follows from Theorem 4.2.4. □

4.3 The fundamental group of the circle

4.3.1 Covering spaces and the fundamental group

We will introduce now a useful tool for computing and studying the fundamental group of a space. This construction was probably inspired by Riemann's work on elliptic integrals, and the definition of Riemann surfaces.

Definition. A continuous surjective map $p : E \to B$ is called a *covering map* if every point $b \in B$ has an open neighborhood U in B such that $p^{-1}(U)$ is a disjoint union of open sets in E and the restriction of p to each of these open sets is a homeomorphism onto U. E is said to be a *covering space* of B. This open neighborhood U is said to be *evenly covered* by p. The disjoint open sets in $p^{-1}(U)$ are called *slices*.

Example 1. The map $p : \mathbb{R} \to S^1$, $p(x) = e^{2\pi i x}$ is a covering map.

Example 2. Define the space

$$X = \mathbb{R} \sqcup (S^1 \times \mathbb{Z})/\sim$$

where $k \sim (1,k)$. This is the real axis with a tangent circle at each integer. Let Y be the Figure 8 space defined in §1.3.7, which can be thought of as $S^1 \sqcup S^1/1 \sim 1$. The map $p : X \to Y$, $p(x) = e^{2\pi i x}$ for all $x \in \mathbb{R}$, where the image lies in the first S^1, and $p(z,k) = z$ for all $z \in S^1, k \in \mathbb{Z}$, where the image lies in the second S^1 is a covering map.

Proposition 4.3.1. (1) Let $p : E \to B$ be a covering map. If B_0 is a subspace of B and $E_0 = p^{-1}(B_0)$ then $p|E_0 : E_0 \to B_0$ is a covering map.
(2) Let $p : E \to B$ and $p' : E' \to B'$ be covering maps. Then

$$p'' : E \times E' \to B \times B'$$

where $p''(e,e') = (p(e), p'(e'))$ is a covering map.

Proof. (1) If $b_0 \in B_0$, there is an open set U in B such that $p^{-1}(U)$ is the disjoint union of open sets U_α, $\alpha \in \mathcal{A}$ such that $p|_{U_\alpha} : U_\alpha \to U$ is a homeomorphism for each α. The sets $U_\alpha \cap E_0$, $\alpha \in \mathcal{A}$ are open and disjoint in E_0, and each of them is homeomorphic to $U \cap B_0$. This proves the first part.

(2) For the second part, if $p^{-1}(U) = \cup_\alpha U_\alpha$ with U_α disjoint and each of them covers U and is homeomorphic to it via p, and $p^{-1}(U') = \cup_\beta U'_\beta$, where U'_β are disjoint and each of them covers U' and is homeomorphic to U' via p', then $(p,p')^{-1} = \cup_{\alpha,\beta} U_\alpha \times U'_\beta$, and the products $U_\alpha \times U'_\beta$ are disjoint and each covers $U \times U'$ and is homeomorphic to it via (p,p'). $\square$

Example 3. The map

$$p : \mathbb{R}^2 \to S^1 \times S^1,$$

$p(x,y) = (e^{2\pi i x}, e^{2\pi i y})$ is a covering map.

Definition. Let $p : E \to B$ be a continuous surjective map, and let X be a topological space. If $f : X \to B$ is continuous, then a *lifting* of f is a map $\tilde{f} : X \to E$ such that $p \circ \tilde{f} = f$.

Theorem 4.3.1. (Path lifting lemma) Let $p : (E, e_0) \to (B, b_0)$ be a covering map. Then any path $f : [0,1] \to B$ beginning at b_0 has a unique lifting to a path $\tilde{f} : [0,1] \to E$ beginning at e_0.

Proof. Cover the path by open sets U that are evenly covered by p. Because $[0,1]$ is compact, so is $f([0,1])$, and by the Lebesgue number theorem (Theorem 2.4.5), there is $n \geq 1$ such that the partition of $[0,1]$ into n equal intervals has the image of each of this intervals lie inside an open set evenly covered by p. Consider the first interval, $[0, \frac{1}{n}]$, and let U_1 be an open set that contains it and is evenly covered by p. Then the lift of f restricted to this interval must lie in the slice V_1 that contains e_0 (since the slices are open and disjoint, and the path $\tilde{f}$ is connected). Because $p : V_1 \to U_1$ is a homeomorphism, $\tilde{f}|[0, \frac{1}{n}]$ must equal $p^{-1} \circ f|[0, \frac{1}{n}]$. Thus the restriction of f to this interval has a unique lifting that starts at e_0.

We continue inductively. Let us assume that lifting $\tilde{f}|[0, \frac{k}{n}]$ exist and is unique. The lifting of $f|[\frac{k}{n}, \frac{k+1}{n}]$ must start at $\tilde{f}(\frac{k}{n})$. Let $[\frac{k}{n}, \frac{k+1}{n}] \subset U_k$ with U_k evenly covered by V_k. Then the lift of $f|[\frac{k}{n}, \frac{k+1}{n}]$ must lie in the slice V_k that covers U_k and contains $\tilde{f}(\frac{k}{n})$. We obtain again $\tilde{f}|[\frac{k}{n}, \frac{k+1}{n}] = p^{-1} \circ f|[\frac{k}{n}, \frac{k+1}{n}]$.

When $k = n$ we obtain the entire lift $\tilde{f}$, and the construction is unique. $\square$

Theorem 4.3.2. (Homotopy lifting lemma) Let $p : (E, e_0) \to (B, b_0)$ be a covering map. Let $H : [0,1] \times [0,1] \to B$ be a homotopy of the paths f and g with $f(0) = g(0) = b_0$, relative to the endpoints. Then there is a unique lifting of H to $\tilde{H} : [0,1] \times [0,1] \to E$ which is a homotopy relative to the endpoints of the liftings of f and g that start at e_0. Consequently $\tilde{f}(1) = \tilde{g}(1)$.

Proof. The proof parallels that of Theorem 4.3.1. Using the fact that $[0,1] \times [0,1]$ and hence $H([0,1] \times [0,1])$ are compact, using the Lebesgue number theorem we can find $n \geq 1$ such that after dividing $[0,1] \times [0,1]$ into equal squares, the image of each square lies inside an open set evenly covered by p.

We start with the square $[0, \frac{1}{n}] \times [0, \frac{1}{n}]$. There is an open set U_1 evenly covered by p such that $H([0, \frac{1}{n}] \times [0, \frac{1}{n}]) \subset U_1$. Then the lift of H restricted to this square must lie in the slice V_1 that contains e_0. Moreover, $\tilde{H}|[0, \frac{1}{n}] \times [0, \frac{1}{n}]$ must coincide with $p^{-1} \circ H|[0, \frac{1}{n}] \times [0, \frac{1}{n}]$ because $p : V_1 \to U_1$ is an isomorphism. Hence the H restricted to this square has a lifting and this is unique.

Next we pass to $[0, \frac{1}{n}] \times [\frac{1}{n}, \frac{2}{n}]$. As above there is a unique lift $\tilde{H}$ of H restricted to this square such that $\tilde{H}(0, \frac{1}{n})$ is as specified by the definition of $\tilde{H}$ on $[0, \frac{1}{n}] \times [0, \frac{1}{n}]$. Because of Theorem 4.3.1, the two lifts of $H|[0,\frac{1}{n}] \times \{\frac{1}{n}\}$ that come from the two neighboring squares must coincide. So $H|[0,\frac{1}{n}] \times [0,\frac{2}{n}]$ has a unique lifting to an $\tilde{H}$ such that $\tilde{H}(0,0) = e_0$.

If we travel through all the squares in lexicographical order ($(i,j) < (k,l)$ if $i < k$ or $i = k$ and $j < l$), we obtain the lift $\tilde{H}$ in an inductive manner.

Because H is continuous, and because the sets $p^{-1}(b_0)$ and $p^{-1}(f(1))$ are discrete, it follows that $H|\{0\} \times [0,1]$ and $H|\{1\} \times [0,1]$ are constant, and hence H is a homotopy relative to the endpoints. □

Definition. Let $p : (E, e_0) \to (B, b_0)$ be a covering map. Given f a loop in B based at b_0, let $\tilde{f}$ be its lifting to E starting at e_0. The *lifting correspondence* is the map $\Phi_{e_0} : \pi_1(B, b_0) \to p^{-1}(b_0)$, $\Phi_{e_0}([f]) = \tilde{f}(1)$.

In view of Theorem 4.3.2, this map is well defined.

Theorem 4.3.3. Let $p : (E, e_0) \to (B, b_0)$ be a covering map. The lifting correspondence Φ_{e_0} is surjective. If E is simply connected then Φ_{e_0} is a bijection.

Proof. Let e_1 be a point in Φ_{e_0}. Consider a path $\tilde{f} : [0,1] \to E$, such that $\tilde{f}(0) = e_0$ and $\tilde{f}(1) = e_1$. Then $f : [0,1] \to B$, $f = p \circ \tilde{f}$ is a loop in B with $f(0) = b_0$, whose lift to E is $\tilde{f}$. Then $\Phi_{e_0}([f]) = e_1$, which proves surjectivity.

Assume now that E is simply connected. Let f and g be loops and B with liftings $\tilde{f}$ and $\tilde{g}$ starting at e_0, such that $\tilde{f}(1) = \tilde{g}(1)$. Because E is simply connected, $\tilde{f} \sim \tilde{g}$. Let $\tilde{H}$ be the homotopy relative to the end points. Then $H = p \circ \tilde{H}$ is a homotopy relative to the base point between f and g. This proves injectivity. □

Example 4. Let us use this result to compute the fundamental group of the the Lie group $SO(3)$ of rotations in the three-dimensional space.

Note that a rotation is determined by a pair $(\vec{v}, \psi)$, where $\vec{v}$ is a unit vector and ψ is the rotation angle (between 0 and 2π). However, $(\vec{v}, \psi)$ and $(-\vec{v}, -\psi)$ determine the same rotation, so the map $F : S^2 \times [0, 2\pi) \to SO(3)$, $(\vec{v}, \psi) \mapsto$ rotation of angle ψ about $\vec{v}$ is $2-1$. Moreover, if $\psi = 0$, then the vector $\vec{v}$ is ambiguous.

Consider now the space $S^3 = \mathbb{R}^3 \cup \{\infty\}$. Define the map $p : S^3 \to SO(3)$, by $p(\vec{v})=$ counter-clockwise rotation about $\vec{v}$ by the angle $4 \arcsin \frac{\|\vec{v}\|}{1+\|\vec{v}\|}$. When $\vec{v} = 0$ or ∞, then the axis of rotation is ambiguous, but the rotation angle is 0, respectively 2π, and so this does not matter, in both cases the rotation is the identity map.

The map p is $2-1$, 0 and ∞ go to the identity map, and if $\vec{v} \neq 0, \infty$, then $\vec{v}$ determines the same rotation with the vector $\vec{w}$ which is of opposite orientation, and whose length is such that

$$\arcsin \frac{\|\vec{v}\|}{1+\|\vec{v}\|} + \arcsin \frac{\|\vec{w}\|}{1+\|\vec{w}\|} = \frac{\pi}{2}.$$

It is not hard to check that p is continuous. Because S^3 is simply connected, Theorem 4.3.3 implies that $\pi_1(SO(3))$ has two elements. The only group with 2-elements is $\mathbb{Z}_2$. We conclude that

$$\pi_1(SO(3)) = \mathbb{Z}_2.$$

The fact that $SO(3)$ is *not* simply connected is responsible for the existence of the spin in quantum mechanics. One of the elements of $\mathbb{Z}_2$ is spin up, the other is spin down.

Example 5. Because the map $S^n \to \mathbb{R}P^n$ which identifies the antipodes is $2-1$, and because S^n is simply connected if $n \geq 2$, it follows that

$$\pi_1(\mathbb{R}P^n) = \mathbb{Z}_2, \quad n \geq 2.$$

4.3.2 The computation of the fundamental group of the circle

Theorem 4.3.4. The fundamental group of S^1 is isomorphic to the additive group of integers.

Proof. Consider the covering map $p : (\mathbb{R}, 0) \to (S^1, 1)$, $p(x) = e^{2\pi ix}$. From Theorem 4.3.3 it follows that the lifting correspondence is bijective, and we see that $p^{-1}(1) = \mathbb{Z}$. In particular, the loops $f : (S^1, 1) \to (S^1, 1)$, $f_n(z) = z^n$, $n \in \mathbb{Z}$ give all equivalence classes in $\pi_1(S^1, 1)$. Because S^1 is a group (the group $U(1)$ of rotations of the plane about the origin), it follows from Problem 4 in Homework 5 that $[z^m][z^n] = [z^{m+n}]$ *is* the multiplication rule in $\pi_1(S^1, 1)$. Hence the map Φ_1 is a group isomorphism. □

Corollary 4.3.1. The manifolds S^1 and S^2 are not homeomorphic.

This result can be used to compute the fundamental groups of other spaces.

Example. $\pi_1(\mathbb{C}\backslash\{0\}) = \mathbb{Z}$.
This follows from the fact that $\mathbb{C}\backslash\{0\} \sim S^1$ (see the proof to Corollary 4.2.1).

Corollary 4.3.2. The manifolds $\mathbb{R}^2$ and $\mathbb{R}^3$ are not homeomorphic.

Proof. Suppose that $h : \mathbb{R}^2 \to \mathbb{R}^3$ is a homeomorphism. Then $h : \mathbb{R}^2\backslash\{\mathbf{0}\} \to \mathbb{R}^3\backslash\{h(\mathbf{0})\}$ is a homeomorphism as well. But

$$\pi_1(\mathbb{R}^2\backslash\{\mathbf{0}\}) = \mathbb{Z} \quad \text{and} \quad \pi_1(\mathbb{R}^3\backslash\{\mathbf{0}\}) = \{0\},$$

by Theorem 4.3.4 and Corollary 4.2.1, so the two spaces cannot be homeomorphic. □

4.3.3 Applications of the fundamental group of the circle

For simplicity, in what follows we will denote by $\bar{B}^2$ the closed unit disk. Also, a map will be called *null homotopic* if it is homotopic to a constant map.

Lemma 4.3.1. Let $f : S^1 \to S^1$ be a continuous map. The following are equivalent:
(1) f is null homotopic;
(2) there is a continuous map $F : \bar{B}^2 \to S^1$ such that $F|S^1 = f$;
(3) $f_* : \pi_1(S^1, 1) \to \pi_1(S^1, f(1))$ is the zero map.

Proof. (1) implies (2). Let $H : S^1 \times [0, 1] \to S^1$ be the homotopy between f and the trivial loop $g(e^{2\pi it}) = 1$ for all t. Define $F(re^{2\pi it}) = H(e^{2\pi it}, r)$.

(2) implies (3). The composition of F and the inclusion i

$$(S^1, 1) \xrightarrow{i} (\bar{B}^2, 1) \xrightarrow{F} (S^1, f(1))$$

is just f. Hence $f_* = F_* \circ i_*$, by Proposition 4.2.3. But F_* is the zero map since the fundamental group of $\bar{B}^2$ is the trivial group.

(3) implies (1). The loop defined by the map 1_{S^1} is mapped by f to a null homotopic loop. This means that $f \circ 1_{S^1} = f$ is null homotopic. □

Theorem 4.3.5. Given a non-vanishing continuous vector field on $\bar{B}^2$, there is a point of S^1 where the vector field points directly inwards and a point of S^1 where it points directly outwards.

Proof. Let $\mathbf{v}(z)$ be the vector field. Normalize it to a unit vector field by taking $\mathbf{v}(z)/\|\mathbf{v}(z)\|$. This can be interpreted as a continuous function $F : \bar{B}^2 \to S^1$. By Lemma 4.3.1 the function $f = F|S^1$ induces the trivial homomorphism at the level of the fundamental group.

Consider the standard covering map $p : \mathbb{R} \to S^1$, view f as a map from $[0, 1]$ to S^1 and take the unique lifting $\tilde{f} : [0, 1] \to \mathbb{R}$ such that $\tilde{f}(0) \in [0, 1]$ (which exists by Theorem 4.3.1). By Theorem 4.3.3, $\tilde{f}(1) = \tilde{f}(0)$. The function $g(t) = \tilde{f}(t) - t$ has the property that $g(0) \geq 0$, $g(1) \leq 0$; by the Intermediate Value Property there is $t \in [0, 1]$ such that $g(t) = 0$. Then $\tilde{f}(t) = t$, showing that $p(t)$ is a fixed point for f. At this fixed point the vector field points directly outwards.

To find a point where the vector field points inwards, replace the vector field by its negative and apply this result. □

Theorem 4.3.6. (The 2-dimensional Brouwer fixed-point theorem) If $f : \bar{B}^2 \to \bar{B}^2$ is continuous, then there is $z \in \bar{B}^2$ such that $f(z) = z$.

Proof. We will argue by contradiction. Assume that f has no fixed point. Then we can define a continuous vector field by assigning to each $z \in \bar{B}^2$ the vector from z to $f(z)$. By Theorem 4.3.5 there is $z \in S^1$ where the vector field points directly outwards. But this is impossible since it would mean that $f(z)$ lies outside of the disk. Hence the conclusion. □

Here is a physical application to Theorem 4.3.5. Consider a flat elastic patch that is a convex region placed on top of a plane. If we stretch this patch and then release, there is a point that does not move. Indeed, the force field of tensions in the rubber patch points inwards at each point of the boundary, so there must be a point where this vector field is zero. In fact this is the fixed point of Brower's theorem, since as the patch shrinks it ends up inside the region it covered when extended, but we see that moreover this point does not move during the contraction.

Proposition 4.3.2. Let A be a 3×3 matrix with positive entries. Then A has an eigenvector with positive entries.

Proof. The set $D = [0, \infty)^3 \cap S^2$ is topologically a closed 2-dimensional disk. Define $f : D \to D$, $f(\mathbf{v}) = A\mathbf{v}/\|A\mathbf{v}\|$. By Brouwer's fixed point theorem f has a fixed point $\mathbf{v_0}$. Then $A\mathbf{v_0} = \|A\mathbf{v_0}\|\mathbf{v_0}$, thus $\mathbf{v_0}$ is the desired eigenvector. □

Remark 4.3.1. The Perron-Frobenius theorem in $\mathbb{R}^n$ states that any square matrix with positive entries has a unique eigenvector with positive entries (up to a multiplication by a positive constant), and the corresponding eigenvalue has multiplicity one and is strictly greater than the absolute value of any other eigenvalue.

Theorem 4.3.7. (The Gauss-d'Alembert fundamental theorem of algebra) Every non-constant polynomial with complex coefficients has at least one complex zero.

Proof. Let $f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ by the polynomial. By rescaling the variable we may assume that $|a_0| + |a_1| + \cdots + |a_{n-1}| < 1$. We will show that $f(\bar{B}^2)$ contains 0.

Arguing by contradiction we assume that $0 \notin f(\bar{B}^2)$. Then we can define the continuous function $G : \bar{B}^2 \to S^1$, $G(z) = f(z)/|f(z)|$. By Lemma 4.3.1, $g = G|S^1$ induces the trivial map at the level of the fundamental group.

On the other hand

$$H(z, s) = \frac{z^n + sa_{n-1}z^{n-1} + \cdots + sa_1 z + sa_0}{|z^n + sa_{n-1}z^{n-1} + \cdots + sa_1 z + sa_0|}$$

is a homotopy between g and z^n. But z^n is not null homotopic because it induces the map $k \mapsto nk$ at the level of the fundamental group. This is a contradiction, which proves that $0 \in f(\bar{B}^2)$. We are done. $\square$

Theorem 4.3.8. (The Borsuk-Ulam theorem) Every continuous map from the 2-dimensional sphere to the plane maps some pair of antipodal points to the same point.

