ELEMENTS OF HOMOTOPY THEORY

Carlos Prieto*

* Instituto de Matemáticas, UNAM

2010

Date of version: January 8, 2010

Preface

Topology is qualitative geometry. Ignoring dimensions, several geometric objects give rise to the same topological object. Homotopy theory considers even more geometric objects as equivalent objects. For instance, in homotopy theory, a solid ball of any dimension and a point are considered as equivalent, also a solid torus and a circle are equivalent from the point of view of homotopy theory. This loss of precision is compensated by the effectiveness of the algebraic invariants that are defined in terms of homotopy.

The problem of deciding if two spaces are homeomorphic or not is, no doubt, the central problem in topology. We call this the *homeomorphism problem.* It was not until the creation of algebraic topology that it was possible to give a reasonable answer to such a problem. Now it is not only because of the conceptual simplicity of point-set topology and its adequate symbology, but thanks to the powerful tool provided by algebra and its most convenient functorial relationship to topology that this effectiveness is achieved. For instance, if two spaces have different algebraic invariants, then they cannot be equivalent from the homotopical viewpoint. Therefore, they cannot be homeomorphic either.

What we now know as algebraic topology was probably started with the *Analysis Situs*, Paris, 1895, and its five *Compléments* (Complements), Palermo, 1899; London, 1900; Paris, 1902; Paris, 1902, and Palermo, 1904, of Henri Poincaré. In the first, he notes that "*geometry is the art of reasoning well with badly made figures.*" And further he says: "*Yes, without doubt, but with one condition. The proportions of the figures might be grossly altered, but their elements must not be interchanged and must preserve their relative situation. In other terms, one does not have to worry about quantitative properties, but one must respect the qualitative properties, that is to say precisely those which are the concern of Analysis Situs.*" Indeed, other works of Poincaré contain as much interesting topology as the just referred to. This is the case for his memoir on the qualitative theory of differential equations, that includes the famous formula of the Poincaré index. This formula describes in topological terms the famous Euler formula, and constitutes one of the first steps in algebraic topology. In these works Poincaré considers already maps on manifolds such as, for instance, vector fields, whose indexes determine the Euler characteristic in his index formula. It is Poincaré too who generalizes the question on the classification of manifolds having in mind the classification of the orientable surfaces considered by Moebius in *Theorie der elementaren Verwandtschaft* (Theory of elementary relationship), Leipzig, 1863. This classification problem was also solved by Jordan in *Sur la déformation des Surfaces* (On the deformation of surfaces), Paris, 1866, who, by classifying surfaces solved an important homeomorphism problem.

Jordan studied also the homotopy classes of closed paths, that is, the first notions of the fundamental group, inspired by Riemann, who already had analyzed the behavior of integrals of holomorphic differential forms and therewith the concept of homological equivalence between closed paths.

This book has the purpose of presenting the topics that, from my own point of view, are the basic topics of algebraic topology that some way should be learnt by any undergraduate student interested in this area or affine areas in mathematics.

The design of the text is as follows. We start with a small Chapter 1, which deals with some basic concepts of general topology, followed by four substantial chapters, each of which is divided into several sections that are distinguished by their double numbering (1.1, 1.2, 2.1, . . .). Definitions, propositions, theorems, remarks, formulas, exercises, etc., are designated with triple numbering (1.1.1, 1.1.2, . . .). Exercises are an important part of the text, since many of them are intended to carry the reader further along the lines already developed in order to prove results that are either important by themselves or relevant for future topics. Most of these are numbered, but occasionally they are identified inside the text by italics (*exercise*).

After the chapter on basic concepts, the book starts with the concept of a topological manifold, then one constructs all closed surfaces and stresses the importance of algebraic invariants as tools to distinguish topological spaces. Other low dimensional manifolds are analyzed; in particular, one proves that in dimension one the only manifolds are the interval, the circle, the half line, and the line. Then using the Heegaard decomposition, it is shown how 3-dimensional manifolds can be analyzed. We state (with no proof) the important Freedman theorem on the classification of simply connected 4-manifolds and finish presenting other manifolds that are important in different branches of mathematics, such as the Stiefel and the Grassmann manifolds.

Further on, the elements of homotopy theory are presented. In particular, the mappings of the circle into itself are analyzed introducing the important concept of degree. Homotopy equivalence of spaces is introduced and studied, as a coarser concept than that of homeomorphism.

The fundamental group is the first properly algebraic invariant introduced and then one gives the proof of the Seifert–van Kampen theorem that allows to compute the fundamental group of a space if one knows the groups of some of its parts. We use this important theorem to compute the fundamental groups of all closed surfaces, as well as of some orientable 3-manifolds, whose Heegaard decomposition is known. Thereafter, covering maps are introduced. This is an essential tool to analyze, from a different point of view, the fundamental group.

In the last chapter a short introduction to knot theory is presented, where one sees instances of the usefulness of the several algebraic invariants. On the one hand, the Jones polynomial is presented, and on the other, as an application of the fundamental group, the group of a knot is defined.

In this point I want recognize the big impact in this book of all experts that directly or indirectly have influenced my education as a mathematician and as a topologist. At the UNAM Guillermo Torres and Roberto Vázquez were decisive. Later on, in my doctoral studies in Heidelberg, Germany, I had the privilege of receiving directly the teaching of Albrecht Dold and Dieter Puppe; and indirectly, through several German topology books, of other people, among which I owe a mention to Ralph Stöcker and Heiner Zieschang's [23], as well as to that of Tammo tom Dieck [7].

Mexico City, Mexico Carlos Prieto
Winter 2010

Contents

Introduction

The object of this book is to introduce algebraic topology. But the first question is, what is topology? This is not an easy question to answer. Trying to define a branch of mathematics in a concise sentence is complicated. However, we may describe topology as that branch of mathematics that studies continuous deformations of geometrical objects. The purpose of topology is to classify objects, or at least to give methods in order to distinguish between objects that are not homeomorphic; that is, between objects that cannot be obtained from each other through a continuous deformation. It also provides techniques to study topological structures in objects arising in very different fields of mathematics. The concepts of "deformation," "continuity," and "homeomorphism" will be fundamental throughout the text. However, even not having yet precise definitions, we shall introduce in what follows several examples, that intuitively illustrate those concepts.

Figure 0.1 shows three topological spaces; namely, a sphere surface with its poles deleted, another sphere with its polar caps removed (including the polar circles), and a cylinder with its top and bottom removed (including the edges). These three objects can clearly be deformed to each other; hence topology would not distinguish them.

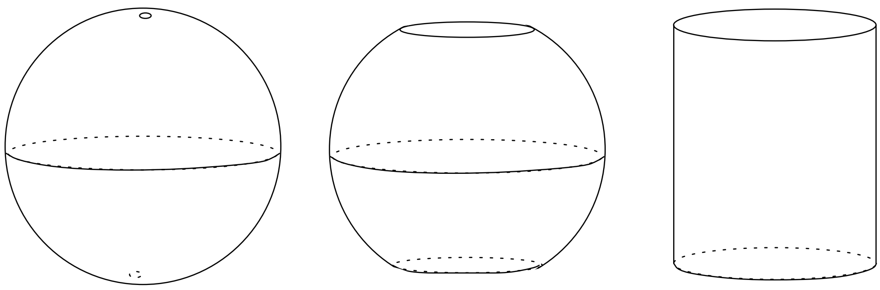


Figure 0.1 A sphere with no poles, a sphere with no polar caps and circles, and a cylinder with no edges

Figure 0.2 shows the surface of a torus (a "doughnut") and a sphere surface with a handle attached. Each one of these figures is clearly a deformation of the other. However, it is intuitively clear that the topological object shown in three forms in Figure 0.1 cannot be deformed to the topological object shown in two forms in Figure 0.2.

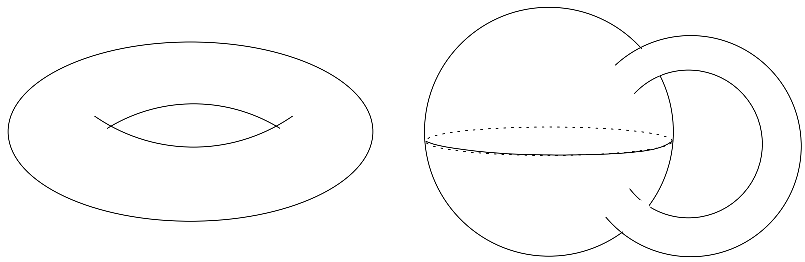


Figure 0.2 A torus and a sphere with one handle

Being a little more precise, we shall say that two objects (topological spaces) will be *homeomorphic* when there exists a one-to-one correspondence mapping points that are *close* in one of them to points that are close in the other. We can add to the list of those spaces that are not homeomorphic to each other, that we mentioned above, the following examples:

(a) Take $N = \{0, 1, \ldots, n-1\}$, $M = \{0, 1, \ldots, m-1\}$, $n < m$. Considered as topological spaces in any possible way, they will never be homeomorphic, since a necessary condition in order for two spaces to be homeomorphic, is that they have, as sets, the same cardinality.

(b) The same argument of (a) shows that a one-point set cannot be homeomorphic to an interval.

(c) More sophisticated arguments are required to show that a closed interval is not homeomorphic to a cross; that is, the topological spaces depicted in the upper part of Figure 0.3 are not homeomorphic to each other. A way of deciding this could be the following: Whatever point that we delete from the interval decomposes it in at most two connected portions. However, there is one point in the cross that when deleted, decomposes it into four pieces (components). For that reason, no point in the first space can exist that would correspond to this special point in the second space under a homeomorphism. Therefore, they cannot be homeomorphic.

(d) The surface of a torus is not homeomorphic to that of a sphere. This might be shown if we observe that a circle (simple closed curve) can be drawn on the surface of the torus that cannot be continuously contracted to a point. However, it is very clear that any circle drawn in the surface of a sphere can

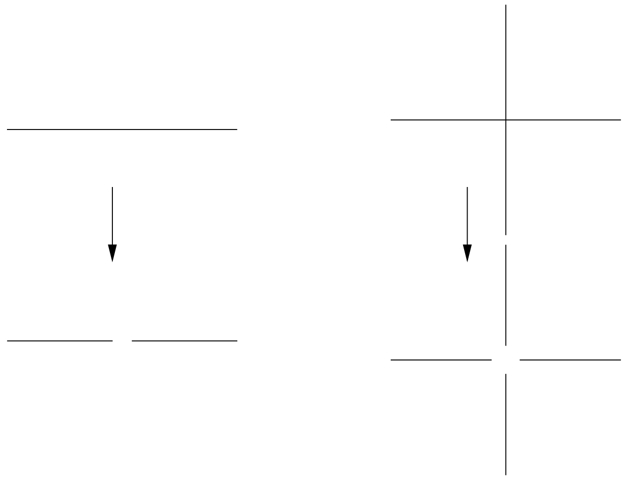


Figure 0.3 An interval and a cross are not homeomorphic

be deformed to a point, as can be appreciated in Figure 0.4. Another way of seeing this might be observing that any circle on the sphere is such that when deleted, the sphere is decomposed in two regions, while the circle on the torus does not have this property. (In other words, the famous theorem of the Jordan curve holds on the sphere, while it does not hold on the torus.)

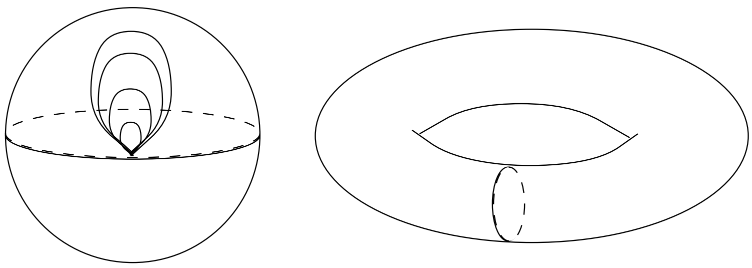


Figure 0.4 Any loop can be contracted on the sphere. A loop that does not contract on the torus

(e) The *Moebius band* that can be obtained from a strip of paper twisting it one half turn and then glueing it along its ends, is not homeomorphic to the *trivial band* obtained from a similar strip by glueing its ends without twisting it (see Figure 0.5).

The argument for showing this can be similar to that used in Example (d); namely, there is a circle on the Moebius band, that if removed from the band, does not disconnect it (it can be cut with scissors along the equator and would not fall apart in two pieces). However, in the trivial band any

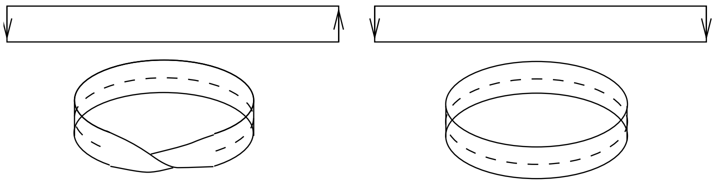


Figure 0.5 The Moebius band and the trivial band

circle parallel to and different from the edges (or any other circle different from the edges) when removed decomposes the band in two components (see Figure 0.6).

The first *exercises* for the reader are the following:

(f) Take the Moebius band and cut it along the equator. What space do you obtain? Will it be a Moebius band again? Or maybe will it be the trivial band?

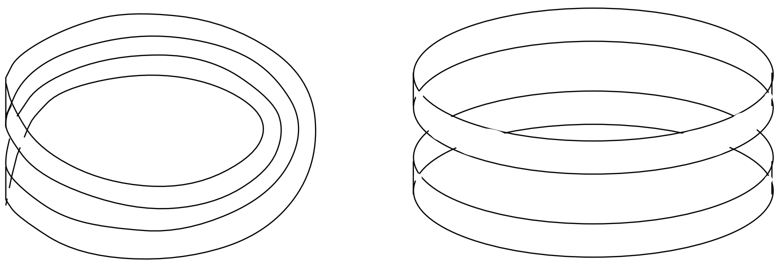


Figure 0.6 The Moebius band is not homeomorphic to the trivial band

(g) Similarly to the given construction of the Moebius band we may take a paper strip and glue its ends, but this time after twisting a full turn. Will this space be homeomorphic to the Moebius band? Or will this space be homeomorphic to the trivial band? What is the relationship between this space and that of (f) (see Figure 0.6.)

One of the central problems in topology consists, precisely, in studying topological spaces in order to be able to distinguish them. In all the previous examples every time we have decided that two spaces are not homeomorphic, it has been on the base of certain *invariants* that can be assigned to them. For instance, in (a) this invariant is the cardinality, in (c) it is the number of components obtained after deleting a point, and in (d) and (e) it is the number of components obtained after removing a circle. One of the objectives pursued by topology is to assign to each space invariants that are relatively easy to compute and allow to distinguish among them.

When we mentioned the intuitive concept of homeomorphism, we used the intuitive concept of nearness of two points, that is, we talked about the possibility of deciding if a point is close to a given point, or equivalently if it is in a *neighborhood* of the given point. In the first chapters of this book we shall make this concept precise.

A *knot* K is is a simple closed curve in the 3-space, i.e., it is the image $k(\mathbb{S}^1)$ under a "decent" inclusion $k : \mathbb{S}^1 \hookrightarrow \mathbb{R}^3$ of the unit circle in the plane into the 3-dimensional Euclidean space as intuitively shown in Figure 0.7.

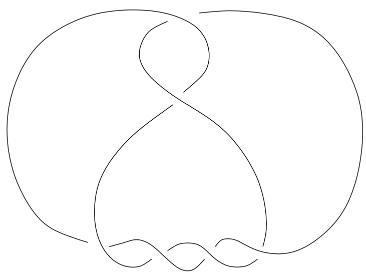


Figure 0.7 A knot in 3-space

Knot theory is an important field of mathematics with striking applications in several ambits of science. The central problem of the theory consists in determining when two given knots are *equivalent*; that is, when is it possible to deform inside the space one knot to the other without breaking it apart. A quite recent result in the theory proved by Gordon and Luecke [12] states that a knot K is determined by its complement, that is, that two knots K and K' are equivalent if and only if their complements $\mathbb{R}^3 - K$ and $\mathbb{R}^3 - K'$ are homeomorphic. In other words, they transformed the problem of classifyng knots into a homeomorphism problem of certain *open sets* in $\mathbb{R}^3$.

Chapter 1 Basic concepts

In this introductory chapter we present some basic topics which will be needed in the next chapters. They are common matter of point-set topology but, for the sake of completeness, we put them here. Thus we shall study some special examples of very basic topological spaces, we recall some special constructions under the general framework of attaching spaces, and we study the elementary aspects of group actions.

1.1 Some examples

In this section, we shall give the basic definitions needed for the theory of manifolds.

Before we start with the formal definitions, we come back to the homeomorphism problem described in the introduction. We shall ask the question if some of the well-known topological spaces are homeomorphic or not.

Let us start with the so-called *stereographic projection*

$$p : \mathbb{S}^n - N \longrightarrow \mathbb{R}^n \,, \tag{1.1.0}$$

where $N = (0, 0, \ldots, 0, 1) \in \mathbb{S}^n \subset \mathbb{R}^{n+1}$, given by

$$p(x) = \left(\frac{x_1}{1 - x_{n+1}}, \ldots, \frac{x_n}{1 - x_{n+1}} \right) \,,$$

where $x = (x_1, x_2, \ldots, x_{n+1}) \in \mathbb{S}^n - N$. The map p is a homeomorphism with inverse given by

$$p^{-1}(y) = \left(\frac{2y_1}{|y|^2 + 1}, \ldots, \frac{2y_n}{|y|^2 + 1}, \frac{|y|^2 - 1}{|y|^2 + 1} \right) \,,$$

where

$$y = (y_1, \ldots, y_n) \in \mathbb{R}^n \,.$$

The stereographic projection shows that $\mathbb{R}^n$ and $\mathbb{S}^n$ are "almost" equal, because up to a point they are homeomorphic. However $\mathbb{R}^n$ and $\mathbb{S}^n$ are not homeomorphic. For instance, in the case $n = 0$ this is clear since $\mathbb{R}^0$ is just one point, while $\mathbb{S}^0$ has two points. For $n > 0$, $\mathbb{R}^n$ is not compact because it is not bounded, while $\mathbb{S}^n$ is compact, since it can be seen as a closed and bounded set in $\mathbb{R}^{n+1}$ (as a matter of fact, $\mathbb{S}^n$ is the one-point compactification of $\mathbb{R}^n$, see [21, 6.3]).