Proof. Let us assume that this is not true, namely that there is a map $f : S^2 \to \mathbb{R}^2$ such that $f(\mathbf{x}) \neq f(-\mathbf{x})$ for all $\mathbf{x} \in S^2$. Then we can define the map $G : S^2 \to S^1$ as the unit vector parallel to the vector from $f(\mathbf{x})$ to $f(-\mathbf{x})$. Intersecting S^2 by $\mathbb{R}^2 \times \{0\}$ gives a circle, and by restriction we obtain a map $g : S^1 \to S^1$ with $g(-z) = -g(z)$. Because g extends to the upper hemisphere, which is a disk, g is null-homotopic.

Returning to the theorem, think of g as a map from $[0,1]$ to S^1 and lift it to a map $\tilde{g} : [0,1] \to \mathbb{R}$. Because g is null homotopic, $\tilde{g}(0) = \tilde{g}(1)$. By the 1-dimensional Borsuk-Ulam theorem (Theorem 2.3.6), there are diametrically opposite points z and $-z$ for which $\tilde{g}$ takes the same value. It follows that $g(z) = g(-z)$ which contradicts $g(-z) = -g(z)$. Hence our initial assumption is false. The conclusion follows. $\square$

4.4 The structure of covering spaces

In what follows all topological spaces will be assumed path connected, unless otherwise specified.

4.4.1 Existence of covering spaces

In this section we will introduce a method for constructing covering spaces. The idea should probably be attributed to B. Riemann; this is how Riemann surfaces are constructed. In a historical perspective, the question was that of studying an integral of the form $\int R(x,y)dx$, where R is a rational function, and y is defined implicitly in terms of x by a polynomial equation $P(x,y) = 0$. Riemann's idea, based on insights of Abel and Jacobi, was to shift to complex variables $x \mapsto z$, $y \mapsto w$. The function w is defined implicitly by $P(z,w) = 0$, and as such lives naturally on a Riemann surface. Then we are supposed to study line integrals on a Riemann surface. Aha, so Riemann surfaces and line integrals come together! But then line integrals themselves are used to build the Riemann surface: choose a reference point in the plane, and then a point b in the plane is covered by several points on the Riemann surface, and those points are defined by the different values of the line integrals on various paths joining the reference point with b. Cauchy's theorem tells us that different values are obtained when paths have holes between them, and so we arrive at topology. Now we take out the complex analysis, and keep just the topological skeleton. And we have the construction below.

This method of constructing covering spaces works for all topological spaces of interest to us. However, one should keep in mind that there are pathological topological spaces for which the construction fails. We will impose the space to be semilocally simply connected:

Definition. A space B is called *semilocally simply connected* if for each $b \in B$, there is a neighborhood U of b such that the homomorphism

$$i_* : \pi_1(U,b) \to \pi_1(B,b)$$

induced by inclusion is trivial.

Example All manifolds have this property, and in particular any space in which all points have simply connected neighborhoods.

Theorem 4.4.1. Let B be path connected, locally path connected, and semilocally simply connected. Let $b_0 \in B$. Given a subgroup H of $\pi_1(B, b_0)$, there exists a path connected, locally path connected space E with a covering map $p : E \to B$, and a point $e_0 \in p^{-1}(b_0)$ such that

$$p_*(\pi_1(E, e_0)) = H.$$

Proof. • *Construction of E*

On the set of all paths in B beginning at b_0 we define an equivalence by

$$\alpha \sim \beta \text{ if and only if } \alpha(1) = \beta(1) \text{ and } [\alpha * \bar{\beta}] \in H.$$

We denote the equivalence class of α by $\hat{\alpha}$. We define E to be the set of equivalence classes, and e_0 the equivalence class of the constant path e_{b_0}. The covering map $p : E \to B$ is defined by the equation

$$p(\hat{\alpha}) = \alpha(1).$$

Since B is path connected, p is onto.

From the fact that H contains the identity element it follows that this equivalence relation is *coarser* than homotopy relative to the endpoints.

Also, if $\alpha \sim \beta$ then $\alpha * \delta \sim \beta * \delta$ for any path δ for which this makes sense.

Now we give E a topology. A basis for this topology consists of

$$B(U, \alpha) = \{\widehat{\alpha * \delta} \,|\, \delta \text{ is a path in } U \text{ beginning at } \alpha(1)\}.$$

Let us prove that this is indeed a basis for a topology.

(1) First note that $\hat{\alpha} \in B(U, \alpha)$, since we can choose δ to be the trivial path. By varying α, we find that the sets $B(U, \alpha)$ cover E, hence the first condition is satisfied.

Let us now show that if $\hat{\beta} \in B(U_1, \alpha_1) \cap B(U_2, \alpha_2)$ then there is $B(U_3, \alpha_3)$ such that

$$\hat{\beta} \in B(U_3, \alpha_3) \subset B(U_1, \alpha_1) \cap B(U_2, \alpha_2).$$

Note that $\beta(1)$ belongs to the intersection of U_1 and U_2. We choose $U_3 = U_1 \cap U_2$ and $\alpha_3 = \beta$. We trivially have

$$B(U_3, \beta) \subset B(U_1, \beta) \cap B(U_2, \beta),$$

and the conclusion would follow if we showed that $B(U_i, \beta) \subset B(U_i, \alpha_i)$, $i = 1, 2$. In fact much more is true, as the following lemma shows.

Lemma 4.4.1. If $\hat{\beta} \in B(U, \alpha)$ then $B(U, \alpha) = B(U, \beta)$.

Proof. Write $\beta = \alpha * \delta$ and note that $\alpha = \beta * \bar{\delta}$. Thus $\hat{\alpha} \in B(U, \beta)$.

On the other hand, if $\gamma \in B(U, \beta)$, then $\gamma = \beta * \delta'$, and since $\beta = \alpha * \delta$, then $\gamma = \alpha * \delta * \delta'$. Thus $\gamma \in B(U, \alpha)$. Thus $B(U, \beta) \subset B(U, \alpha)$. The first part of the proof implies the reverse inclusion and we are done. ☐

• *$p : E \to B$ is a covering map*

We first check that p is continuous and maps open sets to open sets.

Indeed, let U be an open set in B and $\hat{\alpha}$ a point in $p^{-1}(U)$. We will show that $\hat{\alpha}$ has a neighborhood that maps inside U. Let V be a path connected neighborhood such that $p(\alpha) \in V \subset U$. Then $B(V, \alpha)$ is a neighborhood of $\hat{\alpha}$ that is mapped inside U. This proves continuity. As for the rest, note that $p(B(V, \alpha)) = V$.

Now let us show that every point $b \in B$ has an open neighborhood that is evenly covered by p. Here is where we use the fact that B is *semilocally simply connected.* Choose the open neighborhood of U such that the homomorphism $\pi_1(U, b) \to \pi_1(B, b)$ induced by inclusion is trivial.

We will show that the set $p^{-1}(U)$ is the union of all sets of the form $B(U, \alpha)$, where α ranges over all paths in B from b_0 to b. As we saw, p maps $B(U, \alpha)$ to U, so the preimage contains all these sets. On the other hand, if $\hat{\beta}$ is in $p^{-1}(U)$, then $\beta(1) \in U$. Let $\alpha = \beta * \bar{\delta}$, where δ is a path in U from b to $\beta(1)$. Then $\hat{\beta} \in B(U, \alpha)$. This proves the converse inclusion.

If $\hat{\beta} \in B(U, \alpha_1) \cap B(U, \alpha_2)$ then $B(U, \alpha_1) = B(U, \alpha_2) = B(U, \beta)$ by Lemma 4.4.1. So the sets $B(U, \alpha)$ either coincide or are disjoint.

We now show that $p : B(U, \alpha) \to U$ is a homeomorphism. We know that the map is onto, so let us check that it is one-to-one. If

$$p(\widehat{\alpha * \delta}) = p(\widehat{\alpha * \delta'})$$

then $\delta(1) = \delta'(1)$, and the loop $\delta * \bar{\delta}'$ is null homotopic (in B), because it lies in U. Then $[\alpha * \delta] = [\alpha * \delta']$, and the same is true for the coarser equivalence relation: $\widehat{\alpha * \delta} = \widehat{\alpha * \delta'}$. Using also the fact that p maps open sets to open sets, we conclude that its restriction to $B(U, \alpha)$ is a homeomorphism.

• *E is path connected*

Let $\hat{\alpha}$ be a point in E. Define $f : [0, 1] \to E$, $f(s) = \widehat{\alpha_s}$, where $\alpha_s(t) = \alpha(st)$. Then f is a path from $e_0 = \widehat{e_{b_0}}$ to $\hat{\alpha}$, which proves that E is path connected. But we need to check that f is a continuous function. So let us check that for any open set of the form $B(U, \beta)$, $f^{-1}(B(U, \beta))$ is open in $[0, 1]$. If s_0 is a point in the preimage, then α_{s_0} is in the same equivalence class as some $\beta * \delta$ with δ a path in U. Because $\beta * \delta$ and α_{s_0} have the same endpoint, it follows that $\alpha_{s_0}(1) \in U$. Because U is open, $\alpha_s(1) = \alpha(s) \in U$ for s in an open interval containing s_0. Augmenting to $\beta * \delta$ the path from $\alpha(s_0)$ to α_s (running along α), we deduce that $\widehat{\alpha_s} \in B(U, \beta)$ for all these s. This shows that $p^{-1}(B(U, \beta))$ contains an entire open interval around s_0, and since s_0 was arbitrary, this set is open. This proves continuity.

• *$p_*(\pi_1(E, e_0)) = H$.*

Consider a loop α in B based at b_0. Then we can lift it to a path $\tilde{\alpha}$ in E by considering the paths obtained by travelling partially around α. Note that the lift is unique. Explicitly, $\tilde{\alpha}(s) = \alpha_s$ for $s \in [0, 1]$, where $\alpha_s(t) = \alpha(st)$.

$\alpha \in p_*(\pi_1(E, e_0))$ can be rephrased by saying that $\tilde{\alpha}$ is a loop. The fact that $\tilde{\alpha}$ is a loop is equivalent to the fact that $\tilde{\alpha}(0) = \tilde{\alpha}(1)$, i.e. $\widehat{e_{b_0}} = \hat{\alpha}$. This is equivalent, by definition, to $\alpha \in H$. Hence the conclusion. □

The next result shows that additional requirement that B be semilocally simply connected is necessary, at least in the case $H = \{e\}$.

Proposition 4.4.1. Assume that there is a covering map $p : E \to B$ with E simply connected. Then B is semilocally simply connected.

Proof. Choose U and open set that is evenly covered by p. Then U has the desired property. Indeed, if α is a loop in U, its lift to one of the slices is null-homotopic in E. Projecting the homotopy to B we obtain a homotopy in B between α and the trivial loop. Thus $i_*(\alpha) = 0$, as desired. $\square$

4.4.2 Equivalence of covering spaces

The main result in the previous section allows us to construct many covering spaces, but some of these constructions are equivalent.

Definition. Two covering maps $p : E \to B$ and $p' : E' \to B$ are called *equivalent* if there is a homeomorphism $h : E \to E'$ that makes the following diagram commute:

$$\begin{array}{ccc} E & \xrightarrow{h} & E' \\ & {\scriptstyle p}\searrow \quad \swarrow{\scriptstyle p'} & \\ & B & \end{array}$$

Equivalence is the analogue of group isomorphism; the spaces are the same, they are only denoted differently. In order to adress the equivalence problem for the covering spaces of a given space, we need the following result.

Theorem 4.4.2. (The general lifting lemma) Let $p : (E, e_0) \to (B, b_0)$ be a covering map, and $f : (Y, y_0) \to (B, b_0)$ be a continuous map, with Y path connected and locally path connected. The map f can be lifted to $\tilde{f} : (Y, y_0) \to (E, e_0)$ if and only if

$$f_*(\pi_1(Y, y_0)) \subset p_*(\pi_1(E, e_0)).$$

Furthermore, if such lifting exists, it is unique.

Proof. The condition is necessary since $p \circ \tilde{f} = f$ implies

$$f_*(\pi_1(Y, y_0)) = p_*(\tilde{f}_*(\pi_1(Y, y_0)) \subset p_*(\pi_1(E, e_0)).$$

To show that the condition is sufficient, recall the Path Lifting Lemma 4.3.1, which is the particular case of this result when Y is an interval. Since Y is path connected, we can cover it with a "star" of paths starting at y_0, lift f restricted to each path individually, then check that the result yields the desired lifting of f on the whole Y.

For the rigorous construction, let $y \in Y$ and α a path from y_0 to y. Then $f \circ \alpha$ is a path from $f(y_0) = b_0$ to $f(y)$. Lift this to a path β from e_0 to some $z \in E$, and define $\tilde{f}(y) = z$. Note that by the uniqueness of the path lifting, we are obliged to define the lift of f this way, which proves the uniqueness of $\tilde{f}$.

We have to show that $\tilde{f}$ is well defined and continuous. Let us assume that we have another path α' from y_0 to y. Then $f \circ \alpha'$ is a different path from b_0 to $f(y)$, which lifts to some path β' from e_0 to some $z' \in E$. Note that $(f \circ \alpha) * \overline{(f \circ \alpha')}$ is a loop in $f_*(\pi_1(Y, y_0))$, hence in $p_*(\pi_1(E, e_0))$. Thus we can lift it to a loop in E. For this loop to "close up", we need $z = z'$, so $\tilde{f}$ is well defined.

Let us check continuity. Given a point $y \in Y$, let $b = f(y)$, U an open neighborhood of b evenly covered by p, and V the slice containing $\tilde{f}(y)$. The set $f^{-1}(U)$ is an open neighborhood of y, and on this set $\tilde{f} = (p|V)^{-1} \circ f$ which is continuous being a composition of continuous functions. The theorem is proved. $\square$

Theorem 4.4.3. Let $p : (E, e_0) \to (B, b_0)$ and $p' : (E', e'_0) \to (B, b_0)$ be covering maps. Then there is a homeomorphism $h : E \to E'$ such that $h(e_0) = e'_0$ establishing an equivalence of covering maps if and only if the groups

$$H = p_*(\pi_1(E, e_0)) \quad \text{and } H' = p'_*(\pi_1(E', e'_0))$$

are equal. If h exists, it is unique.

Proof. If h exists, then

$$p'_*(\pi_1(E', e'_0)) = p'_*(h_*(\pi_1(E, e_0)) = p_*((\pi_1(E, e_0))$$

so $H = H'$.

For the converse, we apply the General lifting lemma 4.4.2. The trick is to turn "on one side" the diagram defining equivalence:

$$\begin{array}{ccc} & & E' \\ & \overset{h}{\nearrow} & \downarrow p' \\ E & \xrightarrow{p} & B \end{array}$$

As such, for $Y = E$, $f = p$ and the covering map $p' : E' \to B$. Let $h : (E, e_0) \to (E', e'_0)$ be the lift of p obtained this way. We claim that it is the desired homeomorphism.

To prove that this is a homeomorphism, let $h' : (E', e'_0) \to (E, e_0)$ be the map obtained by switching the roles of E and E'. Note that the map $h' \circ h$ satisfies the conditions of the General lifting lemma for $Y = E$, $f = p$ and $p : E \to B$. So does the map 1_E. By uniqueness $h' \circ h = 1_E$. Similarly $h \circ h' = 1_{E'}$.

Finally, the uniqueness of h is guaranteed by the General lifting lemma, since the fact that the diagram is commutative is a way of saying that h is a lifting of p.

□

In fact we can do better than this, we can prove a result which does not involve base points.

Theorem 4.4.4. Let $p : (E, e_0) \to (B, b_0)$ and $p' : (E', e'_0) \to (B, b_0)$ be covering maps. Then there is an equivalence relation $h : E \to E'$ if and only if the groups

$$H = p_*(\pi_1(E, e_0)) \quad \text{and } H' = p'_*(\pi_1(E', e'_0))$$

are conjugate in $\pi_1(B, b_0)$.

Recall that two subgroups H_1 and H_2 of a group G are called conjugate if there is $\alpha \in G$ such that $H_2 = \alpha H_1 \alpha^{-1}$.

Proof. As in the case of the previous result, let us prove first the necessity. If $e'_1 = h(e_0)$, then by Theorem 4.4.3 $p_*(\pi_1(E, e_0)) = p'_*(\pi_1(E', e'_1))$. If β is a path from e'_1 to e'_0 in E', then, because e_0, e'_0 and e'_1 sit above the same point in B, $\alpha = p \circ \beta$ is a loop, and $p'_*(\pi_1(E', e'_1)) = \alpha p'_*(\pi_1(E', e'_0))\alpha^{-1}$.

To do the converse, we do this construction in reverse. If $H' = \alpha H \alpha^{-1}$ let β be the lift of α starting at e'_0, and set $e'_1 = \beta(1)$. Then $p'_*(\pi_1(E', e'_1)) = \alpha^{-1}\pi'_*(\pi_1(E', e'_0))\alpha = \pi_*(\pi_1(E, e_0))$, so we are in the conditions of Theorem 4.4.3, and the conclusion follows. □

Thus, the covering spaces of a given space are in one-to-one correspondence with the conjugacy classes of subgroups of its fundamental group.

Example 1. The projective space $\mathbb{R}P^n$ has the fundamental group equal to $\mathbb{Z}_2$, so it has only two covering spaces, the projective space itself and the sphere S^n.

Example 2. The covering spaces of the plane without a point $\mathbb{C}\backslash\{0\}$ are in one-to-one correspondence with the subgroups of $\mathbb{Z}$. These subgroups are of the form $m\mathbb{Z}$, $m \geq 0$. The corresponding covering spaces are the Riemann surfaces of the function $z \mapsto \sqrt[m]{z}$ if $m > 0$, and the Riemann surface of the function $z \mapsto \ln z$ if $m = 0$.

Example 3. The fundamental group of the torus $S^1 \times S^1$ is $\mathbb{Z} \oplus \mathbb{Z}$, which is abelian, so its covering spaces are in $1-1$ correspondence with the subgroups of this group, namely with $m\mathbb{Z} \oplus n\mathbb{Z}$, $m, n = 0, 1, 2, 3, \ldots$. These covering spaces are tori, if $m, n > 0$, cylinders, if $m = 0$ or $n = 0$ but not both, or $\mathbb{R}^2$ if $m = n = 0$.

The subgroups of the fundamental group form a lattice under inclusion. To this lattice corresponds a *lattice of covering spaces*, as the next result shows.

Proposition 4.4.2. Let H, H' be subgroups of $\pi_1(B, b_0)$ such that H' is a subgroup of H. Let also $p : (E, e_0) \to (B, b_0)$ and $p' : (E', e_0') \to (B, b_0)$ be the coverings associated to H respective H'. Then there is a unique covering map $p'' : (E', e_0') \to (E, e_0)$, that makes the following diagram commutative:

$$\begin{array}{ccc} E' & \xrightarrow{p''} & E \\ & {\scriptstyle p'}\searrow \quad \swarrow{\scriptstyle p} & \\ & B & \end{array}$$

Proof. The existence and uniqueness of the map p'' follow from the General lifting lemma (Theorem 4.4.2) applied to $f = p'$, $Y = E'$ and the covering map $p : E \to B$.

To show that p'' is a covering map, choose $e \in E$. Let U be an open subset of B that is evenly covered by E and contains $p(e)$ and V the slice of E that lies above it containg e. Let U' be an open set of B containing $p(e)$ that is evenly covered by p'. Taking $U \cap U'$ for both U and U', we may assume that $U = U'$. In that case for each slice W of E' that lies above U and maps through p'' to V the map $p''^{-1} : V \to W$ is defined as $p''^{-1} = p'^{-1} \circ p$. $\square$

At the very top of the lattice of covering spaces sits the covering space corresponding to the trivial subgroup of the fundamental group, which is simply connected. It is also the unique simply connected covering space up to equivalence.