1.1.1 Exercise. Let now $q : \mathbb{S}^n - S \longrightarrow \mathbb{R}^n$ be the stereographic projection, but now from the south pole $S = (0, 0, \dots, 0, -1) \in \mathbb{S}^n \subset \mathbb{R}^{n+1}$. Notice that q is given by

$$q(x) = \left(\frac{x_1}{1 + x_{n+1}}, \dots, \frac{x_n}{1 + x_{n+1}} \right) . \tag{1.1.1}$$

Show that the composite $q \circ p^{-1} : \mathbb{R}^n - 0 \longrightarrow \mathbb{R}^n - 0$ is the *inversion* with respect to the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n - 0$, which is given by the map $\iota : \mathbb{R}^n - 0 \longrightarrow \mathbb{R}^n - 0$, with the property that the points y and $\iota(y)$ together with the origin O are collinear and on the same side of O. Moreover, the product of their distances to O is 1, namely $|y| \cdot |\iota(y)| = 1$ (see Figure 1.1). (*Hint:* It is enough to prove that for $x \in \mathbb{S}^n$ the scalar product $\langle p(x), q(x) \rangle = 1$.)

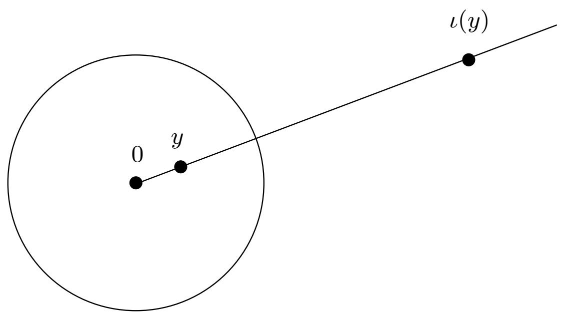


Figure 1.1 Inversion with respect to the unit circle

Questions such as to decide if there are homeomorphisms $\mathbb{B}^n \approx \mathbb{S}^n$, $\mathbb{S}^m \times \mathbb{S}^n \approx \mathbb{S}^{m+n}$, $\mathbb{R}^m \approx \mathbb{R}^n$, $\mathbb{B}^m \approx \mathbb{B}^n$, $\mathbb{S}^m \approx \mathbb{S}^n$ if $m \neq n$ in the last three cases, are much more complicated and their solution can be obtained using more subtle methods. They can be answered using the techniques of algebraic topology, which yields a negative answer to each of the given questions.

In what follows, we shall state and prove an interesting result about homeomorphisms between subspaces of $\mathbb{R}^n$.

1.1.2 Definition. A subset X of $\mathbb{R}^n$ is said to be *convex* if for any two points x_0, $x_1 \in X$ the *line segment* $[x_0, x_1]$ that joins them lies in X, namely, for every $t \in I$, the point $(1 - t)x_0 + tx_1 \in X$.

1.1.3 **Theorem**. *Let X be a convex compact subset of $\mathbb{R}^n$ whose interior X° is nonempty. Then X is homeomorphic to the n-ball $\mathbb{B}^n$ by means of a homeomorphism, which maps the boundary ∂X homeomorphically onto $\mathbb{S}^{n-1}$.*

Proof: Fix an interior point $x_0 \in X^\circ$. Given any point $x \in \partial X$, it has the property that it is the only boundary point which lies in the semiline that starts at x_0 and goes through x. Namely

$$\{(1-t)x_0 + tx \mid t > 0\} \cap \partial X = \{x\}\,.$$

Otherwise, X would not be convex. Take the map

$$\psi : \partial X \longrightarrow \mathbb{S}^{n-1}\,,$$

given by $\psi(x) = (x - x_0)/|x - x_0|$. This is a continuous bijective map. Since both spaces are compact Hausdorff, ψ is a homeomorphism. If we extend ψ radially we get a homeomorphism

$$\varphi : X \longrightarrow \mathbb{B}^n\,,$$

given by $\varphi((1-t)x_0 + tx) = t\psi(x)$, $t \in I$, as desired. $\square$

1.1.4 **Corollary**. *Let $X \subset \mathbb{R}^n$ be a nonempty, open, bounded, convex set. Then X is an n-cell, that is, $X \approx \overset{\circ}{\mathbb{B}}{}^n$.*

Proof: The assumption implies that the closure $\overline{X} = X \cup \partial X$ is a compact, convex set with nonempty interior. Thus, by 1.1.3, there is a homeomorphism $\varphi : \overline{X} \longrightarrow \mathbb{B}^n$ such that $\varphi(\partial X) = \mathbb{S}^{n-1}$. Hence, φ, by restriction, induces a homeomorphism $X = \overline{X} - \partial X \longrightarrow \mathbb{B}^n - \mathbb{S}^{n-1} = \overset{\circ}{\mathbb{B}}{}^n$. $\square$

1.1.5 EXERCISE. Prove that the previous corollary equally holds if $X \subset \mathbb{R}^n$ is a nonempty open convex set.

1.1.6 EXAMPLES.

(a) The cube I^n is a convex, compact subset of $\mathbb{R}^n$, whose interior is nonempty. Thus it is homeomorphic to $\mathbb{B}^n$. Moreover, the open cube $(0,1)^n$ is an n-cell.

(b) The product $\mathbb{B}^m \times \mathbb{B}^n$ is a convex, compact subset of $\mathbb{R}^m \times \mathbb{R}^n = \mathbb{R}^{m+n}$, with nonempty interior. Thus it is homeomorphic to $\mathbb{B}^{m+n}$. Also the product $\overset{\circ}{\mathbb{B}}{}^m \times \overset{\circ}{\mathbb{B}}{}^n$ is an $(m+n)$-cell (cf. 2.1.7).

1.1.7 EXERCISE. Show explicit homeomorphisms for 1.1.6 (a) and (b).

For the time being we shall state a deep theorem, the domain invariance theorem, whose proof will postponed to Chapter 3 (see Theorem 3.5.5). The result has strong consequences in questions like the ones given above. As a consequence, we shall deduce from the domain invariance theorem, two very interesting results, the dimension and the boundary invariance theorems.

1.1.8 **Theorem**. (Domain invariance) *Take subsets* $X, Y \subset \mathbb{R}^n$ *such that* X *is homeomorphic to* Y*. Then, if* X *is open in* $\mathbb{R}^n$*,* Y *must be open too.*

Notice that if we assume that there is a homeomorphism $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ which maps X onto Y, the result would be trivial. The statement of the theorem of invariance of domain requires only the existence of a homeomorphism $\psi : X \longrightarrow Y$. Indeed, the statement of the theorem is equivalent to saying that the property of being open in $\mathbb{R}^n$ is a topological invariant of the subsets of $\mathbb{R}^n$.

In what follows we shall deduce several consequences of Theorem 1.1.8.

1.1.9 **Theorem**. (Dimension invariance) *If* $m \neq n$*, then* $\mathbb{R}^m \not\approx \mathbb{R}^n$*,* $\mathbb{S}^m \not\approx \mathbb{S}^n$ *and* $\mathbb{B}^m \not\approx \mathbb{B}^n$*.*

Proof: Assume $m < n$. Then $\mathbb{R}^m \subset \mathbb{R}^n$ is not open in $\mathbb{R}^n$. However $\mathbb{R}^n \subset \mathbb{R}^n$ is open in $\mathbb{R}^n$. Hence by Theorem 1.1.8, $\mathbb{R}^m$ and $\mathbb{R}^n$ cannot be homeomorphic.

If we assume that there is a homeomorphism $\varphi : \mathbb{S}^m \approx \mathbb{S}^n$, then taking off a point from each one of the spheres, say p from $\mathbb{S}^m$ and $q = \varphi(p)$ from $\mathbb{S}^n$, by restriction we obtain a homeomorphism $\varphi' : \mathbb{S}^m - p \approx \mathbb{S}^n - q$. Via the stereographic projections given in 1.1.0 and 1.1.1, φ' determines a homeomorphism $\psi : \mathbb{R}^m \approx \mathbb{R}^n$. Therefore $m = n$ by the first part of the proof.

Finally take $m < n$ and a homeomorphism $\varphi : \mathbb{B}^m \longrightarrow \mathbb{B}^n$. Hence $\mathring{\mathbb{B}}^n \subset \mathbb{R}^n$ is open and homeomorphic to $\varphi^{-1}(\mathring{\mathbb{B}}^n) \subset \mathbb{B}^m \subset \mathbb{R}^m \subset \mathbb{R}^n$, which is not open in $\mathbb{R}^n$. This contradicts 1.1.8. □

Another interesting application of the domain invariance theorem is the following result.