Definition. The unique simply connected covering space of a path connected, locally path connected, semilocally simply connected space B is called the *universal covering space* of B.

Remark 4.4.1. All covering spaces in the lattice of a given space B have the same universal covering space, namely the covering space of B. It is in this sense that the covering is universal.

4.4.3 Deck transformations

We assume again that *all spaces are path connected and locally path connected.* This section is about the symmetries of a covering space.

Definition. Given a covering map $p : E \to B$, a *deck transformation (or covering transformation)* is a homeomorphism $h : E \to E$ such that $p \circ h = p$.

Another way to say this is that a deck transformation is an equivalence of a covering with itself. Deck transformations form a group, the *deck transformation group*, which we denote by $\mathrm{Aut}(p)$.

A deck transformation induces a permutation of the elements of each fiber $p^{-1}(b)$. Thus the group of deck transformation acts on the fibers. Because of the uniqueness of path lifting, if h is not the identity map, then the action of h has no fixed points. For the same reason, the action is free (i.e. the only map that acts as identity on a fiber is $h = 1_E$). If the action is transitive in one fiber (meaning that for every $e, e' \in p^{-1}(b)$ there is $h \in \mathrm{Aut}(p)$ such that $h(e) = e'$), then it is transitive in all fibers.

Definition. A covering for which the group of deck transformations acts transitively in each fiber is called *regular*.

Example 1. The universal covering of the circle $p : \mathbb{R} \to S^1$, $p(x) = e^{2\pi i x}$ is regular. The group of deck transformations is the group of integer translations in $\mathbb{R}$, which is $\mathbb{Z}$. In fact we have used the group of deck transformations of this cover in the computation of the fundamental group of the circle in §4.3.2.

Given a covering map $p : (E, e_0) \to (B, b_0)$, recall the lifting correspondence $\Phi_{e_0} : \pi_1(B, b_0) \to p^{-1}(b_0)$ defined in §4.3.1.

Lemma 4.4.2. Let $H = p_*(\pi_1(E, e_0))$. The lifting correspondence induces a bijective map

$$\Phi_{e_0} : \pi_1(B, b_0)/H \to p^{-1}(b_0)$$

where $\pi_1(B, b_0)/H$ is the collection of the right cosets of the group H.

Proof. Assume that $\Phi_{e_0}(f) = \Phi_{e_0}(g)$ for some loops f and g based at b_0. Then $f * \bar{g}$ is a loop based at b_0 that lifts to a loop in E based at e_0. Thus $[f * \bar{g}] \in H$, and $[f] = [f * \bar{g}][g]$ implies $[f] \in H[g]$. □

Working with $\bar{f} * g$ instead, we obtain the same conclusion with righ cosets replaced by left cosets.

The group of deck transformations defines a map $\Psi_{e_0} : \mathrm{Aut}(p) \to p^{-1}(b_0)$, by $\Psi_{e_0}(h) = h(e_0)$. Since h is uniquely determined once its value at e_0 is known, Ψ_{e_0} is injective.

Lemma 4.4.3. The image of Ψ_{e_0} is equal to $\Phi_{e_0}(N(H)/H)$, where $N(H)$ is the normalizer of H, namely the group of all deck transformations h for which $hHh^{-1} = H$.

Proof. Recall that $\Phi_{e_0}([\alpha]) = \tilde{\alpha}(1)$, where $\tilde{\alpha}$ is the lift of α starting at e_0. Hence we have to show that there is $h \in \mathrm{Aut}(p)$ such that $h(e_0) = \tilde{\alpha}(1)$ if and only if $\alpha \in N(H)$.

By Theorem 4.4.3, h exists if and only if $H = p_*(\pi_1(E, h(e_0)))$. But, as in the proof of Theorem 4.4.4, $H = [\alpha] * p_*(\pi_1(E, e_0)) * [\alpha]^{-1}$. Thus h exists if and only if $H = [\alpha] * H * [\alpha]^{-1}$, i.e. $\alpha \in N(H)$. □

Note that although $\pi_1(B, b_0)/H$ does not have a group structure, $N(H)/H$ does. This is because H is normal in $N(H)$.

Theorem 4.4.5. The bijection

$$\Phi_{e_0}^{-1} \circ \Psi_{e_0} : \mathrm{Aut}(p) \to N(H)/H$$

is a group isomorphism.

Proof. The fact that this map is bijective follows from Lemma 4.4.3. To check that it is a group homomorphism, note that $\Phi_{e_0}^{-1} \circ \Psi_{e_0}$ maps h to the class of a loop α in $N(H)/H$ such that $\tilde{\alpha}(1) = h(e_0)$. If $h, h' \in \mathrm{Aut}(p)$, and if α and α' are representatives of the corresponding classes of paths, then the unique lift of α' that starts at $h(e)$ ends at $h'(h(e))$, so the path $\alpha * \alpha'$ is the path that corresponds to $h' \circ h$, which shows that

$$\Phi_{e_0}^{-1} \circ \Psi_{e_0}(h' \circ h) = \Phi_{e_0}^{-1} \circ \Psi_{e_0}(h') * \Phi_{e_0}^{-1} * \Psi_{e_0}(h).$$

Also it is not hard to see that the identity deck transformation is mapped to the (class of the) trivial loop. This proves that the map is a group homomorphism and the conclusion follows. □

Corollary 4.4.1. The group H is normal in $\pi_1(B, b_0)$ if and only if the covering is regular. In this case $N(H) = \pi_1(B, b_0)$ and there is an isomorphism

$$\Phi_{e_0}^{-1} \circ \Psi_{e_0} : \mathrm{Aut}(p) \to \pi_1(B, b_0)/H.$$

Corollary 4.4.2. If E is the universal cover of B, then $\mathrm{Aut}(p)$ is isomorphic to $\pi_1(B, b_0)$.

Definition. Given a group of homeomorphisms acting on a space X, we define the *orbit space* to be the quotient space by the equivalence relation $x \sim g(x)$, $x \in X$, $g \in G$. The equivalence class of x is called the *orbit* of x.

Definition. The action of a group G of homeomorphisms on a space X is called properly discontinuous, if for every $x \in X$ there is an open neighborhood U of x such that for every $g_1, g_2 \in G$, $g_1 \neq g_2$, the sets $g_1(U)$ and $g_2(U)$ are disjoint.

Theorem 4.4.6. Let G be a group of homeomorphisms of a path connected, locally path connected space E. Then the quotient map $\pi : E \to E/G$ is a covering map if and only if the action of G is properly discontinuous. In this case π is regular, and G is its group of deck transformations.

Moreover, if $p : E \to B$ is a regular covering map, and G its group of deck transformations, then there is a homeomorphism $f : E/G \to B$ such that $p = f \circ \pi$.

Proof. The fact that on E/G we put the quotient topology makes π both continuous and open.

First, if π is a covering map, then the action is properly discontinuous. Indeed, if $b \in E/G$, then $\pi^{-1}(b) = \{g(x) \,|\, g \in G\}$ for some x. If V is an open neighborhood of b that is evenly covered, then $\pi^{-1}(V)$ is the disjoint union of slices of the form $g(U)$, where U is some open neighborhood of x, and therefore the sets $g(U)$, $g \in G$ are disjoint from each other.

Conversely, if the action is properly discontinuous, then for each open set U such that the sets $g(U)$, $g \in G$ are disjoint, the set $\pi(U)$ is evenly covered (with slices the sets $g(U)$). So π is a covering map.

Now let us show that G is indeed the group of covering transformations. Certainly any $g \in G$ defines a covering equivalence, because the orbit of x is the same as the orbit of $g(x)$. On the other hand, if h maps x_1 to x_2, then these belong to the same orbit, so there is g such that $h(x_1) = g(x_1)$. The uniqueness in Theorem 4.4.3 implies that $h = g$.

Finally, π is regular because G acts transitively on orbits.

To prove the second part, notice that p is constant on each orbit, and so it induces a continuous map $f : E/G \to B$. Also π factors to a continuous map $B \to E/G$. It is not hard to see that this is the inverse of f, and we are done. □

Example 2. The real projective space $\mathbb{R}P^n$ arises as the properly discontinuous action of $\mathbb{Z}_2$ on the sphere S^n.

Example 3. A family of topological spaces that arise as quotients of a space by a group action that is properly discontinuous are the *lens spaces*. They arise as quotients of S^3 by a finite cyclic group. Specifically, let n, k be positive integers, and consider

$$S^3 = \{(z_1, z_2) \,|\, |z_1|^2 + |z_2|^2 = 1\} \subset \mathbb{C}^3\}$$

with the $\mathbb{Z}_n$-action given by

$$\hat{m}(z_1, z_2) = \left(e^{2\pi i m/n} z_1, e^{2\pi i m k/n} z_2\right), \quad \hat{m} \in \mathbb{Z}_n.$$

This action is properly discontinuous, because one of z_1 or z_2 has the absolute value at least $\sqrt{2}/2$, and when you rotate it in the plane you move it at some distance from itself.

The quotient is a 3-dimensional manifold $L(n, k)$, called lens space. Note that $L(2, 1)$ is the 3-dimensional real projective space $\mathbb{R}P^3$.

By Corollary 4.4.2, $\pi_1(L(n, k)) = \mathbb{Z}_n$. In particular, if $n \neq n'$ then $L(n, k)$ is not homeomorphic to $L(n', k')$. We thus deduce the existence of infinitely many 3-dimensional manifolds that are not homeomorphic to each other.

Example 4. The computation of the fundamental group of the figure eight.

Let B be the figure eight. Consider the x and y-axes in the plane, and at each point of integer coordinates glue another copy of the real axis, with the zero at the point. Now you have infinitely many real axes. At each integer of these axes glue another copy of the real axis (at its origin). Keep repeating forever. At each stage you have a topological space, and it is not hard to see that the final result can be given the structure of a topological space itself. Call this space E. It is simply connected.

Now consider the maps that send this space to itself so that the restriction to any line is an isometry to the image of that line (which is another line). This is a group G that acts properly discontinuously. The quotient E/G is homeomorphic to B, and we obtain the covering map $E \to E/G = B$. The group of deck transformations is isomorphic to $\pi_1(B)$ and this group is F_2, the free group on 2 elements.

4.5 The Seifert-van Kampen theorem

4.5.1 A review of some facts in group theory

Given a family of groups $\{G_\alpha\}_{\alpha \in \mathcal{A}}$, consider all words whose "letters" are elements of these groups. Next we reduce these words by the following two operations:

(1) identify all identity elements.

(2) any consecutive letters belonging to the same group should be replaced by their product.

As such we obtain reduced words, in which consecutive letters belong to different groups. Two words are multiplied by juxtaposition followed by reduction. The common identity element acts as identity for this multiplication.

Definition. The *free product* of the groups G_α, $\alpha \in \mathcal{A}$, denoted by $*_{\alpha \in \mathcal{A}} G_\alpha$, is the group of reduced words with the multiplication defined above.

Note that the group G_α embeds naturally in the free product as the set of all words consisting of one letter which letter is in G_α.

Proposition 4.5.1. Let $G = *_\alpha G_\alpha$. Given group homomorphisms $h_\alpha : G_\alpha \to H$, where H is some group, there is a unique group homomorphism $h : G \to H$ such that $h|_{G_\alpha} = h_\alpha$. Conversely, if G contains all G_α and this property holds for any group H, then G is the free product.

Definition. The free product of several copies of $\mathbb{Z}$ is called a *free group*.

Free groups are universal in the following sense. Let G be an arbitrary group, and $\{g_\alpha\}$ a set of generators (which could be all elements of G). Then for each generator there is a group homomorphism $h_\alpha : \mathbb{Z} \to G$, $h_\alpha(k) = g_\alpha^k$. By the previous proposition there is a group homomorphism $h : *_\alpha \mathbb{Z} \to G$. If N is the kernel of h, then

$$*_\alpha \mathbb{Z}/N \simeq G.$$

Definition. The description of a group G as the quotient of a free group by a normal subgroup is called a *presentation* of G. The generators of the normal subgroup are called *relators*.

If the normal subgroup has finitely many elements, the presentation is called *finite*. A group that is finitely generated and finitely presented is usually written as

$$G = \langle g_1, g_2, \ldots, g_m \,|\, \beta_1, \beta_2, \ldots, \beta_n \rangle$$

where the g_i's are the generators, and the β_i's are the relators. Note that the β_i's are words in the g_i's.

Example.

$$\begin{aligned}
\mathbb{Z} &= \langle g \,|\quad \rangle \\
\mathbb{Z}_n &= \langle g \,|\, g^n \rangle \\
\mathbb{Z}^2 &= \left\langle g_1, g_2 \,|\, g_1 g_2 g_1^{-1} g_2^{-1} \right\rangle \\
K &= \left\langle a, b, c \,|\, a^2, b^2, c^2, abc^{-1}, bca^{-1}, cab^{-1} \right\rangle .
\end{aligned}$$

where K is the Klein 4-group, i.e. the group of symmetries of a rectangle.

As we will see below, the Seifert-van Kampen Theorem yields a group presentation of the fundamental group. There is a downside to this, there does not exist an algorithm that decides if two presentations give the same group. This is a theorem!

The groups that make up the free product do not intersect. But what if these groups overlap? We consider only the case of two groups, and make the following definition.

Definition. Given the groups G_1, G_2 and H, and the group homomorphisms $\phi_1 : H \to G_1$, $\phi_2 : H \to G_2$, we define the *free product with amalgamation* to be

$$G_1 *_H G_2 = (G_1 * G_2)/N$$

where N is the smallest normal subgroup of $G_1 * G_2$ containing all elements $\phi_1(h)\phi_2(h)^{-1}$, $h \in H$.

Note that $G_k \hookrightarrow G_1 * G_2 \to G_1 *_H G_2$ defines a natural map $j_k : G_k \to G_1 *_H G_2$, $k = 1, 2$.

One usually thinks of H as being a common subgroup of G_1 and G_2, and ϕ_1, ϕ_2 as being the inclusion homomorphisms.

Example 2. If

$$\begin{aligned}
G_1 &= \langle u_1, \ldots, u_k \,|\, \alpha_1, \ldots, \alpha_l \rangle \\
G_2 &= \langle v_1, \ldots, v_m \,|\, \beta_1, \ldots, \beta_n \rangle \\
H &= \langle w_1, \ldots, w_p \,|\, \gamma_1, \ldots, \gamma_q \rangle
\end{aligned}$$

then the amalgamated product is

$$G_1 *_H G_2 = \langle u_1, \ldots, u_k, v_1, \ldots, v_m \,|\, \alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_n, i_{1*}(w_1)i_{2*}(w_1)^{-1}, \ldots, i_{1*}(w_p)i_{2*}(w_p)^{-1}\rangle$$

The free product with amalgamation also has a universality property, stated in the following proposition.

Proposition 4.5.2. Given the groups G_1, G_2 and H, and the group homomorphisms $\phi_k : H \to G_k$, $k = 1, 2$, consider $G_1 *_H G_2$. Given a group G and group homomorphisms $\psi_k : G_k \to G$, $k = 1, 2$, such that $\psi_1 \circ \phi_1 = \psi_2 \circ \phi_2$, there is a unique group homomorphism $\Phi : G_1 *_H G_2 \to G$ such that $\Phi \circ j_k = \psi_k$, $k = 1, 2$. Moreover, G' is a group for which this property holds for every group G, then $G' = G_1 *_H G_2$.

4.5.2 The statement and proof of the Seifert-van Kampen theorem

Theorem 4.5.1. (Seifert-van Kampen theorem, modern version) Let $X = U_1 \cup U_2$ where U_1 and U_2 are open in X, such that $U_1, U_2, U_1 \cap U_2$ are path connected, and let $x_0 \in U_1 \cap U_2$. Consider the homomorphisms $i_{k*} : \pi_1(U_1 \cap U_2, x_0) \to \pi_1(U_k, x_0)$ and $j_{k*} : \pi_1(U_k, x_0) \to \pi_1(X, x_0)$ induced by the inclusion maps, $k = 1, 2$. If G is a group and $\phi_k : \pi_1(U_k, x_0) \to G$, $k = 1, 2$, are group homomorphisms such that $\phi_1 \circ i_{1*} = \phi_2 \circ i_{2*}$, then there is a unique homomorphism $\Phi : \pi_1(X, x_0) \to G$ such that $\Phi \circ j_{1*} = \phi_1$ and $\Phi \circ j_{2*} = \phi_2$.

The result states that if ϕ_1 and ϕ_2 coincide on $U_1 \cap U_2$ then they induce a homomorphism from $\pi_1(X, x_0)$ to G.

Proof. First let us prove the following result.

Lemma 4.5.1. Suppose $X = U_1 \cup U_2$ where U_1 and U_2 are open subsets of X such that $U_1 \cap U_2$ is path connected. Consider $x_0 \in U_1 \cap U_2$. If $i_k : U_k \to X$, $k = 1, 2$ are the inclusion maps then $i_{1*}(\pi_1(U_1, x_0))$ and $i_{2*}(\pi_1(U_2, x_0))$ generate $\pi_1(X, x_0)$.

Proof. Let f be a loop in X. Using the Lebesgue number theorem (Theorem 2.4.5), we can find a subdivision $a_0 = 0 < a_1 < \cdots < a_n = 1$ of $[0, 1]$ such that $f([a_l, a_{l+1}])$ lies entirely in one of the sets U_k. By joining consecutive intervals whose images lie in the same U_k, we may assume that for each l, $f([a_l, a_{l+1}])$ and $f([a_{l-1}, a_l])$ lie in different U_k's. It follows that $f(a_l) \in U_1 \cap U_2$ for $l = 1, 2, \ldots, n-1$. Also $f(a_0) = f(a_n) = x_0$. Consider paths γ_l, $l = 0, 1, 2, \ldots, n$ from x_0 to $f(a_l)$ inside $U_1 \cap U_2$ ($\gamma_0 = \gamma_n = e_{x_0}$). If we denote $f_l = \gamma_l * f|_{[a_l, a_{l+1}]} * \overline{\gamma_{l+1}}$, then

$$[f] = [f_1] * [f_2] * \cdots * [f_{n-1}].$$

Of course, $[f_l]$ is either in U_1 or U_2, which proves the lemma. □

Because of Lemma 4.5.1, Φ is completely determined if we know where it maps loops that lie entirely in U_1 or U_2. This proves uniqueness.

Let us prove existence. To avoid ambiguities, we use indices for homotopy equivalence classes, to specify in which space the homotopy equivalence takes place. So for a path f, $[f]_X$ stands for the homotopy class rel endpoints of f in X, while $[f]_{U_1}$ stands for the homotopy class rel endpoints of f in U_1.

1. We define first a map ρ from the set of loops that lie entirely in one of the U_k to G by

$$\rho(f) = \begin{cases} \phi_1([f]_{U_1}) & \text{if } f \text{ lies in } U_1 \\ \phi_2([f]_{U_2}) & \text{if } f \text{ lies in } U_2. \end{cases}$$

The map ρ is well defined since if f lies in $U_1 \cap U_2$, then

$$\phi_k([f]_{U_k}) = \phi_k i_{k*}([f]_{U_1 \cap U_2})$$

and we know that $\phi_1 i_{1*} = \phi_2 i_{2*}$. The map ρ satisfies the following conditions

(1) If $[f]_{U_k} = [g]_{U_k}$ for some k, then $\rho(f) = \rho(g)$.

(2) If both f and g lie in the same U_k, then $\rho(f * g) = \rho(f) \cdot \rho(g)$.

This second condition holds because ϕ_k, $k = 1, 2$ are homomorphisms.

2. Next, we extend ρ to the set of paths that lie entirely in U_1 or U_2. First consider for each $x \in X$ a path α_x from x_0 to x (if $x_0 = x$ take the constant path). For a path f from x to y define

$$\rho(f) = \rho(\alpha_x * f * \overline{\alpha_y}).$$

The conditions (1) and (2) are again satisfied. Indeed, if $[f]_{U_k} = [g]_{U_k}$, then $[\alpha_x * f * \overline{\alpha_y}] = [\alpha_x * g * \overline{\alpha_y}]$, with the homotopy being constant on α_x and α_y. For (2), notice that if f runs from x to y and g from y to z, then $\alpha_x * f * \overline{\alpha_y} * \alpha_y * g * \overline{\alpha_z}$ is homotopic relative to the endpoints (in U_k) to $\alpha_x * f * g * \overline{\alpha_z}$.

3. We extend ρ further to arbitrary paths in X. Removing the constraints about U_1 and U_2 we have the new conditions for the extended map

(1) $[f] = [g]$ implies $\rho(f) = \rho(g)$.

(2) $\rho(f * g) = \rho(f) * \rho(g)$ wherever $f * g$ is defined.

The trick is the same as in the lemma. Consider a subdivision of the interval $[0, 1]$ such that each interval of the subdivision is mapped entirely to one of the two sets U_1 or U_2. Using this subdivision, break f as $f_1 * f_2 * \cdots * f_n$ (f_j is the restriction of f to the j-th subinterval). Define

$$\rho(f) = \rho(f_1)\rho(f_2) \cdots \rho(f_n).$$

But, is this independent of the subdivision? Any subdivision with the given property can be transformed into any other by adding or subtracting points. So it is sufficient to show that $\rho(f)$ does not change if we add one point to the subdivision. The path f_j in the middle of whom the point was added now breaks into the paths $f_{j,1}$ and $f_{j,2}$. These paths belong to the same U_k, so we already know that $\rho(f_k) = \rho(f_{k,1}) * \rho(f_{k,2})$. And that's all we need. So the map is well defined. It is an extension of the old ρ because if the path lies entirely in one of the U_k's we can use the trivial partition.

Let us now check condition (1). Consider a homotopy $F : [0, 1] \times [0, 1] \to X$ between f and g (relative to the endpoints). Consider the subdivisions $0 = a_0 < a_1 < \cdots < a_m = 1$, $b_0 = 0 < b_1 < \cdots < b_n = 1$ of the interval $[0, 1]$ such that for each (p, q), $F([a_p, a_{p+1}] \times [b_q, b_{q+1}])$ lies entirely in one of U_k.

Take one of the slices $[0, 1] \times [b_q, b_{q+1}]$, and let $f' = F|_{[0,1]\times\{b_q\}}$, $g' = F|_{[0,1]\times\{b_{q+1}\}}$. If we show that $\rho(f') = \rho(g')$, then by going slice-by-slice we obtain that $\rho(f) = \rho(g)$.

Now we can further fix p and consider the functions

$$f''(p) = F|[0, a_p] \times \{b_q\} * F|\{a_p\} \times [b_q, b_{q+1}] * F|[a_p, 1] \times b_{q+1}$$

and

$$g''(p) = F|[0, a_{p+1}] \times \{b_q\} * F|\{a_{p+1}\} \times [b_q, b_{q+1}] * F|[a_{p+1}, 1] \times b_{q+1}.$$

Then we can write $f'' = f_1'' * f_2'' * \cdots f_p'' * \cdots * f_m''$ and $g'' = f_1'' * f_2'' * \cdots g_p'' * \cdots * f_m$. Then f_p'' and g_p'' are homotopic rel boundary inside the same U_k, and so $\rho(f'') = \rho(g'')$. Varying p we obtain $\rho(f') = \rho(g')$, and ultimately $\rho(f) = \rho(g)$.

Finally, let us check condition (2). For $f * g$ we can consider a subdivision that contains the endpoint of f and the starting point of g. Then

$$f * g = f_1 * f_2 * \cdots * f_m * g_1 * g_2 * \cdots * g_n$$

and using the definition we obtain that $\rho(f * g) = \rho(f) \cdot \rho(g)$.

4. We now define for each loop in X based at x_0,

$$\Phi([f]) = \rho(f).$$

Conditions (1) and (2) show that Φ is well defined and is a group homomorphism. Let us now show that $\Phi \circ j_{k*} = \phi_k$, $k = 1, 2$. If f is a loop in U_k, then

$$\Phi(j_{k*}([f]_{U_k})) = \Phi([f]) = \rho(f) = \phi_k([f]_{U_k}).$$

by definition. And we are done. □

Proposition 4.5.2 implies the following classical version of this theorem.

Theorem 4.5.2. (Seifert-van Kampen Theorem, classical version) If $X = U_1 \cap U_2$ with U_1, U_2 open and $U_1, U_2, U_1 \cap U_2$ path connected, and if $x_0 \in U_1 \cap U_2$ then

$$\pi_1(X, x_0) = \pi_1(U_1, x_0) *_{\pi_1(U_1 \cap U_2, x_0)} \pi_1(U_2, x_0),$$

where the amalgamation is defined via the maps induced by inclusions.

Corollary 4.5.1. Given the open subsets U_1, U_2 of X such that $U_1, U_2, U_1 \cap U_2$ are path connected, if

$$\begin{aligned} \pi_1(U_1, x_0) &= \langle u_1, \ldots, u_k \,|\, \alpha_1, \ldots, \alpha_l \rangle \\ \pi_1(U_2, x_0) &= \langle v_1, \ldots, v_m \,|\, \beta_1, \ldots, \beta_n \rangle \\ \pi_1(U_1 \cap U_2, x_0) &= \langle w_1, \ldots, w_p \,|\, \gamma_1, \ldots, \gamma_q \rangle \end{aligned}$$

then

$$\pi_1(X, x_0) = \left\langle u_1, \ldots, u_k, v_1, \ldots, v_m \,|\, \alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_n, i_{1*}(w_1)(i_{2*}(w_1)^{-1}, \ldots, i_{1*}(w_p)(i_{2*}(w_p)^{-1} \right\rangle$$

Corollary 4.5.2. Given the open subsets U_1, U_2 of X such that $U_1, U_2, U_1 \cap U_2$ are path connected, if $U_1 \cap U_2$ is simply connected then

$$\pi_1(X, x_0) = \pi_1(U_1, x_0) * \pi_1(U_2, x_0).$$

Corollary 4.5.3. Given the open subsets U_1, U_2 of X such that $U_1, U_2, U_1 \cap U_2$ are path connected, if U_2 is simply connected then

$$\pi_1(X, x_0) = \pi_1(U_1, x_0)/N$$

where N is the smallest normal subgroup of $\pi_1(U_1, x_0)$ containing the image of $i_{1*} : \pi_1(U_1 \cap U_2, x_0) \to \pi_1(U_1, x_0)$.

4.5.3 Fundamental groups computed using the Seifert-van Kampen theorem

In the following we will be using Proposition 4.2.6, which asserts that if one space is a deformation retract of another, then the two spaces have isomorphic fundamental groups.

Example 1. The fundamental group of a wedge of circles. The wedge of n circles is obtained by taking the quotient of their disjoint union by the map that identifies the points of coordinate 1 from each of these circles. The standard notation is $\bigvee_{i=1}^n S^1$.

Let us first consider the case $n = 2$. Take as U_1 and U_2 open subsets of $S^1 \bigvee S^1$ such that U_k contains the kth circle, $k = 1, 2$, and this circle is a deformation retract of it. Then $U_1 \cap U_2$ is a contractible space. Applying the Seifert-van Kampen theorem we conclude that

$$\pi_1(S^1 \bigvee S^1) = \pi_1(S^1) * \pi_1(S^1) = \mathbb{Z} * \mathbb{Z}.$$

Proceeding by induction, and taking the set U_1 to be an open neighborhood of $\bigvee_{i=1}^{n-1} S^1$ and U_2 an open neighborhood of the last circle, which are both deformation retract of the corresponding subsets, we deduce that

$$\pi_1\left(\bigvee_{i=1}^{n} S^1\right) = *_{i=1}^n \mathbb{Z}.$$

Example 2. The fundamental group of the complement in S^3 (or $\mathbb{R}^3$) of the trefoil knot.

$$S^3 = \{(z_1, z_2) \,|\, |z_1|^2 + |z_2|^2 = 1\} \subset \mathbb{C}^2.$$

The trefoil knot $K_{2,3}$ can be thought of as the curve

$$\phi(t) = \left(\frac{1}{\sqrt{2}}e^{4\pi it}, \frac{1}{\sqrt{2}}e^{6\pi it}\right).$$

This curve lies on the torus

$$S^1 \times S^1 = \left\{\left(\frac{1}{\sqrt{2}}e^{2\pi it}, \frac{1}{\sqrt{2}}e^{2\pi is}\right) \,|\, t, s \in [0, 1]\right\}.$$

The torus separates the solid tori $T_1 = \{(z_1, z_2) \,|\, |z_1| \geq |z_2|\}$ and $T_2 = \{(z_1, z_2) \,|\, |z_1| \leq |z_2|\}$. Consider two open sets U_1 and U_2 such that $T_k \backslash K_{3,2}$ is a deformation retract of U_k, $k = 1, 2$. Note that the annulus $(S^1 \times S^1)\backslash K_{3,2}$ is a deformation retract of $U_1 \cap U_2$.

We have

$$\pi_1(U_k) = \pi_1(T_k\backslash K_{3,2}) = \pi_1(T_k) = \pi_1(S^1) = \mathbb{Z}, \quad k = 1, 2,$$
$$\pi_1(U_1 \cap U_2) = \pi_1((S^1 \times S^1)\backslash K_{3,2}) = \pi_1(\text{annulus}) = \mathbb{Z}.$$

Note also that the inclusion $U_1 \cap U_2 \hookrightarrow U_1$ induces at the level of the fundamental group the map $\mathbb{Z} \to \mathbb{Z}$, $m \to 3m$ (this is because the generator of the fundamental group of the annulus "wraps" three times around the torus in the direction of the generator of the fundamental group of T_1). Similarly the inclusion $U_1 \cap U_2 \hookrightarrow U_2$ induces at the level of the fundamental groups the map $\mathbb{Z} \to \mathbb{Z}$, $m \to 2m$.

We obtain for the fundamental group of the complement of the trefoil knot the presentation

$$\pi_1(S^3\backslash K_{3,2}) = \langle x, y \,|\, x^3 = y^2 \rangle$$

Note that the complement of the unknot has a circle as a deformation retract. Consequently, its fundamental group is $\mathbb{Z}$. Let us check that $\pi_1(S^3\backslash K_{3,2}) \neq \mathbb{Z}$. Indeed, if we consider the group G of symmetries of a regular hexagon generated by the rotation ρ about the center by $120°$ and the reflection σ about one main diagonal (the dihedral group), then because

$$\rho^3 = \sigma^2 = \text{ identity}$$

there is a group epimorphism of $\pi_1(S^3\backslash K_{3,2})$ onto G. But G is non-commutative, hence so is $\pi_1(S^3\backslash K_{3,2})$, showing that this group is not $\mathbb{Z}$. This proves that the trefoil knot is knotted.

Proposition 4.5.3. (The addition of a disk) Let X be a space, and $g : S^1 \to X$ a continuous map. Consider the space $X \cup_g \overline{B^2}$ obtained by factoring the disjoint union of X with the unit disk by the equivalence relation $e^{2\pi it} \sim g(e^{2\pi it})$. Then the map $i_* : \pi_1(X, x_0) \to \pi_1(X \cup_g \overline{B^2}, x_0)$ induced by inclusion is surjective and its kernel is the normal subgroup of $\pi_1(X, x_0)$ containing the image of $g_*|S^1$.

Proof. The fact that i_* is surjective follows from the fact that every loop is homotopic to one that does not intersect the disk. The proof of this fact follows the idea from the proof of Theorem 4.2.3. One approximates the loop by a piecewise linear one, which is homotopic to it and misses at least one point p of the disk, then one pushes the loop away from p until it leaves the disk.

Let U be the open set $X \cup_g (\overline{B^2}\backslash\{0\}$ and $V = \frac{1}{2}B^2$. Then U and V are path connected, and $U \cap V = \frac{1}{2}B^2\backslash\{0\}$ is also path connected. Moreover, X is a deformation retract of U (just push the punctured disk towards the unit circle), V is simply connected, and $U \cap V$ is homotopically equivalent to a circle. By the Seifert-van Kampen theorem we have

$$\begin{aligned}\pi_1(X \cup_g \overline{B^2}, x_0) &= \pi_1(U, x_0) *_{\pi_1(U\cap V, x_0)} \pi_1(\overline{B^2}, x_0)\\ &= \pi_1(X, x_0) *_{\pi_1(S^1, x_0)} \{0\}.\end{aligned}$$

The amalgamation is defined by the relations $g_*([f]) = 0$ for all $[f] \in \pi_1(S^1, x_0)$ because $[f]$ is mapped to 0 by the inclusion of the circle into the disk. □

Example 3. Given a group presentation with finitely many relators and finitely many generators, there is a topological space that has this as its fundamental group.

Here is how to construct this space. Let

$$G = \langle g_1, g_2, \ldots, g_m \,|\, \beta_1, \beta_2, \ldots, \beta_n\rangle\,.$$

Consider the wedge of n circles, and identify the generator of the jth circle with g_j, $j = 1, 2, \ldots, m$. Now for each relator β_k, consider a loop f_k in the wedge of circles that represents it. For each k, take a closed unit disk and glue it to the wedge of circles by the equivalence relation $e^{2\pi it} = f_k(t)$. As a corollary of Proposition 4.5.3, the result of all these operations is a topological space whose fundamental group is G.

4.5.4 The construction of closed oriented surfaces and the computation of their fundamental groups

We will consider just the case of the surfaces shown in Figure 4.2, which, are all closed compact orientable surfaces. These are the sphere with zero, one, two, three, etc. handles. The sphere with g handles will be called the genus g (closed) surface and will be denoted by Σ_g. We will see later that no two are homeomorphic, not even homotopy equivalent.

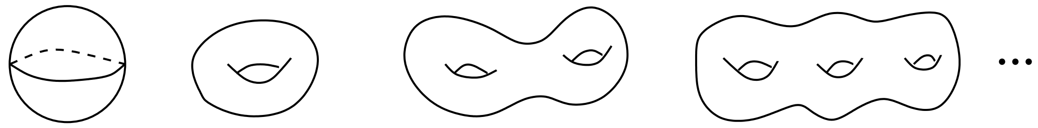

Figure 4.2:

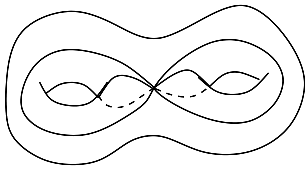

Figure 4.3:

The first two, the sphere and the torus, are well understood. For all the others the same procedure will be applied: They will be obtained by adding a disk to a wedge of circles. We will do this in detail for the third surface on the list, the sphere with two handles, the others being similar.

Consider the wedge of four circles embedded in the surface as shown in Figure 4.3. Exactly how a wedge of two circles cuts the torus into a rectangle, this wedge of four circles, when removed from the surfaces leaves an octagon. Said differently, the surface is made out of two tori glued along a circle. The first torus is cut by two of the circles into a quadrilateral, the second is cut into a quadrilateral as well. But there are also the holes in the tori, which when cut open become segments. So there are two pentagons glued along a side, and the result is an octagon. The gluing pattern on the boundary of the octagon that yields back the surface is shown in Figure 4.4 (note the gluing patterns of the two tori placed one after the other).

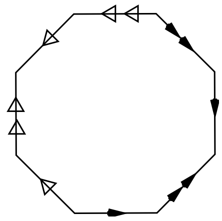

Figure 4.4:

In this figure the four circles are marked with one black arrow, two black arrows, one white arrow, respectively two white arrows. If the generators of the fundamental groups of the four circles are a_1, b_1, a_2, b_2, then when gluing the disk (see Figure 4.4) we introduce the relation $a_1b_1a_1^{-1}b_1^{-1}a_2b_2a_2^{-1}b_2^{-1}$ =identity. We obtain that the fundamental group of the sphere with two handles Σ_2 is

$$\pi_1(\Sigma_2) = \left\langle a_1, b_1, a_2, b_2 \,|\, a_1b_1a_1^{-1}b_1^{-1}a_2b_2a_2^{-1}b_2^{-1} \right\rangle$$

In algebra the expression $xyx^{-1}y^{-1}$ is called the commutator of x and y and is denoted by $[x, y]$. In this case the relator can be written as $[a_1, b_1][a_2, b_2]$.

Let us now show a different approach for computing the fundamental group of the genus 3 surface. This can be obtained by gluing a genus 2 surface with one hole, $\Sigma_{2,1}$ to a torus with one hole $\Sigma_{1,1}$ (see Figure 4.5). The two surfaces can be thought as being open (so the boundary

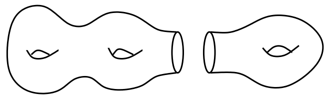

Figure 4.5:

circle is missing), and the gluing is with an overlap along an annulus, so that the conditions of the Seifert-van Kampen theorem are satisfied with $U_1 = \Sigma_{2,1}$ and $U_2 = \Sigma_{1,1}$.

The torus with one hole has a deformation retract which is the wedge of two circles. There are two ways to see this, shown in Figure 4.6, either directly on the torus, where the hole can be "inflated" to the point where the result is a regular neighborhood of the union of the meridian and the longitude, or by looking at the planar representation of the torus with a hole in the middle.

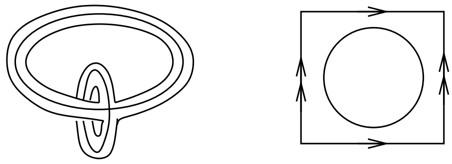

Figure 4.6:

For a genus 2 surface the deformation retract consists of the wedge of four circles shown in Figure 4.7. A way to see this is to recall Figure 4.4, this time with a hole in the middle, and to increase continuously the size of the hole until only the boundary remains.

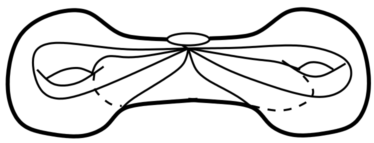

Figure 4.7:

Now we are in position to apply the Seifert-van Kampen theorem. Since $U_1 \cap U_2$ is an annulus, its fundamental group is $\mathbb{Z}$. In view of Example 1 in § 4.5.3, the fundamental group of the genus 3 surface Σ_3 is

$$\pi_1(\Sigma_3) = (\mathbb{Z} * \mathbb{Z}) *_{\mathbb{Z}} (\mathbb{Z} * \mathbb{Z} * \mathbb{Z} * \mathbb{Z}).$$

Let the generators of these $\mathbb{Z}$'s be respectively $a_1, b_1, a_2, b_2, a_3, b_3$.

Now let us understand how the amalgamation takes place. The generator of the fundamental group of the annulus becomes the boundary of the square in the first case, and the boundary of the octagon in the second. By choosing appropriately a_1 and b_1, we can make the boundary of the square to correspond to the element $b_1a_1b_1^{-1}a_1^{-1}$ in the wedge of two circles. Similarly the boundary of the octagon can be made to correspond to the element $a_2b_2a_2^{-1}b_2^{-1}a_3b_3a_3^{-1}b_3^{-1}$, and the Seifert-van Kampen theorem sets these equal. As a result we obtain

$$\pi_1(\Sigma_3) = \langle a_1, b_1, a_2, b_2, a_3, b_3 \mid [a_1, b_1][a_2, b_2][a_3, b_3] \rangle .$$

In fact, the genus g surface Σ_g, obtained by adding g handles to the sphere, can also be obtained by a pairwise gluing of the sides of a $4g$-gon, and hence by adding a disk to a wedge of $2g$ circles.