1.1.10 **Theorem**. (Invariance of boundary) *Let* $\varphi : \mathbb{B}^n \longrightarrow \mathbb{B}^n$ *be a homeomorphism. Then* $\varphi(\mathbb{S}^{n-1}) = \mathbb{S}^{n-1}$*.*

Proof: Take $x \in \mathbb{S}^{n-1}$ and assume that $y = \varphi(x) \in \mathring{\mathbb{B}}^n$. Let $\varepsilon > 0$ be such that the open ball $B^\circ_\varepsilon(y) \subset \mathbb{B}^n$. The inverse image $\varphi^{-1}(B^\circ_\varepsilon) \subset \mathbb{R}^n$ is homeomorphic a B°_ε but since $\varphi^{-1}(B^\circ_\varepsilon) \subset \mathbb{B}^n$ and $\varphi^{-1}(B^\circ_\varepsilon) \cap \mathbb{S}^{n-1} \neq \emptyset$, this is not open in $\mathbb{R}^n$, which contradicts 1.1.8. □

1.1.11 DEFINITION. A topological space B homeomorphic to $\mathbb{B}^n$ is called an n*-ball* and a topological space S homeomorphic to $\mathbb{S}^{n-1}$ is called an $(n-1)$*-sphere*. Let $\varphi : B \longrightarrow \mathbb{B}^n$ be a homeomorphism. The $(n-1)$-sphere $B^\bullet = \varphi^{-1}(\mathbb{S}^{n-1})$ is called the *boundary* of B. (As in more general cases, in what follows we shall denote $B^\bullet$ by ∂B.)

1.1.12 Note. We abuse the language by naming “boundary” this subset of a ball. However, this will not produce confusion since it corresponds to the boundary of a Euclidean ball. Thus it is important to call the boundary of a subset A of a topological space X boundary of A *in* X, to distinguish it from the “intrinsic boundary” of a ball.

1.1.13 **Proposition**. *The boundary concept of 1.1.11 is well defined.*

Proof: If $\psi : B \longrightarrow \mathbb{B}^n$ is another homeomorphism, then by 1.1.10, $\psi^{-1}\varphi(\mathbb{S}^{n-1}) = \mathbb{S}^{n-1}$. Hence $\varphi(\mathbb{S}^{n-1}) = \psi(\mathbb{S}^{n-1})$ and the definition of $B^\bullet$ given in1.1.11 is independent of the choice of the homeomorphism φ. □

1.2 Special constructions

In this section, we shall analyze some special constructions which will play a role in what follows. Consider the continuous maps

$$X \xleftarrow{f} A \xrightarrow{g} Y\,.$$

1.2.1 Definition. The *double attaching space* is defined as the identification space of $X \sqcup Y$, which we denote by $Y_g \cup_f X$, given by identifying $f(a) \in X$ with $g(a) \in Y$ for all $a \in A$. That is,

$$Y_g \cup_f X = X \sqcup Y/f(a) \sim g(a)\,, \;\; a \in A\,.$$

Schematically it looks as shown in Figure 1.2.

1.2.2 Remark. The diagram

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "g"'] & X \arrow[d, "i"] \\ Y \arrow[r, "j"'] & Y_g \cup_f X \end{tikzcd}$$

is a so-called *cocartesian square.* or *pushout diagram* The double attaching space $Y_g \cup_f X$ is also called the *pushout* of (f, g). It is characterized by the pushout universal property (see Exercise 1.2.4):

(PO) If $\varphi : X \longrightarrow Z$ and $\psi : Y \longrightarrow Z$ are continuous maps such that $\varphi \circ f = \psi \circ g$, then there is a unique map $\xi : Y_g \cup_f X \longrightarrow Z$ such that $\xi \circ i = \varphi$ and $\xi \circ j = \psi$.

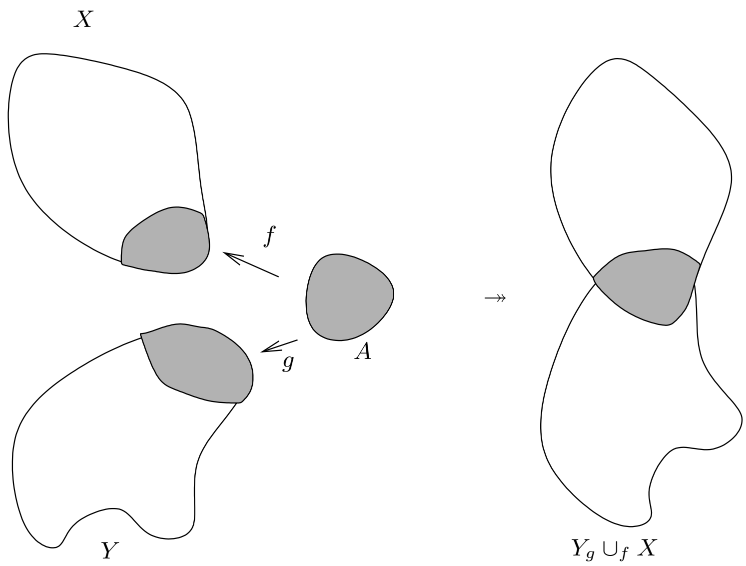


Figure 1.2 The double attaching space

1.2.3 Examples.

(a) If $A \subset Y$ and $g : A \hookrightarrow Y$ is the inclusion, then $Y_g \cup_f X$ is called the *attaching space* of $f : A \longrightarrow X$ and is denoted simply by $Y \cup_f X$.

(b) If $A = \emptyset$, then $Y_g \cup_f X = X \sqcup Y$.

(c) If $A \subset X$, $Y = \{*\}$, and $f : A \hookrightarrow X$ is the inclusion map, then $Y_g \cup_f X = X/A$.

1.2.4 Exercise. Consider the maps $X \xleftarrow{f} A \xrightarrow{g} Y$ and their double attaching space $Y_g \cup_f X$. Show that $Y_g \cup_f X$ is indeed a pushout, namely show that it is characterized by the property (PO) given above.

The construction of the attaching space is a very important construction, since with it many associated constructions can be obtained, that play a relevant role in several branches of topology. Many of them are based in a particular space associated to a given topological space that we define in what follows.

1.2.5 Definition. Let $f : X \longrightarrow Y$ be continuous and consider the diagram

$$X \times I \xhookleftarrow{i_0} X \xrightarrow{f} Y\,,$$

where $i_0(x) = (x, 0)$. The attaching space $Y \cup_f (X \times I)$ is called the *mapping cylinder* of f and is denoted by M_f.

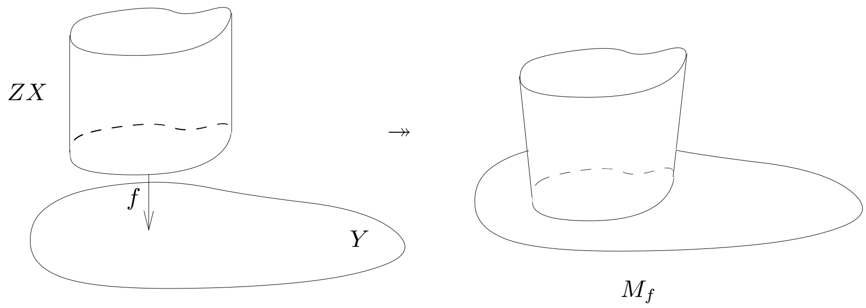


Figure 1.3 The mapping cylinder

If $Y = X$ and $f = \mathrm{id}_X$, then $M_f = X \times I$, and this space is called simply the *cylinder* over X and is denoted by ZX. The quotient space of the cylinder over X, in which the top is collapsed to a point, $ZX/X \times \{1\} = X \times I/X \times \{1\}$ is called the *cone* over X and is denoted by CX. There is a natural inclusion $X \hookrightarrow CX$ given by $x \mapsto q(x, 0)$, where $q : ZX \longrightarrow CX$ is the quotient map.

1.2.6 Definition. Let $f : X \longrightarrow Y$ be continuous and consider the diagram

$$CX \hookleftarrow X \xrightarrow{f} Y \,.$$

The attaching space $Y \cup_f CX$ is called the *mapping cone* of f and is denoted by C_f.

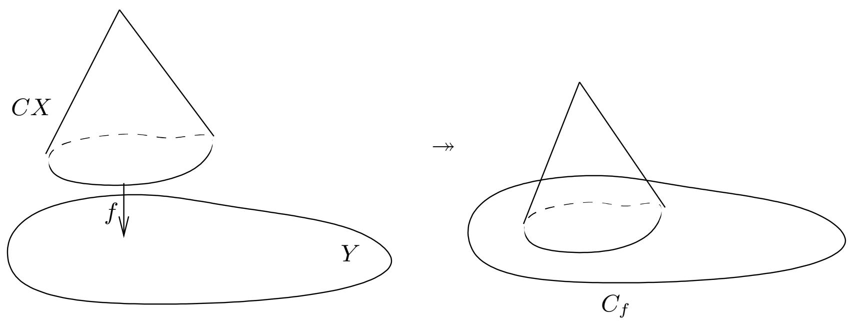


Figure 1.4 The mapping cone

1.2.7 Exercise. Prove that $C\mathbb{S}^{n-1} \approx \mathbb{B}^n$.

In what follows, we define a new space obtained from a given space, that plays an important role in algebraic topology, mainly in homotopy theory.

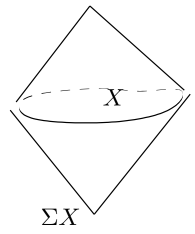


Figure 1.5 The suspension

1.2.8 Definition. Take $Y = *$. The mapping cone of $f : X \longrightarrow *$, namely $* \cup_f CX$, is called the *suspension* of X and is denoted by ΣX.