As a result, we obtain

$$\pi_1(\Sigma_g) = \langle a_1, b_1, a_2, b_2, \ldots, a_g, b_g \mid [a_1, b_1][a_2, b_2] \cdots [a_g, b_g] \rangle \, .$$

The question is how to distinguish these groups. The trick is to abelianize them, that is to factor them be the largest normal subgroups generated by all relations of the form $[x, y]$. We obtain $\pi_1(\Sigma_g) = \mathbb{Z}^{2g}$. The abelianizations of the fundamental groups are themselves very important topological invariants; they are the first homology groups. These will be studies in the next chapter.

Chapter 5

Homology

The homotopy groups, namely the fundamental group and its higher dimensional generalizations defined in the previous chapter, have the downsides that they are difficult to compute (as is the case of the higher dimensional homotopy groups) or they are hard to distinguish from each other (as is the case for the fundamental group obtained as a group presentation). The way around these problems is to introduce a different family of groups, which are much easier to compute and moreover which are abelian, thus very easily distinguishable. The idea behind this construction goes back to H. Poincaré. While in the case of homotopy theory the objects where loops (i.e. images of circles) or images of higher dimensional spheres inside the topological space and the equivalence between them was defined by homotopy, in the new situation the objects are cycles, which are images of manifolds of same dimension inside the same topological space, two cycles being equivalent if between them runs the image of a manifold one dimension higher.

5.1 Simplicial homology

The above idea for constructing a homology theory can be formalized easily in a combinatorial setting if we triangulate the topological space and all cycles. We follow the point of view of Allen Hatcher's book on Algebraic Topology and construct the simplicial homology based on Δ-complexes (originally called semisimplicial complexes by their inventors Eilenberg and Zilber), mainly due to the fact that for the classical simplicial complexes the computations are almost impossible.

In this chapter all spaces are assumed to be compact.

5.1.1 Δ-complexes

At the heart of homology theory lies the notion of a simplex.

Definition. Let $v_0, v_1, \dots, v_n$ be points in $\mathbb{R}^N$, $N > n$ that do not lie in an affine subspace of dimension n or less. The *n-simplex* with vertices $v_0, v_1, \dots, v_n$ is the set

$$\sigma^n = \langle v_0, v_1, \dots, v_n \rangle = \{t_0v_0 + t_1v_1 + \cdots + t_nv_n \in \mathbb{R}^N \mid t_i \geq 0, t_0 + t_1 + \cdots + t_n = 1\}.$$

In other words an n-simplex is the convex hull of $n+1$ points that do not lie in an n-dimensional subspace. The list of coefficients $(t_0, t_1, \dots, t_n)$ of a point $t_0v_0 + t_1v_1 + \cdots + t_nv_n$ in the simplex are called *barycentric coordinates.*

Example 1. A 1-simplex is a line segment, a 2-simples is a triangle (together with its interior), a three simplex is a solid tetrahedron.

We convene that the ordering of the vertices of the simplex defines an orientation, with two orientations being the same if one gets from one to the other by an even permutation of the vertices, and two orderings are of opposite orientations otherwise.

Example 2. $\langle v_0, v_1, v_2\rangle = \langle v_1, v_2, v_0\rangle = -\langle v_0, v_2, v_1\rangle$.

If we remove several vertices of a simplex, the remaining vertices determine a simplex of smaller dimension, which we will call a *sub-simplex* of the original simplex. When removing one vertex, the remaining n vertices determine a sub-simplex called *face*. The union of the faces is the boundary of the simplex. More precisely, we make the following definition.

Definition. Given an n-simplex $\sigma^n = \langle v_0, v_1, \ldots, v_n\rangle$, its *boundary* is the formal sum

$$\partial\sigma^n = \sum_{j=0}^{n}(-1)^j \langle v_0, v_1, \ldots, \hat{v_j}, \ldots, v_n\rangle$$

where the hat means that the respective vertex is removed.

Example 3. $\partial\langle v_0, v_1\rangle = \langle v_1\rangle - \langle v_0\rangle$.

The points of σ^n with all barycentric coordinates positive form the *interior* of σ^n, denoted $\mathrm{Int}(\sigma^n)$. When one coordinate is zero, the point is on a face. The zero dimensional simplex consists only of its interior.

Definition. A Δ-*complex* K consists of a topological space X and a finite collection of continuous maps $\phi_\alpha : \sigma^{n_\alpha} \to X$ from various simplices to X with the following properties:
(1) The restriction of ϕ_α to the interior of σ^{n_α} is one-to-one, and each point of X is the image of exactly one such restriction.
(2) Each restriction of a map ϕ_α to a face is another map ϕ_β.
(3) A set U is open in X if and only if $\phi^{-1}(U)$ is open for each α.

An alternative way of saying this is that X is the quotient of the disjoint union of several simplices, with the equivalence relation defined by gluing maps homeomorphisms that identify one simplex to another. We can assume, for convenience, that these maps are *linear isometries* (by choosing simplices of the right size).

Given a Δ-complex K, the space X from the definition is usually denoted by $|K|$ and is called the *realization* of K. Conversely K is called a Δ-complex structure on X. Note that a topological space can have many different Δ-complex structures.

Remark 5.1.1. The realization $|K|$ of a Δ-complex is a *metric space* with the distance between two points being the shortest length of a piece-wise linear path that connects the points.

In practice, one usually identifies the image of the map ϕ_α with the simplex σ^{n_α}. This means that a Δ-complex is a topological space obtained by gluing finitely many simplices by homeomorphisms between some of their subsimplices. By contrast, in the older notion of a simplicial complex two simplices can share at most one sub-simplex. As such, a Δ-structure on a topological space requires fewer simplices than a simplicial structure, and consequently this concept is more useful for explicit computations.

Example 4. The realization of the 2-dimensional torus, the projective plane, and the Klein bottle as Δ-complexes are shown in Figure 5.1.

Example 4. An n-dimensional sphere can be realized as a Δ-complex by gluing two n-simplices by a (orientation reversing) homeomorphism between their boundaries.

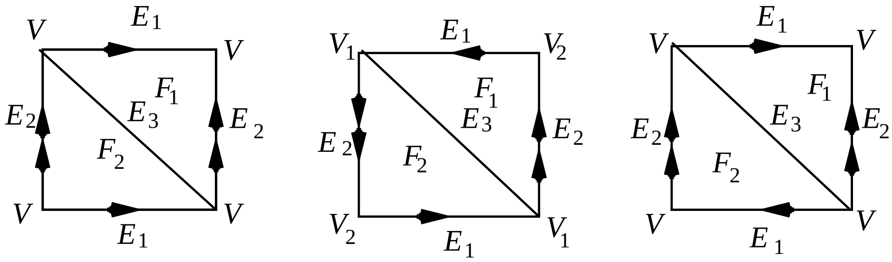

Figure 5.1:

5.1.2 The definition of simplicial homology

For a Δ-complex X, let $C_n(X)$ be the free abelian group with basis the n-dimensional simplices of X. The elements of $C_n(X)$ are called n-chains. For each simplex recall the boundary map,

$$\partial \langle v_0, v_1, \ldots, v_n \rangle = \sum_{j=0}^{n} (-1)^j \langle v_0, v_1, \ldots, \hat{v_j}, \ldots, v_n \rangle .$$

Extend this map to a group homomorphism $\partial_n : C_n(X) \to C_{n-1}(X)$, for each $n \geq 1$. We obtain the sequence of group homomorphisms

$$\cdots \xrightarrow{\partial_{n+1}} C_n(X) \xrightarrow{\partial_n} C_{n-1}(X) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \to 0.$$

Proposition 5.1.1. For each n, $\partial_{n-1}\partial n = 0$.

Proof. We have

$$\partial_{n-1}\partial_n \langle v_0, v_1, \ldots, v_n \rangle = \sum_{j=1}^{n} (-1)^j \partial_{n-1} \langle v_0, v_1, \ldots, \hat{v_j}, \ldots v_n \rangle$$
$$= \sum_{i<j} (-1)^j (-1)^i \langle v_0, v_1, \ldots, \hat{v_i}, \ldots, \hat{v_j}, \ldots, v_n \rangle + \sum_{i>j} (-1)^j (-1)^{i-1} \langle v_0, v_1, \ldots, \hat{v_j}, \ldots, \hat{v_i}, \ldots, v_n \rangle .$$

Note that when exchanging i and j in the second sum, we obtain the negative of the first sum. Hence the result is zero, as desired. □

As a corollary, we have what is called a *chain complex*

$$\cdots \xrightarrow{\partial_{n+1}} C_n(X) \xrightarrow{\partial_n} C_{n-1}(X) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \to 0,$$

in which Im $\partial_n \subset$ Ker ∂_{n-1} for all n.

Definition. We let $B_n(X) = \text{Im } \partial_{n+1}$ and $Z_n(X) = \text{Ker } \partial_n$. The elements of $B_n(X)$ are called *boundaries*, the elements of $Z_n(X)$ are called cycles. The group $H_n(X) = Z_n(X)/B_n(X)$ is called the *n-th simplicial homology group* of X. The cosets of $B_n(X)$ are called homology classes.

For finite Δ-complexes the chain groups $C_n(X)$ are finitely generated free abelian groups, and consequently $B_n(X)$ and $Z_n(X)$ are finitely generated free abelian groups. It follows that $H_n(X)$ is a finitely generated abelian group; in fact the generators of $Z_n(X)$ yield generators for $H_n(X)$, and the generators of $B_n(X)$ yield the relators of $H_n(X)$. Thus in computations we are led again, as for the fundamental group, to a presentation of $H_n(X)$, but in this case to the presentation of an abelian group. A little linear algebra allows us to compute this group precisely, as we will explain in the next section.

5.1.3 Some facts about abelian groups

The fundamental theorem of finitely generated abelian groups can be stated in two equivalent forms

Version A: Every finitely generated abelian group is of the form

$$G = \mathbb{Z}^n \oplus \mathbb{Z}_{p_1^{n_1}} \oplus \mathbb{Z}_{p_2^{n_2}} \oplus \cdots \oplus \mathbb{Z}_{p_k^{n_k}},$$

where $n, k \geq 0$, $n_1, n_2, \ldots, n_k \geq 1$ and $p_1, p_2, \ldots, p_k$ are prime numbers, these numbers are uniquely defined (up to reordering) by G.

Version B: Every finitely generated abelian group is of the form

$$G = \mathbb{Z}^n \oplus \mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \oplus \cdots \oplus \mathbb{Z}_{m_k}$$

where $n, k \geq 0$, m_1 divides m_2, m_2 divides m_3, ..., m_{k-1} divides m_k, and these numbers are uniquely defined by G.

Note that a finitely generated abelian group is the sum of a free abelian group and a finite abelian group. The later is called the torsion. The number of copies of $\mathbb{Z}$ in the free part is called the rank.

In what follows we will be interested in abelian groups that arise as kernels and images of homomorphisms between free abelian groups and as quotients of such groups. The kernels, images, and quotients can be computed as in the case of vector spaces, using row operations on matrices. These operations are:

1. A row is multiplied by -1.
2. Two rows are exchanged.
3. A row is added to or subtracted from another row.

5.1.4 The computation of the homology groups for various spaces

In this section we will call zero-dimensional simplices vertices, one-dimensional simplices edges, two-dimensional simplices faces, and 3-dimensional simplices (solid) tetrahedra. All computations will be performed for a certain realization of the given space as a Δ-complex; we will prove in § 5.2.4 that the homology groups are independent of this realization.

Example 1. The circle. We can realize the circle as a Δ-complex by considering the circle itself to be an edge E with the two endpoints at the vertex $V = 1$. The corresponding chain complex is

$$0 \longrightarrow \mathbb{Z}E \xrightarrow{\partial_1} \mathbb{Z}V \longrightarrow 0,$$

with $\delta_1 E = V - V = 0$. Hence

$$H_1(S^1) = Z_1(S^1) = \mathbb{Z}$$

and

$$H_0(S^1) = Z_0(S^1)/B_0(S^1) = \mathbb{Z}V/\{0\} = \mathbb{Z}.$$

Example 2. The projective plane. $\mathbb{R}P^2$ can be obtained from the upper hemisphere of S^2 by gluing its boundary S^1 by the antipodal map. Using this we can realize $\mathbb{R}P^2$ as a Δ-complex as shown in Figure 5.2. Here the orientation is given by the arrows, which specify the order of the vertices (for example $F_1 = < V_1, V_2, V_2 >$ with the first V_2 being the one on the upper side.

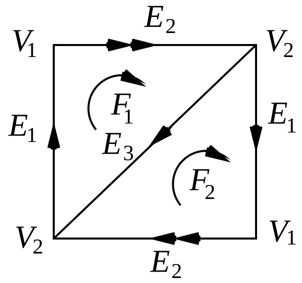

Figure 5.2:

The chain complex is

$$0 \longrightarrow \mathbb{Z}F_1 \oplus \mathbb{Z}F_2 \xrightarrow{\partial_2} \mathbb{Z}E_1 \oplus \mathbb{Z}E_2 \oplus \mathbb{Z}E_3 \xrightarrow{\partial_1} \mathbb{Z}V_1 \oplus \mathbb{Z}V_2 \longrightarrow 0.$$

with the (nontrivial) boundary maps given by

$$\partial_2 F_1 = E_1 + E_2 + E_3, \quad \partial_2 F_2 = E_1 + E_2 - E_3$$

and

$$\partial_1 E_1 = V_1 - V_2, \quad \partial_1 E_2 = V_2 - V_1, \quad \partial_1 E_3 = V_2 - V_2 = 0.$$

Computation of $H_0(\mathbb{R}P^2)$. The kernel of the zero map ∂_0 is $Z_0(\mathbb{R}P^2) = \mathbb{Z}V_1 \oplus \mathbb{Z}V_2$, with basis V_1, V_2. The image of ∂_1 is $B_0(\mathbb{R}P^2) = \mathbb{Z}(V_1 - V_2)$, with basis $V_1 - V_2$. The coordinates of the basis vectors of $B_0(\mathbb{R}P^2)$ in the basis of $Z_0(\mathbb{R}P^2)$ form a matrix which should be brought in reduced row-echelon form with operations over the integers. This matrix is the row matrix $[1, -1]$, which is already in row-echelon form. The structure of $H_0(\mathbb{R}P^2) = Z_0(\mathbb{R}P^2)/B_0(\mathbb{R}P^2)$ is read in this matrix. The leading 1 in this matrix says that one of the two $\mathbb{Z}$ in $Z_0(\mathbb{R}P^2) = \mathbb{Z} \oplus \mathbb{Z}$ is factored out. Hence

$$H_0(\mathbb{R}P^2) = \mathbb{Z}.$$

Computation of $H_1(\mathbb{R}P^2)$. A little linear algebra shows that the kernel of ∂_1, $Z_1(\mathbb{R}P^2)$, is a free abelian group with basis consisting of $w_1 = E_1 + E_2$ and $w_2 = E_3$. On the other hand the image of ∂_2 is generated by $E_1 + E_2 + E_3$ and $E_1 + E_2 - E_3$. In the basis w_1, w_2, these generators are $w_1 + w_2$ and $w_1 - w_2$. The coefficient matrix is

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

In reduced row-echelon form this is

$$\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}.$$

Thus from $Z_1(\mathbb{R}P^2) = \mathbb{Z} \oplus \mathbb{Z}$, the first $\mathbb{Z}$ is factored out, while the second is factored by a $2\mathbb{Z}$. Hence

$$H_1(\mathbb{R}P^2) = (\mathbb{Z}/\mathbb{Z}) \oplus (\mathbb{Z}/2\mathbb{Z}) = \mathbb{Z}_2.$$

Note that $H_1(\mathbb{R}P^2) = \pi_1(\mathbb{R}P^2)$!

Computation of $H_2(\mathbb{R}P^2)$. To find $Z_2(\mathbb{R}P^2)$ we need to determine which linear combinations with integer coefficients of F_1 and F_2 are mapped to zero by ∂_2. This means that we should have

$$n_1(E_1+E_2+E_3)+n_2(E_1+E_2-E_3)=0\in\mathbb{Z}E_1\oplus\mathbb{Z}E_2\oplus\mathbb{Z}E_3.$$

This can be rewritten as

$$(n_1+n_2)E_1+(n_1+n_2)E_2+(n_1-n_2)E_3=0.$$

For this to happen, n_1+n_2 and n_1-n_2 should both be zero, hence $n_1=n_2=0$. Thus $Z_2(\mathbb{R}P^2)=\{0\}$, and consequently

$$H_2(\mathbb{R}P^2)=\{0\}.$$

In conclusion

$$H_n(\mathbb{R}P^2)=\begin{cases}\mathbb{Z} & \text{if } n=0\\ \mathbb{Z}_2 & \text{if } n=1\\ 0 & \text{if } n\geq 2.\end{cases}$$

Example 3. The torus. A realization of the torus as a Δ-complex is shown in Figure 5.3.

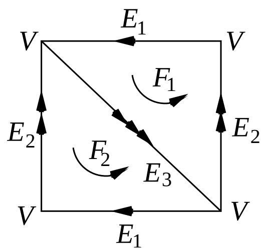

Figure 5.3:

We have the chain complex

$$0\longrightarrow\mathbb{Z}F_1\oplus\mathbb{Z}F_2\xrightarrow{\partial_2}\mathbb{Z}E_1\oplus\mathbb{Z}E_2\oplus\mathbb{Z}E_3\xrightarrow{\partial_1}\mathbb{Z}V\longrightarrow 0.$$

The boundary maps are

$$\partial_2F_1=E_1+E_2+E_3,\quad \partial_2F_2=-E_1-E_2-E_3,$$
$$\partial_1E_1=\partial_1E_2=\partial_1E_3=V-V=0.$$

Computation of $H_0(S^1\times S^1)$. We have $B_0(S^1\times S^1)=\text{Im }\partial_1=\{0\}$ and $Z_0(S^1\times S^1)=\mathbb{Z}V$.
Hence

$$H_0(S^1\times S^1)=\mathbb{Z}$$

Computation of $H_1(S^1\times S^1)$. We have

$$B_0(S^1\times S^1)=\text{Im }\partial_2=\mathbb{Z}(E_1+E_2+E_3).$$

Also

$$Z_0(S^1\times S^1)=\text{Ker }\partial_1=\mathbb{Z}E_1\oplus\mathbb{Z}E_2\oplus\mathbb{Z}E_3.$$

The matrix $[1, 1, 1]$ is already in reduced row-echelon form, with leading term 1. Hence one of the $\mathbb{Z}$'s in $\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$ is factored out. We thus obtain

$$H_1(S^1 \times S^1) = \mathbb{Z} \oplus \mathbb{Z}$$

Note again that $H_1(S^1 \times S^1) = \pi_1(S^1 \times S^1)$.
Computation of $H_2(S^1 \times S^1)$. If we write $\partial_2(n_1F_1 + n_2F_2) = 0$ we obtain

$$(n_1 - n_2)(E_1 + E_2 + E_3) = 0.$$

Hence $n_1 = n_2$. We obtain

$$H_2(S^1 \times S^1) = Z_2(S^1 \times S^1) = \mathbb{Z}(F_1 + F_2) \simeq \mathbb{Z}.$$

In conclusion

$$H_n(S^1 \times S^1) = \begin{cases} \mathbb{Z} & \text{if } n = 0, 2 \\ \mathbb{Z} \oplus \mathbb{Z} & \text{if } n = 1 \\ 0 & \text{if } n > 2. \end{cases}$$

Example 4. The 3-dimensional sphere. Like the 2-dimensional sphere can be obtained by gluing together two disks along their boundaries, the 3-dimensional sphere can be obtained by gluing two balls along their boundaries. The balls can be realized as two tetrahedra T_1 and T_2. Thus S^3 can be realized as a Δ-complex with two 3-dimensional simplices T_1 and T_2 that have the same vertices, edges, and faces. Since a tetrahedron has four faces, six edges, and four vertices, we have a chain complex

$$0 \longrightarrow \mathbb{Z}T_1 \oplus \mathbb{Z}T_2 \xrightarrow{\partial_3} \mathbb{Z}F_1 \oplus \mathbb{Z}F_2 \oplus \mathbb{F}_3 \oplus \mathbb{Z}F_4 \xrightarrow{\partial_2} \mathbb{Z}E_1 \oplus \mathbb{Z}E_2 \oplus \mathbb{Z}E_3 \oplus \mathbb{Z}E_4 \oplus \mathbb{Z}E_5 \oplus \mathbb{Z}E_6$$
$$\xrightarrow{\partial_1} \mathbb{Z}V_1 \oplus \mathbb{Z}V_2 \oplus \mathbb{Z}V_3 \oplus \mathbb{Z}V_4 \longrightarrow 0.$$

To exhibit the boundary maps we have to be a little bit more precise. As such we set

$$T_1 = \langle V_1, V_2, V_3, V_4 \rangle, T_2 = \langle V_1, V_3, V_2, V_4, \rangle$$
$$F_1 = \langle V_2, V_3, V_4 \rangle, F_2 = \langle V_1, V_3, V_4 \rangle, F_3 = \langle V_1, V_2, V_4 \rangle, F_4 = \langle V_1, V_2, V_3 \rangle$$
$$E_1 = \langle V_1, V_2 \rangle, E_2 = \langle V_1, V_3 \rangle, E_3 = \langle V_2, V_3 \rangle, E_4 = \langle V_3, V_4 \rangle, E_5 = \langle V_2, V_4 \rangle, E_6 = \langle V_1, V_4 \rangle.$$

The boundary maps are

$$\begin{aligned} \partial_3 T_1 = \partial_3 \langle V_1, V_2, V_3, V_4 \rangle &= \langle V_2, V_3, V_4 \rangle - \langle V_1, V_3, V_4 \rangle + \langle V_1, V_2, V_4 \rangle - \langle V_1, V_2, V_3 \rangle \\ &= F_1 - F_2 + F_3 - F_4 \\ \partial_3 T_2 = \partial_3 \langle V_1, V_3, V_2, V_4 \rangle &= \langle V_3, V_2, V_4 \rangle - \langle V_1, V_2, V_4 \rangle + \langle V_1, V_3, V_4 \rangle - \langle V_1, V_3, V_2 \rangle \\ &= -F_1 - F_3 + F_2 + F_4, \end{aligned}$$

$$\begin{aligned} \partial_2 F_1 &= \partial_2 \langle V_2, V_3, V_4 \rangle = \langle V_3, V_4 \rangle - \langle V_2, V_4 \rangle + \langle V_2, V_3 \rangle = E_4 - E_5 + E_3 \\ \partial_2 F_2 &= \partial_3 \langle V_1, V_3, V_4 \rangle = \langle V_3, V_4 \rangle - \langle V_1, V_4 \rangle + \langle V_1, V_3 \rangle = E_4 - E_6 + E_2 \\ \partial_2 F_3 &= \partial_3 \langle V_1, V_2, V_4 \rangle = \langle V_2, V_4 \rangle - \langle V_1, V_4 \rangle + \langle V_1, V_2 \rangle = E_5 - E_6 + E_1 \\ \partial_2 F_4 &= \partial_3 \langle V_1, V_2, V_3 \rangle = \langle V_2, V_3 \rangle - \langle V_1, V_3 \rangle + \langle V_1, V_2 \rangle = E_3 - E_2 + E_1, \end{aligned}$$

and

$$\partial_1 E_1 = V_2 - V_1, \quad \partial_1 E_2 = V_3 - V_1, \quad \partial_1 E_3 = V_3 - V_2,$$
$$\partial_1 E_4 = V_4 - V_3, \quad \partial_1 E_5 = V_4 - V_2, \quad \partial_1 E_6 = V_4 - V_1.$$

The computation of $H_0(S^3)$. It is clear that $Z_0(S^3) = \mathbb{Z}V_1 \oplus \mathbb{Z}V_2 \oplus \mathbb{Z}V_3 \oplus \mathbb{Z}V_4$.

To find $B_0(S^3)$, we arrange the coefficients of the images of $E_1, E_2, ..., E_6$ through ∂_1 in a matrix, then we bring this matrix in row-echelon form. This is the same as the transpose of the matrix of ∂_1, namely

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix}.$$

By performing row operations we obtain successively

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & -1 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \to \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This means that $B_0(S^3) = \mathbb{Z}^3$, and that when factoring $Z_0(S^3)$ by $B_0(S^3)$ three of the $\mathbb{Z}$ are factored out of $\mathbb{Z}^4$. We thus have

$$H_0(S^3) = \mathbb{Z}.$$

The computation of $H_1(S^3)$. Let us find the kernel of ∂_1. The matrix of this transformation is

$$\begin{bmatrix} -1 & -1 & 0 & 0 & 0 & -1 \\ 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

which after performing a number of row operations becomes

$$\begin{bmatrix} 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The kernel of this matrix is a free abelian group with basis $[1,-1,1,0,0,0]$, $[1,-1,0,-1,1,0]$, and $[0,1,0,1,0,-1]$, that is $E_1 - E_2 + E_3$, $E_1 - E_2 - E_4 + E_5$, and $E_2 + E_4 - E_6$.

On the other hand, the image of ∂_2 is generated by $E_3+E_4-E_5$, $E_2+E_4-E_6$, $E_1+E_5-E_6$, and $E_1-E_2+E_3$. Of course the standard approach is to write these four vectors in the basis of $Z_1(S^3) = \ker\partial_1$, and then decide what subspace they span. But we can see right away that the first and the last basis elements appear in this list, while $E_1 - E_2 - E_4 + E_5 = (E_1 - E_2 + E_3) - (E_3 + E_4 - E_5)$. Hence $B_1(S^3) = Z_1(S^3)$, and hence

$$H_1(S^3) = \{0\}.$$

The computation of $H_2(S^3)$. Let us find first the kernel of ∂_2. The matrix of this transformation is

$$\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 \end{bmatrix}.$$

In row-echelon form this is the same as

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The kernel of this homomorphism is a free group with basis $[1, -1, 1, -1]$, namely $F_1 - F_2 + F_3 - F_4$. Note that this basis element equals $\partial_3 T_1$, hence $Z_2(S^3) = B_2(S^3)$ and consequently

$$H_2(S^3) = \{0\}.$$

The computation of $H_3(S^3)$. The matrix of ∂_3 is

$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}.$$

The kernel of this matrix (viewed as a linear transformation between $\mathbb{Z}$-modules) is $[1, 1]^T$. So $Z_3(S^3) = \mathbb{Z}(T_1 + T_2) \cong \mathbb{Z}$, and hence

$$H_3(S^3) = \mathbb{Z}/\{0\} = \mathbb{Z}.$$

We conclude that

$$H_n(S^3) = \begin{cases} \mathbb{Z} & \text{if } n = 0, 3 \\ 0 & \text{if } n \neq 0, 3. \end{cases}$$

These computations can be repeated for higher dimensional spheres, but they become more complicated as the dimension grows. There is only one instance where level of dificulty is the same in all dimensions, namely for the computation of $H_n(S^n)$. Following the same procedure as above we can obtain the following result.

Proposition 5.1.2. For every $n \geq 1$, $H_n(S^n) = \mathbb{Z}$.

Since for $m > n$, $H_m(S^n) = \{0\}$, we obtain that spheres of different dimensions have different homology groups.

5.1.5 Homology with real coefficients and the Euler characteristic

The same considerations apply if in § 5.1.2 we consider $C_n(X)$ to consist of the $\mathbb{R}$-vector space with basis the n-dimensional simplices of X. To avoid confusion, we use the notation $C_n(X;\mathbb{R})$. In this new setting we obtain the homology with real coefficients $H_n(X;\mathbb{R})$. What we have been computing so far is the homology with integer coefficients, for which the standard notation is $H_n(X;\mathbb{Z})$. From now on we will always specify whether the coefficients are integer or real.

In the case of homology with real coefficients we are in the realm of vector spaces, and the linear algebra is simpler. Quotients are easier to compute by just subtracting the dimensions. Let us point out that there is no distinction between the homologies with real, rational or complex coefficients when it comes to computations.

Example 1. The circle. The computations from the previous section apply mutatis mutandis to show that

$$H_n(S^1;\mathbb{R}) = \begin{cases} \mathbb{R} & \text{if } n = 0,1 \\ 0 & \text{if } n \geq 2. \end{cases}$$

Example 2. The projective plane. Realizing $\mathbb{R}P^2$ as a Δ-complex in the same manner as in § 5.1.4 we obtain the chain complex of vector spaces

$$0 \longrightarrow \mathbb{R}F_1 \oplus \mathbb{R}F_2 \xrightarrow{\partial} \mathbb{R}E_1 \oplus \mathbb{R}E_2 \oplus \mathbb{R}E_3 \xrightarrow{\partial} \mathbb{R}V_1 \oplus \mathbb{R}V_2 \longrightarrow 0.$$

with boundary maps

$$\begin{aligned} &\partial_2 F_1 = E_1 + E_2 + E_3, \quad \partial_2 F_2 = E_1 + E_2 - E_3, \\ &\partial_1 E_1 = V_1 - V_2, \quad \partial_1 E_2 = V_2 - V_1, \quad \partial_1 E_3 = V_2 - V_2 = 0. \end{aligned}$$

We see that the image of ∂_1 is the 1-dimensional space with basis $V_1 - V_2$, hence $H_0(\mathbb{R}P^2;\mathbb{R}) = \mathbb{R}$, the quotient of a 2-dimensional space by a 1-dimensional subspace.

Also the kernel of ∂_1 is the 2-dimensional space with basis $E_1 - E_2$ and E_3. The image of ∂_2 is also a 2-dimensional space, because the vectors $E_1 + E_2 + E_3$ and $E_1 + E_2 - E_3$ are linearly independent. Hence $H_1(\mathbb{R}P^2;\mathbb{R}) = \{0\}$.

Finally, $H_2(\mathbb{R}P^2;\mathbb{R}) = \{0\}$.

As we have seen in this example, the homology with real coefficients is coarser then the homology with integer coefficients. The ease of computation comes at a price! In fact, by standard linear algebra, the homology with real coefficients can be computed directly from the one with integer coefficients (but not vice-versa).

Proposition 5.1.3. Let X be a topological space that can be realized as a Δ-complex. If $H_i(X;\mathbb{Z}) = \mathbb{Z}^n \oplus T$, where T is the torsion (i.e. a finite abelian group), then $H_i(X;\mathbb{R}) = \mathbb{R}^n$.

Definition. The number $b_i(X) = \dim H_n(X;\mathbb{R})$ is called the ith *Betti number* of X (after Enrico Betti).

Putting together the Betti numbers, one can construct the oldest known topological invariant.

Definition. Let X be a topological space that can be realized as a Δ-complex. The number

$$\chi(X) = \sum_{i \geq 0} (-1)^i \dim H_i(X;\mathbb{R})$$

is called the *Euler characteristic* of X.

Theorem 5.1.1. (Poincaré) Let X be a topological space that can be realized as a Δ-complex. Then

$$\chi(X) = \sum_{i\geq 0}(-1)^i \mathrm{dim} C_i(X;\mathbb{R}).$$

Proof. Note that by the first isomorphism theorem, if $f : V \to W$ is a linear map between vector spaces, then $V/\mathrm{Ker} f \cong \mathrm{Im} f$, and so $V \cong \mathrm{Ker} f \oplus \mathrm{Im} f$. If V is finite dimensional, then

$$\mathrm{dim} V = \dim\,\mathrm{Ker} f + \dim\,\mathrm{Im} f.$$

Using this fact we obtain

$$\sum_{i\geq 0}(-1)i\beta_i(X) = \sum_{n\geq 0}(-1)^i(\mathrm{dim} Z_i(X;\mathbb{R}) - \mathrm{dim} B_i(X;\mathbb{R}))$$

$$= \sum_{i\geq 0}(-1)^i(\mathrm{dim} Z_i(X;\mathbb{R}) + \mathrm{dim} B_{i-1}(X;\mathbb{R})) = \sum_{i\geq 0}(-1)^i \mathrm{dim} C_i(X;\mathbb{R}),$$

and we are done. □

As a corollary we obtain

Theorem 5.1.2. (Euler) Given a connected planar graph, denote by v the number of vertices, by e the number of edges, and by f the number of faces (including the face containing the point at infinity). Then

$$v - e + f = 2.$$

Proof. By Problem 1 in Homework 3 of this semester and Proposition 5.1.3,

$$H_i(S^2;\mathbb{R}) = \begin{cases} \mathbb{R} & \text{if } i = 0, 2 \\ 0 & \text{otherwise.} \end{cases}$$

It follows that $\chi(S^2) = 1 - 0 + 1 = 2$.

Place the planar graph on S^2. The faces might not be triangles, nevertheless we can triangulate them. Adding an edge increases the number of edges by 1 and the number of faces by 1, and so the sum $v - f + s$ does not change. Once we reach a triangulation, we have a Δ-complex, and then Theorem 5.1.1 applies. Hence the conclusion. □

Example 1. As an application of Euler's formula, let us determine the platonic solids. Recall that a platonic solid (i.e. a regular polyhedron) is a polyhedron whose faces are congruent regular polygons and such that each vertex belongs to the same number of edges.

Let m be the number of edges that meet at a vertex and let n be the number of edges of a face. With the usual notations, when counting vertices by edges we obtain $2e = mv$. When counting faces by edges we obtain $2e = nf$. Euler's formula becomes

$$\frac{2}{m}e - e + \frac{2}{n}e = 2,$$

or

$$e = \left(\frac{1}{m} + \frac{1}{n} - \frac{1}{2}\right)^{-1}.$$

The right-hand side must be a positive integer positive. In particular $\frac{1}{m} + \frac{1}{n} > \frac{1}{2}$. The only possibilities are:

1. $m = 3, n = 3$, in which case $e = 6$, $v = 4$, $f = 4$; this is the regular tetrahedron.
2. $m = 3, n = 4$, in which case $e = 12$, $v = 8$, $f = 6$; this is the cube.
3. $m = 3, n = 5$, in which case $e = 30$, $v = 20$, $f = 12$; this is the regular dodecahedron.
4. $m = 4, n = 3$, in which case $e = 12$, $v = 6$, $f = 8$; this is the regular octahedron.
5. $m = 5, n = 3$, in which case $e = 30$, $v = 12$, $f = 20$; this is the regular icosahedron.

We proved the well known fact that there are five platonic solids.

A second example is the piecewise-linear version of the Gauss-Bonnet theorem.

Example 2. Let Σ_g be a sphere with g handles (i.e. a genus g surface) realized as a polyhedron whose faces are convex polygons. At a point $x \in \Sigma_g$ that is not a vertex, Σ_g is locally isometric to the plane, so we say that the (Gaussian) curvature at that point is $K_x = 0$. If x is a vertex, then the (Gaussian) curvature K_x is 2π minus the sum of the angles that meet at that vertex (this measures how far the vertex is from being flat).

Theorem 5.1.3. (PL Gauss-Bonnet theorem)

$$\sum_{x \in \Sigma_g} K_x = 2\pi\chi(\Sigma_g).$$

Proof. First, realize Σ_g as a Δ-complex by first describing it as a regular $2g$-gon with sides identified pairwise, then divide the regular $2g$-gon into triangles. Now compute the Euler characteristic using Poincaré's Theorem (Theorem 5.1.1) to obtain $\chi(\Sigma_g) = 2 - 2g$. If in an arbitrary realization of Σ_g as a polyhedron whose faces are convex polygons, we denote by v the number of vertices, by e the number of edges, and by f the number of faces, then reasoning similarly as in the proof of Euler's theorem and using Theorem 5.1.1 we obtain

$$v - e + f = \chi(\Sigma_g).$$

Multiply this relation by 2π to obtain

$$2\pi v - 2\pi e + 2\pi f = 2\pi\chi(\Sigma_g).$$

If n_k, $k \geq 3$, denotes the number of faces which are k-gons, then $f = n_3 + n_4 + n_5 + \cdots$. Also, counting edges by the faces, and using the fact that each edge belongs to two faces, we have $2e = 3n_3 + 4n_4 + 5n_5 + \cdots$. The relation becomes

$$2\pi v - \pi(n_3 + 2n_4 + 3n_5 + \cdots) = 2\pi\chi(\Sigma_g).$$

Because the sum of the angles of a k-gon is $(k-2)\pi$, the sum in the above relation is equal to the sum of all angles of faces. The conclusion follows. □

Remark. In the setting of differential geometry, the Gauss-Bonnet theorem is expressed as

$$\int_S K dA = 2\pi\chi(\Sigma_g),$$

or, in words, the integral of the Gaussian curvature over a closed surface Σ_g is equal to the Euler characteristic of the surface multiplied by 2π.

We have seen above homology with coefficients in $\mathbb{Z}$ and $\mathbb{R}$. One can generalize this to homology with coefficients in other rings or fields. Of particular interest is homology with $\mathbb{Z}_2$ coefficients, because on the one hand it is simple enough so that it is very easy to compute, and on the other hand it is nontrivial enough so as to distinguish spaces. You might want to try to compute the homology with $\mathbb{Z}_2$ coefficients for the spaces discussed above.

5.2 Continuous maps between Δ-complexes

5.2.1 Δ-maps

So far we have talked about topological spaces that can be realized as Δ-complexes. What about continuous functions between these spaces. It turns out that the appropriate continuous maps in this situation are the Δ-maps, namely those that map simplices to simplices. Moreover, we will show that any continuous map between topological spaces that are Δ-complexes can be approximated by a Δ-map.

Definition. A Δ-*map* $f : K \to L$ is a continuous map $|K| \to |L|$ which lifts to a map from the disjoint union of the simplices that comprise K to the disjoint union of the simplices that comprise L such that its restriction to each simplex is a linear map onto a simplex.

Example 1. The inclusion of the Möbius band in the Klein bottle, as a Δ-map is shown in Figure 5.4.

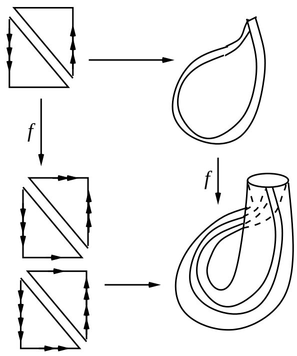

Figure 5.4:

Example 2. A Δ-map that projects the torus onto a circle is shown in Figure 5.5.

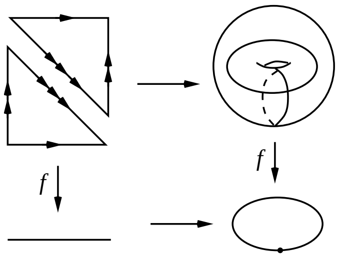

Figure 5.5:

The following result is obvious.

Proposition 5.2.1. The composition of two Δ-maps is a Δ-map.

Definition. Two Δ-complexes are called *isomorphic* if there are Δ-maps $f : K \to L$, $g : L \to K$ such that $f \circ g = 1_L$, $g \circ f = 1_K$.

Given a Δ-map $f : K \to L$, it induces a group homomorphism $(f_\#)_n : C_n(K, \mathbb{Z}) \to C_n(L, \mathbb{Z})$ defined on generators by $(f_\#)_n(\sigma^n) = f(\sigma^n)$ if the dimension of $f(\sigma^n)$ is n and zero otherwise. Note that there is one such map in each dimension. The maps $(f_\#)_n$ satisfy

$$(f_\#)_n \partial_n = \partial_n (f_\#)_{n+1}.$$

Definition. A mapping $\phi_* = (\phi_n)_n$ between chain complexes

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & C_{n+1}(K) & \xrightarrow{\partial_{n+1}} & C_n(K) & \xrightarrow{\partial_n} & C_{n-1}(K) & \longrightarrow & \cdots \\
 & & \phi_{n+1}\downarrow & & \phi_n\downarrow & & \phi_{n-1}\downarrow & & \\
\cdots & \longrightarrow & C_{n+1}(L) & \xrightarrow{\partial_{n+1}} & C_n(L) & \xrightarrow{\partial_n} & C_{n-1}(L) & \longrightarrow & \cdots
\end{array}$$

is called a *chain map* if for all n, $\partial_n \phi_n = \phi_{n-1} \partial_n$.