1.2.9 Exercise. Prove that the suspension of a space X coincides with the quotient of the cylinder $ZX = X \times I$ obtained by collapsing the *top* of the cylinder $X \times \{1\}$ into a point, and the *bottom* of the cylinder $X \times \{0\}$ into another point.

1.2.10 Exercise. Prove that the suspension of a space X is homeomorphic to the double attaching space corresponding to the diagram

$$CX \hookleftarrow X \hookrightarrow CX \,,$$

or, equivalently, to the mapping cone of the inclusion $X \hookrightarrow CX$.

Another interesting construction is the following.

1.2.11 Definition. Let $f : X \longrightarrow X$ be continuous. The quotient space of the cylinder over X, ZX, obtained after identifying the bottom of the cylinder with the top through the identification $(x, 0) \sim (f(x), 1)$, is called the *mapping torus* of f and is denoted by T_f.

1.2.12 Examples.

(a) If $f = \mathrm{id}_X : X \longrightarrow X$, then the mapping torus T_f is called *torus* of X and is denoted by TX.

(b) If $X = I$, then TI is the *trivial band* or *standard cylinder.*

(c) If $X = I$, and $f : I \longrightarrow I$ is such that $f(t) = 1 - t$, then T_f is the *Moebius band.*

(d) If $X = \mathbb{S}^1$, then $T\mathbb{S}^1$ is the *standard torus* and is denoted by $\mathbb{T}^2$.

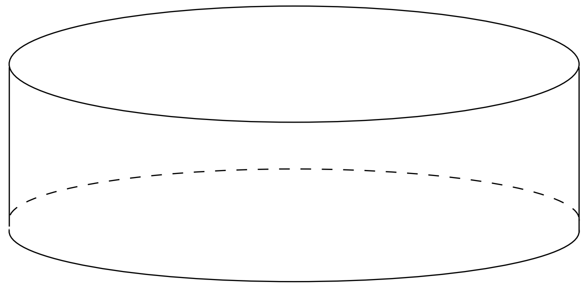


Figure 1.6 The trivial band

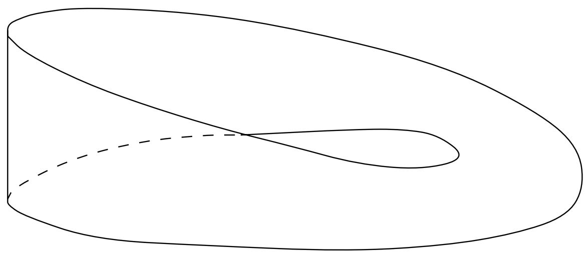


Figure 1.7 The Moebius band

(e) If $X = \mathbb{S}^1 = \{\zeta = \mathrm{e}^{2\pi \mathrm{i}t} \in \mathbb{C} \mid t \in I\}$ and $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ is such that $f(\mathrm{e}^{2\pi\mathrm{i}t}) = \mathrm{e}^{-2\pi\mathrm{i}t}$, then T_f is the *Klein bottle*.

1.2.13 EXERCISE. Prove that the trivial band is homeomorphic to the space obtained from the *square* $I \times I$ by identifying each point $(s, 0)$ with $(s, 1)$, while the Moebius band is homeomorphic to the space obtained from the square by identifying $(s, 0)$ with $(1 - s, 1)$.

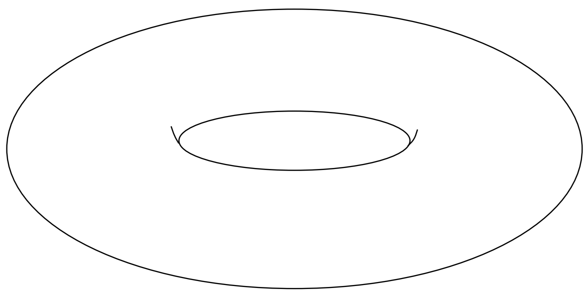


Figure 1.8 The standard torus

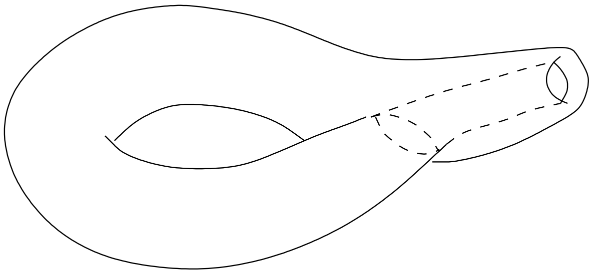


Figure 1.9 The Klein bottle

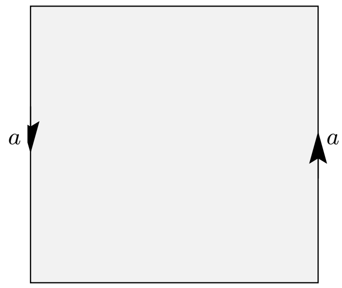


Figure 1.10 The construction of the Moebius band

1.2.14 Exercise. Prove that the standard torus is homeomorphic to the space obtained from the square $I \times I$ by identifying the points $(s, 0)$ with $(s, 1)$, and $(0, t)$ with $(1, t)$, while the Klein bottle is obtained from the square by identifying $(s, 0)$ with $(1 - s, 1)$, and $(0, t)$ with $(1, t)$. This is the classical definition of the bottle. We shall see below in 1.2.20 a different way of defining it.

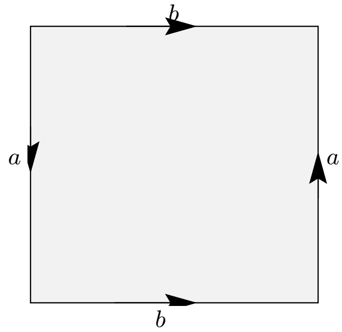


Figure 1.11 The construction of the Klein bottle

1.2.15 Exercise. Prove that the Klein bottle is a quotient space of the Moebius band. Describe the identification.

1.2.16 EXERCISE. Let $q : \mathbb{S}^{n-1} \longrightarrow \mathbb{RP}^{n-1}$ be the quotient map such that $q(x) = q(-x)$, and let

$$\mathbb{B}^n \xleftarrow{\ i\ } \mathbb{S}^{n-1} \xrightarrow{\ q\ } \mathbb{RP}^{n-1} \ .$$

Prove that $\mathbb{RP}^{n-1} \cup_q \mathbb{B}^n \approx \mathbb{RP}^n$.

1.2.17 DEFINITION. Consider the diagram

$$\mathbb{B}^n \xleftarrow{\ i\ } \mathbb{S}^{n-1} \xrightarrow{\ \varphi\ } X \ .$$

The resulting attaching space $X \cup_\varphi \mathbb{B}^n$ is said to be obtained by *attaching a cell* of dimension n to the space X. Frequently this space is denoted by $X \cup_\varphi e^n$. The image of $\mathbb{B}^n$, resp. $\mathring{\mathbb{B}}^n$, in $Y = X \cup_\varphi \mathbb{B}^n$ is called *closed cell*, resp. *open cell* de Y, and φ is called *characteristic map* of the cell of Y.

1.2.18 EXERCISE. Prove that $\mathbb{RP}^n$ is obtained from $\mathbb{RP}^{n-1}$ by attaching a cell of dimension n with the canonical map $q : \mathbb{S}^{n-1} \longrightarrow \mathbb{RP}^{n-1}$ as characteristic map. Conclude that the projective space $\mathbb{RP}^n$ is obtained by successively attaching cells of dimensions $1, 2, \ldots, n$ to the singular space $\{*\}$; that is,

$$\mathbb{RP}^n = \{*\} \cup e^1 \cup e^2 \cup \cdots \cup e^n \,.$$

(*Hint:* $\mathbb{RP}^n$ is obtained from $\mathbb{B}^n$ by identifying in its boundary $\mathbb{S}^{n-1} = \partial\mathbb{B}^n$ each pair of antipodal points in one point.)

1.2.19 EXERCISE. Take $\mathbb{B}^n \hookleftarrow \mathbb{S}^{n-1} \hookrightarrow \mathbb{B}^n$. Prove that the corresponding attaching space is homeomorphic to $\mathbb{S}^n$, namely, $\mathbb{S}^n$ is obtained from $\mathbb{B}^n$ by attaching a cell of dimension n with the inclusion as characteristic map. Equivalently, $\mathbb{S}^n$ is obtained from $\mathbb{S}^{n-1}$ by attaching two cells of dimension n, each with $\mathrm{id}_{\mathbb{S}^{n-1}}$ as characteristic map. Conclude that the n-sphere can be decomposed as

$$\mathbb{S}^n = \mathbb{S}^0 \cup (e_1^1 \cup e_2^1) \cup (e_1^2 \cup e_2^2) \cup \cdots \cup (e_1^n \cup e_2^n) \,.$$

The procedure of decomposing a space and then putting it together again is called *cutting and pasting*. We have the following.

1.2.20 EXAMPLES.

1. Consider the torus $\mathbb{T}^2$ embedded in 3-space in such a way that it is symmetric with respect to the origin (namely, so that $x \in \mathbb{T}^2$ if and only if $-x \in \mathbb{T}^2$, see Figure 1.12 (a)). We shall show using the method of cutting and pasting, that if we identify each point $x \in \mathbb{T}^2$ with its opposite $-x$, then we obtain

the Klein bottle. The result of this identification is the same as if we slice the torus along the inner and outer equators and keep the upper (closed) part, so that we obtain a ring (see Figure 1.12 (b)), and then identify with each other antipodal points of the outer circle and those of the inner circle. This is marked with double and single arrows. Now cut the ring into two halves along the dotted lines. Up to a homeomorphism we obtain two rectangles, where we use different types of arrows to codify what has to be identified (see Figure 1.12 (c)). Flip the top rectangle (Figure 1.12 (d)) and identify the edges marked with the single solid arrow (Figure 1.12 (e)). We obtain a square such that after realizing the identifications marked therein, we obtain the Klein Bottle (cf. Exercise 1.2.14).

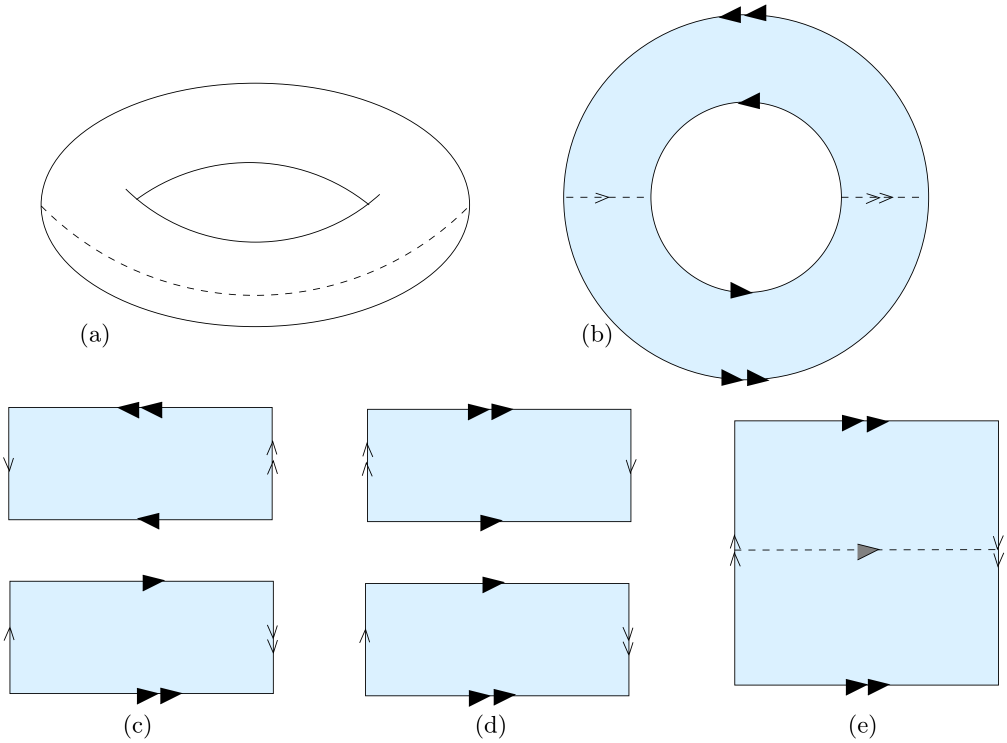


Figure 1.12 Cutting and pasting the torus

2. Consider the square $I \times I$ and identify on the boundary the points of the form $(1, t)$ with $(1 - t, 1)$, and those of the form $(0, 1 - t)$ with $(t, 0)$. This is illustrated in Figure 1.13 (a), where edge a is identified with edge a' and edge b with edge b', preserving the counterclockwise orientation (cf. 2.2.16 below). Now cut the square along the diagonal form $(0, 0)$ to $(1, 1)$. Calling the new edges c and c', one obtains two triangles with sides a, b, c and a', b', c' oriented as shown in Figure 1.13 (b). We may now glue both triangles along the edges a and a' preserving the orientations (turning and flipping the first

triangle), and so we obtain a quadrilateral with edges b, c, b', c' shown in Figure 1.13 (c). Clearly this quadrilateral is homeomorphic (via an affine map) to a square with edges equally denoted shown in Figure 1.13 (d). Now observe that in the resulting square one identifies the vertical sides with the same orientation, namely $(0, t)$ with $(1, t)$, and the horizontal edges with the opposite direction, namely $(t, 0)$ with $(1 - t, 1)$. The result of this last identification is the Klein bottle defined in 1.2.14.

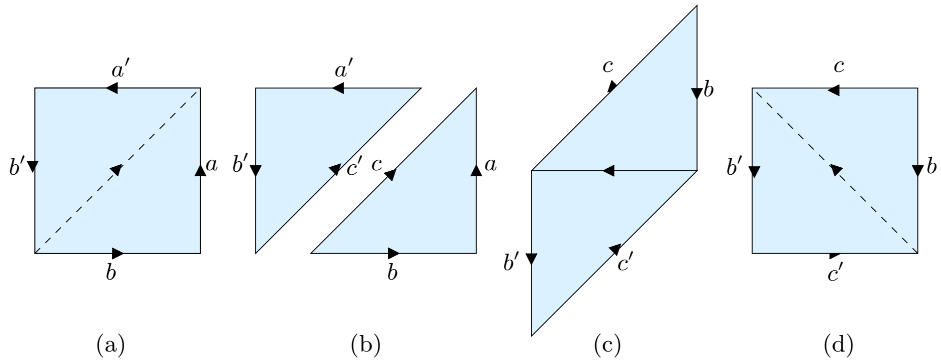


Figure 1.13 Cutting and pasting the Klein bottle

1.2.21 Exercise. Using the method of cutting and pasting, show that the result of cutting the Moebius band along the equator yields the trivial band, namely a space homeomorphic to $\mathbb{S}^1 \times I$. Notice that if you realize that in the 3-space, one obtains a band with a double twisting, as shown on Figure 1.14.

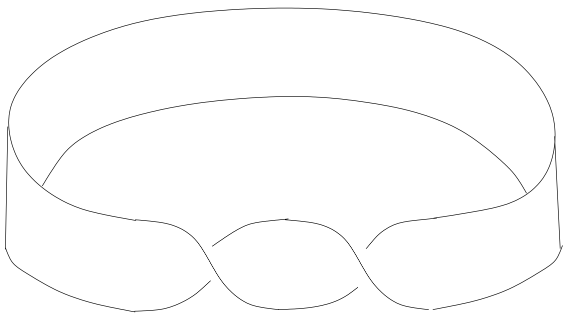


Figure 1.14 Cutting the Moebius band yields the trivial band

1.3 Path connectedness

A fundamental concept in algebraic topology is that of a path, which we shall study thoroughly in Chapter 4. In what follows, as it is usual, we shall denote by

I the real *unit interval* $[0,1] = \{t \in \mathbb{R} \mid 0 \leq t \leq 1\}$. Given a topological space X, and two points $x_0, x_1 \in X$, a *path* in X from x_0 to x_1 is a map $\omega : I \longrightarrow X$ such that $\omega(0) = x_0$ and $\omega(1) = x_1$. We denote this fact by $\omega : x_0 \simeq x_1$. It is easy to show that the relation $\simeq$ is an equivalence relation. Hence a space X is decomposed in equivalence classes, which we call *path-components* of X. We denote the set of path-components of X by $\pi_0(X)$.

We start by recalling that a topological space X is *connected* if it is impossible to find a *separation* (A, B) of X, namely two nonempty disjoint open sets $A, B \subset X$ such that $A \cup B = X$. Otherwise, X is *disconnected* and such a pair (A, B) is called a *disconnection* of X. In symbols, a disconnection satisfies

$$X = A \cup B\,, \qquad A \cap B = \emptyset\,, \qquad A \neq \emptyset \neq B\,.$$

It is easy to show that *X is disconnected if and only if there is a continuous surjective map $\delta : X \longrightarrow \{0,1\}$.*

1.3.1 Definition. A topological space X is said to be *path connected* if given any two points $x_0, x_1 \in X$, there is a path in X from x_0 to x_1. In other words, X is path connected if and only if it has one path-component.

1.3.2 **Theorem**. *A subspace $X \subseteq \mathbb{R}$ is path connected if and only if X is an interval (open, half-open, or closed) or it is a ray (open or closed).*

Proof: Let X be an interval or a ray and assume it is not connected. Then there is a continuous surjective map $\delta : X \longrightarrow \{0,1\}$. Let $x_0, x_1 \in X$ be such that $\delta(x_0) = 0$ and $\delta(x_1) = 1$. Since X is an interval, then it contains the interval $[x_0, x_1]$. The intermediate value theorem implies that there is some $x \in [x_0, x_1]$ such that $\delta|_{[x_0,x_1]}(x) = \frac{1}{2}$, which is a contradiction.