Proposition 5.2.2. $(f_\#)_*$ is a chain map.

Proof. If the dimension of $f(\sigma^n)$ is n or is less than or equal to $n-2$, then the equality $(f_\#)_{n-1}(\partial_n \sigma^n) = \partial_n((f_\#)_n \sigma^n)$, in the first case because this is the same as $f(\partial_n \sigma^n) = \partial_n(f(\sigma_n)$, and in the second because both sides are zero.

The only interesting case is when the dimension of $f(\sigma^n)$ is exactly $n-1$. Because of linearity the vertices of σ^n are mapped to the vertices of $f(\sigma^n)$ and because the latter has dimension $n-1$, two of the vertices of σ^n are mapped to the same vertex. Set $\sigma_n =< v_0, v_1, \ldots, v_n >$ and $f(\sigma^n) =< w_0, w_1, \ldots, w_{n-1} >$. By reordering the vertices (and maybe changing orientations) we may assume that $f(v_j) = w_j$, $j = 0, 1, \ldots, n-1$, and $f(v_n) = w_{n-1}$. Then, eliminating the simplices that have two equal vertices equal to w_{n-1} in the image of f and hence are zero, we have

$$(f_\#)_{n-1} \partial \sigma_n = (-1)^{n-1} < w_0, w_1, \ldots, w_{n-1} > + (-1)^n < w_0, w_1, \ldots, w_{n-1} >= 0.$$

And this is of course equal to $\partial_n (f_\#)_n \sigma^n$ because $(f_\#)_n \sigma^n = 0$. $\square$

Theorem 5.2.1. Any chain map induces a sequence of homomorphisms

$$(\phi_*)_n : H_n(K) \to H_n(L).$$

Proof. If $\partial_n z = 0$ then $\partial_n \phi_n(z) = \phi_{n-1}(\partial_n z) = 0$. If $z = \partial_{n+1} w$ then $\phi_n z = \phi_n(\partial w) = \partial_{n+1} \phi_n(w)$.

Hence we can define

$$(\phi_*)_n(\gamma + \partial_{n+1} C_{n+1}(K, \mathbb{Z})) = \phi_n(\gamma) + \partial_{n+1} C_{n+1}(L, \mathbb{Z}).$$

$\square$

In particular, because $f_\#$ is a chain map, for each n, $(f_\#)_n$ induces a map $(f_*)_n : H_n(K, \mathbb{Z}) \to H_n(L, \mathbb{Z})$.

A short-hand writing is $f_* : H_*(K, \mathbb{Z}) \to H_*(L, \mathbb{Z})$ understanding that there is such a map in each dimension.

Proposition 5.2.3. (a) If $f : K \to K$ is the identity map, then $(f_*)_n : H_n(K, \mathbb{Z}) \to H_n(K, \mathbb{Z})$ is the identity isomorphism for each n.
(b) If $f : K \to L$ and $g : L \to M$ are Δ-maps between Δ-complexes, then $((g \circ f)_*)_n = (g_*)_n (f_*)_n$, for all n.

Proof. The proof is easy and is left as an exercise. □

Example 3. Consider the projection of the torus onto a circle from Example 2 above. Clearly $(f_*)_2 = 0$, so we only need to be concerned with $(f_*)_0$ and $(f_*)_1$. The map $(f_\#)_0$ maps the unique vertex V of the torus, to the unique vertex V' of the circle. Since

$$H_0(S^1 \times S^1, \mathbb{Z}) = Z_0(S^1 \times S^1, \mathbb{Z}) = \mathbb{Z}V$$

and

$$H_0(S^1, \mathbb{Z}) = Z_0(S^1, \mathbb{Z}) = \mathbb{Z}V',$$

the map $(f_*)_0 : \mathbb{Z}V \to \mathbb{Z}V'$ is an isomorphism.

In dimension 1 the torus has three edges, say E_1, E_2, E_3, and the circle has just one edge, E. It is not hard to see that $(f_\#)_1(E_1) = (f_\#)_1(E_3) = E$ and $(f_\#)_1(E_2) = 0$ (the vertical edge is mapped to V' by the Δ-map).

By the computations in § 5.1.4,

$$H_1(S^1 \times S^1, \mathbb{Z}) = \mathbb{Z}(E_1 + B_1(S^1 \times S^1, \mathbb{Z})) \oplus \mathbb{Z}(E_2 + B_1(S^1 \times S^1, \mathbb{Z})) \cong \mathbb{Z}E_1 \oplus \mathbb{Z}E_2$$

and

$$H_1(S^1, \mathbb{Z}) = Z_1(S^1, \mathbb{Z}) = \mathbb{Z}E.$$

Hence $(f_*)_1 : \mathbb{Z} \oplus \mathbb{Z} \to \mathbb{Z}$ is just the projection onto the first coordinate.

Example 4. Let us consider the map $f : S^1 \to S^1$, $f(z) = z^2$. For this to be a Δ-map, one can realize the first circle as a Δ-complex with two vertices V_1, V_2 and two edges E_1, E_2 as shown in Figure 5.6. The second circle as a Δ-complex with one edge E and one vertex V, as in § 5.1.4. Let us examine the homology of the first Δ-complex.

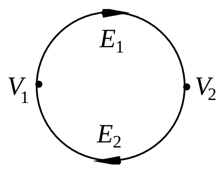

Figure 5.6:

The chain complex is

$$0 \to \mathbb{Z}E_1 \oplus \mathbb{Z}E_2 \xrightarrow{\partial_1} \mathbb{Z}V_1 \oplus \mathbb{Z}V_2 \to 0,$$

where $\partial_1(E_1) = -\partial_1(E_2) = V_2 - V_1$.

Since

$$\mathbb{Z}V_1 \oplus \mathbb{Z}V_2 = \mathbb{Z}V_1 \oplus \mathbb{Z}(V_2 - V_1),$$

it follows that $H_0(S^1, \mathbb{Z})$ is one-dimensional with basis (the equivalence class of) V_1. And since $f_\#(V_1) = V$, we obtain that $(f_*)_0 : \mathbb{Z} \to \mathbb{Z}$ is the identity map.

In dimension one, an easy computation shows that

$$H_1(S^1, \mathbb{Z}) = Z_1(S^1, \mathbb{Z}) = \mathbb{Z}(E_1 + E_2).$$

Because $f(_\#)_1(E_1 + E_2) = 2E$, we obtain that $(f_*)_1 : \mathbb{Z} \to \mathbb{Z}$ is multiplication by 2.

5.2.2 Simplicial complexes, simplicial maps, barycentric subdivision.

We will now focus on a more restrictive notion than that of a Δ-complex, that of a simplicial complex.

Definition. A Δ-complex K for which every simplex is embedded in $|K|$ is called a *simplicial complex.* A Δ-map between simplicial complexes is called a *simplicial map.*

Example 1. In Figure 5.7 we see how the torus can be realized as a simplicial complex. It has 4 0-dimensional simplices, 12 1-dimensional simplices, and 8 2-dimensional simplices.

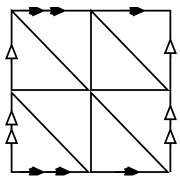

Figure 5.7:

Definition. The *barycenter* of a simplex $\sigma^k = \langle v_0, v_1, \ldots, v_k \rangle$ is the point

$$\dot{\sigma}^k = \frac{1}{k+1} v_0 + \frac{1}{k+1} v_1 + \cdots + \frac{1}{k+1} v_k.$$

Definition. The *barycentric subdivision* of a symplex $\langle v_0, v_1, \cdots, v_k \rangle$ is the collection of the $(k+1)!$ simplices

$$\left\langle \dot{\sigma}_{\pi(0)}, \dot{\sigma}_{\pi(1)}, \ldots, \dot{\sigma}_{\pi(k)} \right\rangle$$

where π is a permutation of $\{0, 1, 2, \ldots, k\}$ and for each $i = 0, 1, 2, \ldots, k$, $\sigma_{\pi(i)} = \left\langle v_{\pi(0)}, v_{\pi(1)}, \ldots, v_{\pi(i)} \right\rangle$.

Example 2. The barycentric subdivisions for the one- and two-dimensional simplices are shown in Figure 5.8.

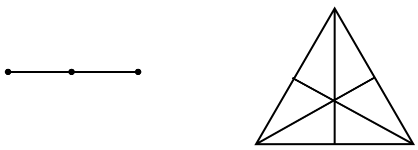

Figure 5.8:

Proposition 5.2.4. If K is a Δ-complex, then $K^{(1)}$ is a simplicial complex.

The *mth barycentric subdivision* is obtained by applying this operation m times to each simplex of a Δ-complex. The mth barycentric subdivision of a Δ-complex K is again a Δ-complex, denoted by $K^{(m)}$.

Each vertex v induces a map $v : C_n(K) \to C_{n+1}(K)$, by $v \langle v_0, v_1, \ldots, v_n \rangle = \langle v, v_0, v_1, \ldots, v_n \rangle$ when this makes sense, and 0 otherwise.

Lemma 5.2.1. For any vertex v and n-chain c such that $vc \neq 0$, one has

$$\partial_{n+1}(vc) = c - v\partial_n c.$$

Proof. It suffices to check the equality on simplices. For $c = \langle v_0, v_1, \ldots, v_n \rangle$, we have

$$\begin{aligned}\partial_{n+1} v \langle v_0, v_1, \ldots, v_n \rangle &= \partial_{n+1} \langle v, v_0, v_1, \ldots, v_n \rangle \\ &= \langle v_0, v_1, \ldots, v_n \rangle - \langle v, v_1, v_2, \ldots, v_n \rangle + \langle v, v_0, v_2, \ldots, v_n \rangle \cdots \\ &= c - v(\langle v_1, v_2, \ldots, v_n \rangle - \langle v_0, v_2, \ldots, v_n \rangle + \cdots)\end{aligned}$$

and this equals $c - v\partial_n c$, as desired. □

Definition. The *first chain derivation* is defined inductively on the dimension of chains by

$$\begin{aligned}\phi_0(v) &= v \\ \phi_n(\sigma^n) &= \dot{\sigma}^n \phi_{n-1}(\partial \sigma^n).\end{aligned}$$

Example 3. Figure 5.9 shows the first chain derivations of a 1-simplex and a 2-simplex, where in each case, the chain derivation of the simplex on the left is the sum of the simplices on the right with specified orientations.

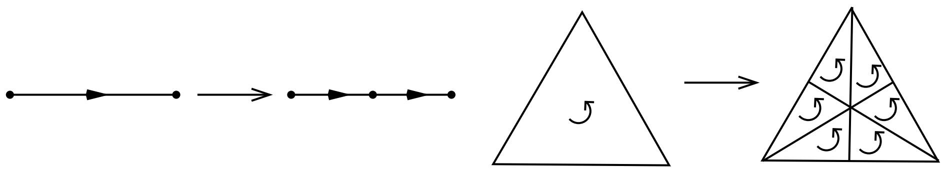

Figure 5.9:

One can think of the chain derivation as replacing each simplex with the union of the pieces obtained after applying a barycentric subdivision. By iterating this procedure m times we obtain what is called the *mth chain derivation.*

Theorem 5.2.2. Each chain derivation is a chain mapping.

Proof. We have to prove that the following diagram is commutative:

$$\begin{array}{ccccccc}\cdots & \xrightarrow{\partial_{n+1}} & C_n(K) & \xrightarrow{\partial_n} & C_{n-1}(K) & \xrightarrow{\partial_{n-1}} & \cdots \\ & & \downarrow{\scriptstyle \phi_n} & & \downarrow{\scriptstyle \phi_{n-1}} & & \\ \cdots & \xrightarrow{\partial_{n+1}} & C_n(K^{(1)}) & \xrightarrow{\partial_n} & C_{n-1}(K^{(1)}) & \xrightarrow{\partial_{n-1}} & \cdots\end{array}$$

We induct on n. For $n = 0$ there is nothing to check. When $n = 1$ it is easy to check by hand that $\partial_1\phi_1 = \phi_0\partial_1$.

Assuming that $\partial_{n-1}\phi_{n-1} = \phi_{n-2}\partial_{n-1}$ we can write

$$\begin{aligned}\partial_n\phi_n(\sigma^n) = \partial_n(\dot{\sigma}^n\phi_{n-1}\partial_n\sigma^n) &= \phi_{n-1}\partial_n\sigma^n - \dot{\sigma}^n\partial_{n-1}\phi_{n-1}\partial_n\sigma^n \\ &= \phi_{n-1}\partial_n\sigma^n - \dot{\sigma}_n\phi_{n-2}\partial_{n-1}\partial_n\sigma^n = \phi_{n-1}\partial_n\sigma^n. \quad □\end{aligned}$$

Proposition 5.2.5. Let K be a Δ-complex with first chain derivation $\phi = (\phi_n)_{n\geq 0}$. Then there exists a chain mapping

$$\psi = \{\psi_n : C_n(K^{(1)}) \to C_n(K)\}$$

such that $\psi_n \circ \phi_n$ is the identity map on $C_n(K)$, for all $n \geq 0$.

Proof. Let $f : K^{(1)} \to K$ be a Δ-map defined as follows. First we define f on vertices, by setting $f(v) = v$ if v is a vertex of K and $f(\dot{\sigma}^n)$ is a vertex of the simplex σ^n of which $\dot{\sigma}^n$ is the barycenter. Extend the construction linearly over each simplex.

Let ψ be the chain map induced by f. Then ψ has the desired property. ☐

Definition. Two chain maps $(\phi_n)_{n\geq 0}$ and $(\psi_n)_{n\geq 0}$ are called *chain homotopic* if there exist linear homomorphisms $D_n : C_n(K) \to C_{n+1}(L)$, $n \geq 0$ (with $D_{-1} = 0$) such that

$$\partial_{n+1} D_n + D_{n-1}\partial_n = \phi_n - \psi_n, \quad \text{for all } n \geq 0.$$

On a diagram this looks like

$$\begin{tikzcd}
C_{n+1}(K) \arrow[r, "\partial_{n+1}"] \arrow[d, "\phi_{n+1}-\psi_{n+1}"'] & C_n(K) \arrow[r, "\partial_n"] \arrow[d, "\phi_n-\psi_n"] \arrow[dl, "D_n"'] & C_{n-1}(K) \arrow[d, "\phi_{n-1}-\psi_{n-1}"] \arrow[dl, "D_{n-1}"'] \\
C_{n+1}(L) \arrow[r, "\partial_{n+1}"'] & C_n(L) \arrow[r, "\partial_n"'] & C_{n-1}(L)
\end{tikzcd}$$

The idea behind this definition comes from actual homotopies. If a homotopy $H : \sigma^n \times [0,1] \to L$ is also a Δ-map, then its image is a prism. The map D_n associates to σ_n this prism, written as the sum of the simplices that compose it. The 1- and 2-dimensional cases are shown in Figure 5.10.

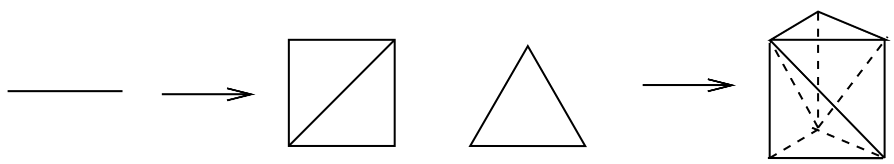

Figure 5.10:

Theorem 5.2.3. If $(\phi_n)_{n\geq 0}$ and $(\psi_n)_{n\geq 0}$ are chain homotopic, then for each n, $(\phi_*)_n = (\psi_*)_n$, where ϕ_* and ψ_* are the maps induced on homology.

Proof. For a cycle $z_n \in C_n(K)$ we have

$$\begin{aligned}
(\phi_*)_n([z_n]) - (\psi_*)_n([z_n]) &= ((\phi_*)_n - (\psi_*)_n)([z_n]) = [(\phi_n - \psi_n)(z_n)] \\
&= [\partial_{n+1} D_n(z_n) + D_{n-1}(\partial_n z_n)] = [\partial_{n+1} D_n(z_n)],
\end{aligned}$$

and this is zero because it is a boundary. ☐

Definition. Two Δ-complexes K and L are called *chain equivalent* if there exist chain mappings $\phi = \{\phi_n : C_n(K) \to C_n(L)\}_{n\geq 0}$ and $\psi = \{\psi_n : C_n(L) \to C_n(K)\}_{n\geq 0}$ such that $\phi \circ \psi$ is chain homotopic to the identity map of $C_*(K)$ and $\psi \circ \phi$ is chain homotopic to the identity map of $C_*(L)$.

Corollary 5.2.1. Chain equivalent Δ-complexes have isomorphic homology groups in all dimensions.

Proof. If $\phi_n : C_n(K) \to C_n(L)$ and $\psi_n : C_n(L) \to C_n(K)$ are chain equivalences, then $\phi_n \circ \psi_n$ is chain homotopic to the identity on $C_n(L)$ and $\psi_n \circ \phi_n$ is chain homotopic to the identity on $C_n(K)$. By Theorem 5.2.3, $(\phi_*)_n \circ (\psi_*)_n = 1_{H_n(L)}$ and $(\psi_*)_n \circ (\phi_*)_n = 1_{H_n(K)}$. □

As a corollary, we compute the homology of a simplex.

Proposition 5.2.6. Let σ^m be an m-dimensional simplex. Then

$$H_n(\sigma^m) = \begin{cases} \mathbb{Z} & \text{if } n = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Pick a vertex v of σ^m and define $D_n : C_n(\sigma^m) \to C_{n+1}(\sigma^m)$ by

$$D_n(s^n) = \begin{cases} vs^n & \text{if } v \text{ does not belong to } s^n \\ 0 & \text{otherwise.} \end{cases}$$

One can check that for $n \geq 1$,

$$id_n = \partial_{n+1} D_n + D_{n-1} \partial_n.$$

This means that $id_n : H_n(\sigma^m) \to H_n(\sigma^m)$ is the zero map, and this can only happen if $H_n(\sigma^m) = 0$ for $n \geq 1$. □

On the other hand $H_0(\sigma^m) = \mathbb{Z}$, because σ^m is path connected so any two points are homologous.

We are now in position to make the first advance towards the proof that homology groups are topological invariants.

Theorem 5.2.4. The homology groups of a Δ-complex and its first barycentric subdivision are isomorphic.

Proof. In view of Corollary 5.2.1, it suffices to show that the Δ-complex and its first barycentric subdivision are chain equivalent.

So we have two Δ-complexes, K and its first barycentric subdivision $K^{(1)}$. Let $\phi = (\phi_n)_{n \geq 0}$ be the first chain derivation of K and let $\psi = (\psi_n)_{n \geq 0}$ be the chain map defined in Proposition 5.2.5. We already know that $\psi \circ \phi$ is chain homotopic to the identity map of $C_*(K)$, because it *is* the identity. We need to show that $\psi \circ \phi$ is chain homotopic to the identity map of $C_*(K^{(1)})$.

For a simplex $\sigma^n = \langle v_0, v_1, \ldots, v_n \rangle \in K^{(1)}$ define

$$\begin{aligned} D_n(\langle v_0, v_1, \ldots, v_n \rangle) = & \langle v_0, \psi(v_0), \psi(v_1), \ldots, \psi(v_n) \rangle - \langle v_0, v_1, \psi(v_1), \ldots, \psi(v_n) \rangle + \cdots \\ & \pm \langle v_0, v_1, \ldots, v_n, \psi(v_n) \rangle . \end{aligned}$$

In this sum the convention is that degenerate simplices are equal to 0, while the other simplices should be divided by barycenters into actual simplices of the barycentric subdivision.