Conversely, let X be a nonempty connected subspace of $\mathbb{R}$. Assume that $x_0, x_1\ X$ and that $x_0 < x < x_1$. If $x \notin X$, then $A = X \cap (-\infty, x)$ and $B = X \cap (x, +\infty)$ form a disconnection of X, which is a contradiction. Thus, with x_0 and x_1, X contains all elements x that lie between them. This a characterization of an interval. $\square$

1.3.3 **Proposition**. *If a space X is path connected, then it is connected.*

Proof: If X is not connected, take a disconnection (A, B) of X and take points $x_0 \in A$ and $x_1 \in B$. Since X is path connected, there is a path $\omega : I \longrightarrow X$ such that $\omega(0) = x_0$ and $\omega(1) = x_1$. Then the sets $A' = \omega^{-1}(A)$ and $B' = \omega^{-1}(B)$ form a disconnection of the unit interval I. This contradicts 1.3.2. $\square$

Using a similar idea to the one used in previous proof, one can show that *if X is connected and $f : X \longrightarrow Y$ is continuous, then $f(X)$ is connected.*

1.3.4 EXERCISE. Show the last statement, namely, if X is connected and $f : X \longrightarrow Y$ is continuous, then $f(X)$ is connected. Furthermore, show that if X is path connected and $f : X \longrightarrow Y$ is continuous, then $f(X)$ is path connected.

Given a set $C \subset X$, we denote its *closure* in X by $\overline{C}$, ant is given by $\overline{C} = \{x \in X \mid U \cap C \neq \emptyset \text{ for every neighborhood } U \text{ of } x\}$. An important theorem on connectedness is the following.

1.3.5 **Theorem**.

(a) *Let $\{X_\lambda \mid \lambda \in \Lambda\}$ be a family of connected subspaces of X, such that the intersection $\bigcap_{\lambda\in\Lambda} X_\lambda \neq \emptyset$. Then $Y = \bigcup_{\lambda\in\Lambda} X_\lambda$ is connected.*

(b) *If $C \subset X$ is a connected subspace and D is such that $C \subseteq D \subset \overline{C}$, then D is a connected subspace too. In particular, $\overline{C}$ is connected.*

Proof: (a) Let $f : Y \longrightarrow \{0, 1\}$ be a continuous map. Since X_λ is connected and the restriction $f_\lambda : X_\lambda \longrightarrow \{0, 1\}$ is continuous, then f_λ is constant. But the intersection of all X_λ is nonempty, thus f is constant and so Y must be connected.

(b) Suppose on the contrary that (A, B) is a disconnection of D in X, that is,

$$D \subset A \cup B\,, \qquad (A \cap B) \cap D = \emptyset\,, \qquad A \cap D \neq \emptyset \neq B \cap D\,.$$

But since A and B are open, then $A \cap C \neq \emptyset \neq B \cap C$, because the points in D which are not in C must be cluster points. Thus (A, B) would be a disconnection of C in X, which is a contradiction. □

If we take a point $x \in X$ and consider the set $C_x = \bigcup\{C \subset X \mid x \in C \text{ and } C \text{is connected}\}$, then C_x is the largest connected set that contains x. It is called the *connected component* of x. By Theorem 1.3.5 (b), C_x is closed.

A space X is said to be *locally (path) connected*, if for each point $x \in X$ and for each neighborhood U of x in X, there is a (path) connected neighborhood V of x in X such that $V \subset U$.

1.3.6 **Theorem**.

(a) *Let X be locally connected and let $U \subset X$ be open. Then the connected components of U are open.*

(b) *Let X be locally path connected and let $U \subset X$ be open and connected. Then U is path connected. Furthermore, the path-components of an open set $V \subset X$ coincide with the connected components of V and they are open.*

Proof: (a) Let $C \subset U$ be the connected component of x in U and take $y \in C$. Then $C = C_y$. Since U is a neighborhood of y in X, there is a connected neighborhood V of y in X such that $V \subset U$. Hence $V \subset C$, because $C = C_y$ is the maximal connected set in X in which y lies. Thus C is open in X.

(b) Take $x \in U$ and let c_x be the path-component of U in which x lies. If $y \in c_x$, then there is a path connected neighborhood V of y such that $V \subset U$. Hence we have $V \subset c_x$ and thus c_x is open. Since the complement $U - c_x$ is a union of path-components, each of which is open, then it is open. But U is connected. Hence $U - c_x = \emptyset$, i.e., $U = c_x$ is path connected.

Combining (a) with the latter, we obtain the second part of (b). □

For $A \subset X$, we denote by ∂A the (topological) *boundary* of A in X, namely $\partial A = \{x \in X \mid U \cap A \neq \emptyset \neq U \cap (X - A)$ for each neighborhood U of $x\}$. If A° denotes the interior, given by $A^\circ = X - \overline{X - A}$, where $\overline{C}$ denotes the closure of C, then $\partial A = \overline{A} - A^\circ$. We have the following.

1.3.7 **Theorem**.

(a) *If A is open in X and it is a connected component of the open set B in X, then $\partial A \subset \partial B$.*

(b) *If A is an open connected subset of X, then A is a connected component of $X - \partial A$.*

Proof: (a) First notice that A is closed in B, hence $A = \overline{A} \cap B$. Since A and B are open in X we have

$$\partial A = \overline{A} - A = \overline{A} - (B \cap A) = \overline{A} \cap (X - B) \subset \overline{B} \cap (X - B) = \partial B\,.$$

(b) Let B be the connected component of $X - \partial A$ which contains A. If $A \neq B$, then as well $B \cap A \neq \emptyset$ as $B \cap (X - A) \neq \emptyset$. Thus the pair $(B \cap A, B \cap (X - \overline{A}))$ would be a disconnection of B, which is a contradiction. □

Another useful theorem is the next.

1.3.8 **Theorem**. *Let X be a connected space and let $A \subset X$ be a connected subspace. Furthermore, let C be a connected component of $X - A$. Then one has the following:*

(a) *If $U \subset X - C$ is open and closed in $X - C$, then $U \cup C$ is connected.*

(b) *$X - C$ is connected.*

Proof: (a) If (P, Q) is a disconnection of $U \cup C$, then C is contained in P, say. Hence $Q \subset U$. Then Q is open and closed in U. Since U is open and closed in $X - C$, Q is open and closed in $X - C$. Thus Q is open and closed in $(X - C) \cup (C \cup U) = X$, but this contradicts the connectedness of X.

(b) If (P', Q') is a disconnection of $X - C$, then we shall prove that $(A \cap P', A \cap Q')$ is a disconnection of A. If we assume that $A \cap P' = \emptyset$, then by (a), $C \cup P'$ is connected and it is contained in $X - A$. Since C is a proper subset of $C \cup P'$, this contradicts the fact that C is a connected component of $X - A$. Therefore $A \cap P' \neq \emptyset$ and similarly $A \cap Q' \neq \emptyset$. □

If the spaces involved are Hausdorff spaces, we have interesting relationships between connectedness and compactness.

1.3.9 **Theorem**. *Let X be a compact Hausdorff space. If C is a connected component of X, then the the open and closed neighborhoods of C form a neighborhood basis of C in X.*

Proof: We shall consider the case of a compact metric space X with metric d. A *t-chain*, $t > 0$, from x to x' in X is a family $\{x = x_0, \ldots, x_k = x'\}$ such that $d(x_j, x_{j+1}) < t$ for $1 \leq j < k$. We define a relation in X by $x \sim_t x'$ if and only if there is a t-chain from x to x'. This is clearly an equivalence relation. Furthermore, the equivalence class $K_x(t)$ of all points x' such that $x \sim_t x'$ is open in X. On the other hand, since $X - K_x(t)$ is the union of the other open equivalence classes, it is open too and thus $K_x(t)$ is closed. Define the set

$$K_x = \bigcap_{t>0} K_x(t)\,.$$

Now consider the connected component of x, C_x, and define the set CO_x as the intersection of all open and closed sets that contain x. Then one has

$$C_x \subseteq CO_x \subseteq K_x\,,$$

and obviously, if K_x is connected, then all inclusions are equalities. Since K_x is intersection of closed sets, it is closed, and if it is disconnected, then one can decompose it as a disjoint union $K_x = A \cup B$ such that A and B are closed and nonempty. Hence there is an $s > 0$ such that the open neighborhoods of A and B, $U_{2s}(A) = \cup_{x \in A} D^\circ_{2s}(x)$ and $U_{2s}(B) = \cup_{x \in B} D^\circ_{2s}(x)$ are disjoint, where $D^\circ_{2s}(x)$ denotes the open ball with center x and radius $2s$.

Define $H = X - U_s(A) \cup U_s(B)$ and assume that $x \in A$ and take $x' \in B$. For each t such that $0 < t < s$, $x \sim_t x'$, so that there is a t-chain $\{x_0, \ldots, x_k\}$ from x

to x'. According to the choice of s one can always assume that one of the links of the corresponding t-chain lies in H. In other words, if $t < s$, then $K_x(t) \cap H \neq \emptyset$. On the other hand, if $t < t$, then $K_x(t) \subseteq K_x(t')$. Hence by the compactness of the sets $K_x(t) \cap H$, $0 < t < s$, their intersection is nonempty, namely $K_x \cap H \neq \emptyset$, thus contradicting the definition of H. Therefore K_x is connected.

Let now V be an open neighborhood of C. Since C is a connected component of X, then $C = C_x = K_x$ for every $x \in C$. If the set $K_x(t) \cap (X - V)$ is nonempty for every t, then by the compactness of $K_x(t) \cap (X - V)$, their intersection $K_x \cap (X - V) \neq \emptyset$, which is again a contradiction.

If X is not metric, the proof is similar, but the t-neighborhoods must be replaced by something more general, by defining an adequate uniform structure. This can be found in [5, Ch. II.§4]. □

1.4 Group actions

An interesting aspect of topology is the one that links it with the algebra. This relationship appears in different ways. An important way refers to what we may call the symmetry of a topological space, that brings us directly to the origins of group theory itself, because initially the groups were precisely conceived as symmetry groups.