A little algebra shows that

$$\partial_{n+1} D_n(\sigma^n) + D_{n-1} \partial_n(\sigma^n) = \phi_n \psi_n(\sigma^n) - \sigma^n,$$

and hence D_n, $n \geq 0$ is a chain homotopy between $\phi \circ \psi$ and the identity map. This completes the proof. □

It is important to notice that the homology groups are isomorphic in a canonical way, the isomorphism being defined by the first chain derivation. As such a cycle in K gives the same cycle in $K^{(1)}$ but with its simplices subdivided.

Theorem 5.2.5. If $f : K \to L$ is a Δ-map then $f : K^{(1)} \to L^{(1)}$ is a simplicial map, and the following diagram commutes

$$\begin{array}{ccc} H_n(K) & \xrightarrow{(f_*)_n} & H_n(L) \\ \cong\downarrow & & \cong\downarrow \\ H_n(K^{(1)}) & \xrightarrow{(f_*)_n} & H_n(L^{(1)}) \end{array}$$

Proof. The corresponding diagram for $f_\#$ at the level of chains is commutative. The conclusion follows. ☐

5.2.3 The simplicial approximation theorem

In view of Theorem 5.2.4 for the formal proofs we can focus only on simplicial complexes. First, some definitions.

Definition. The *open simplex*, *open*(σ), is the collection of all points in σ with positive barycentric coordinates. The *open star* of a vertex $ost(v)$, is the union of all open simplices containing v.

Note that $ost(v)$ is an open neighborhood of v.

Lemma 5.2.2. If $v_0, v_1, \ldots, v_n$ are vertices in a simplicial complex K, then they are the vertices of a simplex if and only if

$$\cap_{i=0}^{n} ost(v_i) \neq \emptyset.$$

Proof. If they are the vertices of a simplex, then the intersection of the open stars contains the interior of that simplex, so is nonempty.

Conversely, if the intersection is nonempty, let x be a point in this intersection. Then all the v_i, $0 \leq i \leq n$, appear with positive barycentric coordinates in x. This can only happen if the v_i, $1 \leq i \leq n$, lie in a simplex. Consequently they are the vertices of a simplex. ☐

Definition. Let K and L be simplicial complexes, and $f : |K| \to |L|$ be a continuous map. We say that K *is star related to* L *relative to* f if for any vertex v in K there is a vertex w in L such that $f(ost(v)) \subset ost(w)$.

The following result was proved by Brouwer.

Theorem 5.2.6. (The simplicial approximation theorem) Let K and L be simplicial complexes, and let $f : |K| \to |L|$ be a continuous function. Then there is a barycentric subdivision $K^{(n)}$ and a continuous function $f_\Delta : |K| \to |L|$ such that
a) f_Δ is a simplicial map from $K^{(n)} \to L$
b) f_Δ is homotopic to f.

Proof. We need the following result.

Lemma 5.2.3. Let K and L be simplicial complexes, and let $f : |K| \to |L|$ be a continuous function such that K is star related to L relative to f. Then there is a simplicial map $f_\Delta : |K| \to |L|$ homotopic to f.

Proof. Since K is star related to L relative to f, there is for each vertex v of a simplex of K a vertex w of simplex of L such that $f(ost(v)) \subset ost(w)$. Define $f_\Delta(v) = w$. Let $\sigma = \langle v_0, \dots, v_n\rangle$ be a simplex of K. Then $\cap_{i=1}^n ost(v_i) \neq \emptyset$. It follows that $\cap_{i=0}^n ost(f_\Delta(v_i)) \neq \emptyset$. By Lemma 5.2.2, the elements of the set

$$\{f_\Delta(v_i) \,|\, i = 0, 1, 2, \dots, n\}$$

are the vertices of a simplex σ' in L (this simplex can have dimension less than n). Hence f_Δ can be extended linearly, to a simplicial map $f_\Delta : K \to L$.

To see that the geometric realization $f_\Delta : |K| \to |L|$ is homotopic to f, note that with this definition $(f(ost(v)) \subset ost(w) = ost(f_\Delta(v))$. As such, whenever $x \in |K|$ and $f(x)$ belongs to a simplex, then $f_\Delta(x)$ belongs to that simplex. Thus we can define a homotopy between f and f_Δ by $H(x,t) = tf(x) + (1-t)f_\Delta(x)$. The lemma is proved. □

Let us return to the theorem. The space $|L|$ is compact and as explained in Remark 5.1.1, it is also a metric space. The open cover $\{ost(v), v \text{ a vertex of } L\}$ has a Lebesgue number $\eta > 0$, meaning that every ball of radius η in $|L|$ is inside the open star of some vertex.

The function f is uniformly continuous so there is a δ such that for every x, $f(B(x,\delta))$ subset $B(f(x),\eta)$. Consider a sufficiently small barycentric subdivision of K such that the open star of each vertex lies inside a ball of radius δ. Then the image of this every open star lies inside the open star of a vertex in L, and so K is star related to L relative to f. The conclusion follows by applying Lemma 5.2.3. □

Definition. A map f_Δ satisfying the properties from the statement of the theorem is called a *simplicial approximation of f.*

Definition. Two simplicial maps $f, g : K \to L$ are said to be *contiguous* if for every simplex $\sigma \in K$, the vertices of $f(\sigma)$ together with the vertices of $g(\sigma)$ form a simplex in L.

Proposition 5.2.7. Suppose $f : |K| \to |L|$ is a continuous function that has simplicial approximations f_Δ and f'_Δ. Then f_Δ and f'_Δ are contiguous.

Proof. Suppose $\sigma = \langle v_0, v_1, \dots, v_n\rangle \in K$. Then for $x \in open(\sigma)$,

$$f(x) \in f\left(\cap_i ost(v_i)\right) \subset \cap_i f(ost(v_i)) \subset \cap_i ost(f_\Delta(v_i)) \cap ost(f'_\Delta(v_i)).$$

Since this intersection is not empty, by Lemma 5.2.2 the vertices of $f_\Delta(\sigma)$ together with those of $f'_\Delta(\sigma)$ are the vertices of a simplex in L. □

Theorem 5.2.7. (a) Contiguous simplicial maps have homotopic realizations.
(b) Conversely, if K and L are simplicial complexes and $f, g : |K| \to |L|$ are homotopic maps with simplicial approximations $f_\Delta, g_\Delta : K^{(m)} \to L$, then there is $m > n$ and simplicial maps maps $f_0 = f_\Delta, f_1, \dots, f_m = g_\Delta$ from $K^{(n)}$ to L such that for each $0 \le i < n$, f_i is contiguous to f_{i+1}.

Proof. (a) If f and g are contiguous, then $H : |K| \to |L|$, $H(x,t) = (1-t)f(x) + tg(x)$ is a homotopy between f and g. Here the line segment $(1-t)f(x) + tg(x)$ is taken in the simplex to which both $f(x)$ and $g(x)$ belong by the condition of f and g to be contiguous.
(b) Let $H : |K| \times [0,1] \to |L|$ be a homotopy with $H(\cdot, 0) = f$, $H(\cdot, 1) = g$. Consider the open cover of $|K|$ given by

$$\{H^{-1}(ost(w)) \,|\, w \text{ vertex in } L\}.$$

Using the Lebesgue number lemma, we deduce that there is a partition of $[0,1]$, $0 = t_1 < t_2 < \ldots < t_{k-1} = 1$ such that for any $x \in |K|$ and $i < m$, $H(x, t_i)$ and $H(x, t_{i+1})$ lie in the same $ost(w)$ for some $w \in L$. Define the functions $h_i : |K| \to |L|$ by $h_i(x) = H(x, t_i)$. Consider the (finitely many) open sets $h_i^{-1}(ost(w)) \cup h_{i+1}^{-1}(ost(w))$, w a vertex in L, $1 \leq i \leq k-2$. Subdivide K enough times such that each of the simplices of $K^{(n)}$ lies entirely in one of these open sets. Let $f_i : K^{(n)} \to L$ be a simplicial map such that $h_i(ost(v)) \cup h_{i+1}(ost(v)) \subset ost(f_i(v))$ for each vertex v of $K^{(n)}$. By definition, f_i is a simplicial approximation to both h_i and h_{i+1}. Hence f_i and f_{i+1} are contiguous. Since $h_0 = f$ and $h_k = g$, f_1 and f_{k-1} are simplicial approximations to f respectively g. By Proposition 5.2.7, f_1 is contiguous to $f_0 = f_\Delta$ and f_{k-1} is contiguous to $f_k = g_\Delta$. The theorem is proved. □

5.2.4 The independence of homology groups on the geometric realization of the space as a Δ-complex

Lemma 5.2.4. Contiguous simplicial maps induce the same homomorphisms at the level of homology groups.

Proof. We will show that the simplicial maps are chain homotopic. Let $\phi, \psi : K \to L$ be contiguous simplicial maps. Define $D_n : C_n(K) \to C_{n+1}(L)$ by

$$D_n(\langle v_0, v_1, \ldots, v_n \rangle = \sum_{i=0}^{n} (-1)^i \langle \phi(v_0), \ldots, \phi(v_i), \psi(v_i), \ldots, \psi(v_n) \rangle).$$

Here again we use the convention that degenerate simplices are mapped to zero. This map is well defined, namely each summand is a simplex, precisely because ϕ and ψ are contiguous.

We compute

$$\begin{aligned}
\partial_{n+1} D_n(\langle v_0, v_1, \ldots, v_n \rangle) &= \partial_{n+1} \left(\sum_{i=0}^{n} (-1)^i \langle \phi(v_0), \ldots, \phi(v_i), \psi(v_i), \ldots, \psi(v_n) \rangle \right) \\
&= \sum_{j \leq i} (-1)^{i+j} \left\langle \phi(v_0), \ldots, \widehat{\phi(v_j)}, \ldots, \phi(v_i), \psi(v_i), \ldots, \psi(v_n) \right\rangle \\
&\quad + \sum_{i \leq j} (-1)^{i+j+1} \left\langle \phi(v_0), \ldots\ldots, \phi(v_i), \psi(v_i), \ldots, \widehat{\psi(v_j)}, \ldots, \psi(v_n) \right\rangle .
\end{aligned}$$

Also

$$\begin{aligned}
D_{n-1} \partial_n(\langle v_0, v_1, \ldots, v_n \rangle) &= D_{n-1} \left(\sum_{j=0}^{n} (-1)^j \langle v_0, \ldots, \widehat{v_j}, \ldots, v_n \rangle \right) \\
&= \sum_{j < i} (-1)^{i+j-1} \left\langle \phi(v_0), \ldots, \widehat{\phi(v_j)}, \ldots, \phi(v_i), \psi(v_i), \ldots, \psi(v_n) \right\rangle \\
&\quad + \sum_{j > i} (-1)^{i+j} \left\langle \phi(v_0), \ldots, \phi(v_i), \psi(v_i), \ldots, \widehat{\psi(v_j)}, \ldots, \psi(v_n) \right\rangle .
\end{aligned}$$

When adding the two everything with $i \neq j$ cancels, and we only have the $i = j$ terms from the first expression. Each of these appears twice, with opposite signs, except for $i = 0$ and $i = n$, one of which appears with plus, and the other with minus. Hence

$$(\partial_{n+1} D_n + D_{n-1} \partial_n)(\langle v_0, v_1, \ldots, v_n \rangle) = \langle \phi(v_0), \phi(v_1), \ldots, \phi(v_n) \rangle - \langle \psi(v_0), \psi(v_1), \ldots, \psi(v_n) \rangle ,$$

showing that D_n, $n \geq 0$, is a chain homotopy between ϕ and ψ. The conclusion follows. □

This result combined with Theorem 5.2.7 allows us to define, for each continuous map between spaces that can be realized as simplicial complexes, a homomorphism between homology groups.

Definition. Let K and L be Δ-complexes and $f : |K| \to |L|$ a continuous map with simplicial approximation f_Δ. Define $(f_*)_n : H_n(K) \to H_n(L)$ by $(f_*) = (f_{\Delta *})_n$.

Theorem 5.2.8. Let K and L be Δ-complexes and $f, g : |K| \to |L|$ be continuous maps that are homotopic. Then $(f_*)_n = (g_*)_n$, $n \geq 0$. Consequently, if the underlying topological spaces of two Δ-complexes are homotopically equivalent, then their homology groups are isomorphic.

Proof. This is a corollary of Lemma 5.2.4 and Theorem 5.2.7. □

Corollary 5.2.2. If two spaces are homeomorphic, then their homology groups are isomorphic.

In particular, let X be a topological spaces such that $X = |K| = |L|$, with K, L Δ-complexes (or simplicial complexes). Then $H_n(K) = H_n(L)$, for $n \geq 0$. Indeed, if we let $f = id : |K| \to |L|$ and $g = id : |L| \to |K|$, with simplicial approximations f_Δ and g_Δ, then $f_\Delta \circ g_\Delta$ is homotopic to $id : |L| \to |L|$ and $g_\Delta \circ f_\Delta$ is homotopic to $id : |K| \to |K|$, so by functoriality $(f_\Delta)_* \circ (g_\Delta)_* = id$ and $(g_\Delta)_* \circ (f_\Delta)_* = id$, which shows that $(f_\Delta)_*$ and $(g_\Delta)_*$ are isomorphisms. Thus the homology groups do not depend on how X is realized as a Δ-complex.

Definition. A space is called *contractible* if it is homotopy equivalent to a point. An alternative way to say this is that the identity map is null-homotopic.

Corollary 5.2.3. If X is a contractible topological space that can be realized as a Δ-complex, then

$$H_n(X) = \begin{cases} \mathbb{Z} & \text{if } n = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proof. If the space is contractible, it has the homology of a point. We realize the point as a Δ-complex with one 0-dimensional simplex. The corresponding simplicial complex is

$$\cdots 0 \to 0 \to \cdots \to 0 \to \mathbb{Z} \to 0.$$

Hence the conclusion. □

5.3 Applications of homology

Theorem 5.3.1. If m and n are positive integers and $m \neq n$, then $\mathbb{R}^m$ and $\mathbb{R}^n$ are not homeomorphic.

Proof. Assume that for some $m < n$, there is a homeomorphism $h : \mathbb{R}^m \to \mathbb{R}^n$. Consider the restriction

$$h : \mathbb{R}^m \backslash \{0\} \to \mathbb{R}^n \backslash \{h(0)\},$$

which is still a homeomorphism. Thus h should induce isomorphism at the level of homology groups. If $m = 1$ this is impossible since the second space is connected, while the first is not.

For $m > 1$, $\mathbb{R}^m \backslash \{0\}$ and $\mathbb{R}^n \backslash \{h(0)\}$ are homotopically equivalent to S^{m-1} respectively S^{n-1}. This means thta S^{m-1} and S^{n-1} are homotopically equivalent. Consequently, they have isomorphic homology groups.

But, by Proposition 5.1.2, $H_{n-1}(S^{n-1}) = \mathbb{Z}$, while $H_{n-1}(S^{m-1}) = 0$ because S^{m-1} can be realized as a Δ-complex with no $n-1$-dimensional simplex. The two groups cannot be isomorphic, a contradiction. Hence our assumption was false, and the conclusion follows. □

We should remark that within the proof we also showed that S^n and S^m are not homotopically equivalent if $m \neq n$. In view of the proof, we can extend homology to some spaces that are not themselves realizable as Δ-complexes. If X is homotopy equivalent to Y and Y can be realized as a Δ-complex then we set $H_n(X) = H_n(Y)$. The definition is independent of Y (up to an isomorphism).

The following generalizes Theorem 4.3.6.

Theorem 5.3.2. (Brouwer fixed point theorem) Every continuous function from a closed ball of $\mathbb{R}^n$ to itself has a fixed point.

Proof. This proof mimics that for the 2-dimensional situation given in § 4.3.3. Assume to the contrary that there is $f : \overline{B}^n \to \overline{B}^n$ which is continuous with no fixed points. For $\mathbf{x} \in \overline{B}^n$, define $g(\mathbf{x})$ as the point where the ray from $f(\mathbf{x})$ to $\mathbf{x}$ meets the boundary S^{n-1} of the ball. The function $g : \mathbf{B}^n \to S^{n-1}$ is continuous (prove it!). Furthermore, if i is the inclusion of S^{n-1} into $\overline{B}^n$, then $g \circ i$ is the identity map. Looking at the composition of isomorphisms

$$H_{n-1}(S^{n-1}) \xrightarrow{i_*} H_{n-1}(\overline{B}^n) \xrightarrow{g_*} H_{n-1}(S^{n-1}),$$

we notice that it must be equal to zero, because the homology group in the middle is the trivial group, by Corollary 5.2.3. But $(g \circ i)_* = 1_{\mathbb{Z}}$. This is a contradiction. The conclusion follows. $\square$

The proof of Proposition 4.3.2 applies mutatis mutandis to the following general situation.

Proposition 5.3.1. Every square matrix with positive entries has an eigenvector with positive entries.

Here is another application to game theory. First some terminology. Consider a game in which two players, 1 and 2, have m respectively n strategies. Let A and B be the be the payoff matrices, meaning that if player 1 chooses strategy i and player 2 strategy j then the payoffs of the two players are A_{ij} respectively B_{ij}. A mixed strategy is a pair of points $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ with nonnegative entries, such that $\sum_{i=1}^n x_i = \sum_{j=1}^m y_j = 1$. As such they represent probability distributions for the strategies of the two players. Consequently, the expectation values of the payoffs for the two players are respectively $\mathbf{x}^T A\mathbf{y}$ and $\mathbf{x}^T B\mathbf{y}$.

A *Nash equilibrium* (after John Nash) is a mixed strategy $(\mathbf{x}, \mathbf{y})$ such that for all $\mathbf{x}', \mathbf{y}'$,

$$\mathbf{x}'^T A\mathbf{y} \leq \mathbf{x}^T A\mathbf{y} \text{ and } \mathbf{x}^T B\mathbf{y}' \leq \mathbf{x}^T B\mathbf{y}.$$

This means that if one of the players plays according to the Nash equilibrium, the other is obliged to play according the the Nash equilibrium in order to maximize the payoff.

Theorem 5.3.3. (Nash) For all payoff matrices A and B there is a Nash equilibrium.

Proof. Assume that there are payoff matrices A and B for which this does not hold. Consider the space of all mixed strategies. It is the product of an $n-1$-dimensional simplex σ^{n-1} with an $m-1$-dimensional simplex σ^{m-1}, hence it is homeomorphic to a closed $m+n-2$-dimensional ball $\overline{B}^n$. Let $\phi(\mathbf{x}') = \mathbf{x}'^T A\mathbf{y}$ and $\psi(\mathbf{y}') = \mathbf{x}^T B\mathbf{y}'$. Because of the fact that there are no Nash equlibria, the vector $((\nabla\phi)_{\mathbf{x}}, (\nabla\psi)_{\mathbf{y}})$ is never zero. Define the function $f : \sigma^{n-1} \times \sigma^{m-1} \to \partial(\sigma^{n-1} \times \sigma^{m-1})$ by associating to each point $(\mathbf{x}, \mathbf{y})$ the point at which the ray from $(\mathbf{x}, \mathbf{y})$ of direction $((\nabla\phi)_{\mathbf{x}}, (\nabla\psi)_{\mathbf{y}})$ intersects the boundary of the space of all mixed strategies.

By Brouwer's fixed point theorem, f has a fixed point. But this is impossible, since the ray points away from $(\mathbf{x}, \mathbf{y})$. The contradiction proves that our assumption was false, and the conclusion follows. $\square$

Remark 5.3.1. There is no known algorithm for finding the Nash equilibrium in polynomial time.