1.4.1 Definition. A topological space G endowed with two continuous mappings

$$G \times G \longrightarrow G\,, \quad (g,h) \longmapsto gh \quad \text{and} \quad G \longrightarrow \mathcal{G}\,, \quad g \longmapsto g^{-1}\,,$$

which provide G with the structure of a group, is called a *topological group*.

In other words, a topological group is a set together with the structure of a topological space and the structure of a group, both of which are compatible.

1.4.2 Examples.

(a) If G is any group and it is furnished with the discrete topology, then it is a topological group simply called a *discrete group*.

(b) If $G = \mathbb{R}^n$ is considered as a topological space as usual and as a group with the mappings $(x, y) \mapsto x + y$ and $x \mapsto -x$, then it is a topological group. In particular, $\mathbb{R} = \mathbb{R}^1$ and $\mathbb{C} = \mathbb{R}^2$ are topological groups in this sense.

(c) If $G = \mathbb{C} - \{0\}$ is considered as a topological subspace of $\mathbb{C} = \mathbb{R}^2$ and has the group structure given by $(w, z) \mapsto wz$ and $z \mapsto z^{-1}$ (complex multiplication and complex inverse), then it is a topological group.

(d) If $\mathbb{S}^1 \subset \mathbb{C}-\{0\}$ is considered as a topological subspace and also as a subgroup, then it is a topological group.

(e) Take $G = \mathrm{GL}_n(\mathbb{R})$ (resp. $G = \mathrm{GL}_n(\mathbb{C})$) to be the set of invertible $n \times n$ matrices with real (resp. complex) entries. G can be seen as a topological subspace of $\mathbb{R}^{n^2}$ (resp. $\mathbb{C}^{n^2}$) by putting each row of the matrix simply one after the other in one line. It can also be considered as a group with the usual matrix multiplication. Since these group operations are clearly continuous, G is a topological group. See next chapter.

1.4.3 EXERCISE. Show that G is a topological group if and only if the mapping $G \times G \longrightarrow G$ given by $(g, h) \mapsto gh^{-1}$ is continuous.

1.4.4 DEFINITION. Let G be a topological group. We say that G *acts* on a topological space X if there is a continuous map

$$G \times X \longrightarrow X\,, \quad (g, x) \longmapsto gx\,,$$

called the *action*, so that the following identities hold:

$$1x = x$$
$$g(hx) = (gh)x\,,$$

where in the first identity 1 denotes the neutral element of the group G, while in the second, the action of the group is applied twice on the left hand side, while on the right hand side it is applied only once after multiplying inside the group.

1.4.5 EXERCISE. Prove that given $g \in G$, the map $t_g : X \longrightarrow X$ such that $t_g(x) = gx$ is a homeomorphism. This map is called *translation* in X by the element g.

As the previous exercise shows, the action of the group corresponds to a set of homeomorphisms of the space onto itself, with the group structure given by composition. This suggests the symmetry of the space.

1.4.6 EXAMPLES.

(a) Let $G = \mathbb{Z}_2$ be the discrete group with two elements $\{1, -1\}$ and take $X = \mathbb{S}^n$. Then one has a group action

$$\mathbb{Z}_2 \times \mathbb{S}^n \longrightarrow \mathbb{S}^n$$

such that $(-1)x = -x$. This action is called *antipodal action* on the n-sphere.

(b) Let $G = \mathbb{S}^1$ be the topological group of unit complex numbers, and take $X = \mathbb{S}^{2n+1} \subset \mathbb{C}^{n+1} = \mathbb{R}^{2n+2}$. Then one has a group action

$$\mathbb{S}^1 \times \mathbb{S}^{2n+1} \longrightarrow \mathbb{S}^{2n+1}\,, \quad (\zeta, z) \longmapsto \zeta z\,,$$

given by componentwise complex multiplication.

(c) Let $G = \mathbb{S}^1$ be again the topological group of unit complex numbers, and take $X = \mathbb{S}^1 \times \mathbb{S}^1$ be the 2-torus. Then one has a group action

$$\mathbb{S}^1 \times \times(\mathbb{S}^1 \times \mathbb{S}^1) \longrightarrow \mathbb{S}^1 \times \mathbb{S}^1\,, \quad (\zeta, (w, z)) \longmapsto (\zeta w, \zeta z)\,,$$

given by componentwise complex multiplication.

(d) Let $G = \mathrm{GL}_n(\mathbb{R})$ be the topological group of invertible real $n \times n$ matrices, and take $X = \mathbb{R}^n$. Then one has a group action

$$\mathrm{GL}_n(\mathbb{R}) \times \mathbb{R}^n \longrightarrow \mathbb{R}^n\,, \quad (M, A) \longmapsto MA\,,$$

where M is a matrix and A is an n-vector written vertically.

Given a group action of a topological group G on a space X. We define an equivalence relation on X be declaring $x \sim y$ if and only if there is $g \in G$ such that $y = gx$. The equivalence classes are the so-called *orbits* and for each $x \in X$ is given by the set $Gx = \{gx \mid g \in G\}$.

1.4.7 Definition. The quotient space of a G-space X under the equivalence relation given above is called the *quotient of X under the action of G*, or more simply, the *orbit space* of the G-space X. This space is usually denoted by X/G.

1.4.8 Examples.

(a) Consider the antipodal action on $\mathbb{S}^n$ given in 1.4.6 (a). Then the orbits in $\mathbb{S}^n$ are the pairs of points $\{x, -x\}$ and the quotient of $\mathbb{S}^n$ under the antipodal action $\mathbb{S}^n/\mathbb{Z}_2$ is the *real projective space* $\mathbb{RP}^n$.

(b) Consider the $\mathbb{S}^1$-action on $\mathbb{S}^{2n+1}$ given in 1.4.6 (b). Then the orbits in $\mathbb{S}^{2n+1}$ are circles and the quotient of $\mathbb{S}^{2n+1}$ under this action $\mathbb{S}^{2n+1}/\mathbb{S}^1$ is the *complex projective space* $\mathbb{CP}^n$.

(c) The group $\mathbb{Z}$ (with $+$ as group multiplication) acts on $\mathbb{R}$ so that if $n \in \mathbb{Z}$ and $x \in \mathbb{R}$, $\alpha(n, x) = x + n$. One can prove that the orbit space $\mathbb{R}/\mathbb{Z}$ is homeomorphic to $\mathbb{S}^1$. Similarly, $\mathbb{Z}\times\mathbb{Z}$ acts on $\mathbb{R}^2 = \mathbb{R}\times\mathbb{R}$ by $\alpha((n_1, n_2), (x_1, x_2)) = (x_1 + n_1, x_2 + n_2)$. In this case, the orbit space $\mathbb{R} \times \mathbb{R}/\mathbb{Z} \times \mathbb{Z}$ is the torus $\mathbb{S}^1 \times \mathbb{S}^1$.

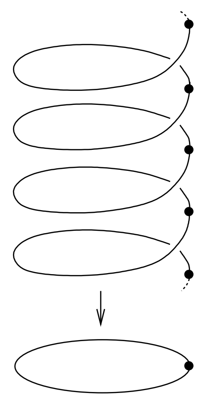


Figure 1.15 The identification $\mathbb{R} \twoheadrightarrow \mathbb{R}/\mathbb{Z}$

1.4.9 EXERCISE. Show that $\mathbb{R}/\mathbb{Z} \approx \mathbb{S}^1$ (*Hint:* The map $\mathbb{R} \longrightarrow \mathbb{S}^1$ such that $t \mapsto \mathrm{e}^{2\pi i t}$ determines the homeomorphism.)

1.4.10 EXERCISE. Use Exercise 1.4.9 to conclude that the orbit space $\mathbb{R}\times\mathbb{R}/\mathbb{Z}\times\mathbb{Z}$ is the torus $\mathbb{S}^1 \times \mathbb{S}^1$. (*Hint:* Recall that if one has an identification $q : X \longrightarrow X'$ and a locally compact space Y, then $q \times \mathrm{id}_Y : X \times Y \longrightarrow X' \times Y$ is also an identification.)

1.4.11 EXERCISE. Show that there is an action of $\mathbb{Z}$ on $\mathbb{R}^2$ such that

$$(n, (x_1, x_2)) \mapsto (n + x_1, (-1)^n x_2)\,.$$

Show furthermore that the orbit space of this action, $\mathbb{R}^2/\mathbb{Z}$, is homeomorphic to the Moebius band.

1.4.12 DEFINITION. An action of a topological group G on a space X is called *transitive* if given any two points $x, y \in X$, there is an element $g \in G$ such that $gx = y$.

1.4.13 EXAMPLE. Let G be a locally compact topological group and let $H \subseteq G$ be a (closed) subgroup. If the set $X = G/H$ of left cosets of H in G has the quotient topology, then there is an action $G \times X \longrightarrow X$ given by $g[g'] = [gg']$, where $[g]$ denotes the coset $gH \subset G$. Clearly this is a transitive action. The space $X = G/H$ is called *homogeneous space.*

1.4.14 **Proposition**. *Let G be a topological group, which as a topological space is a compact Hausdorff space. If $H \subset G$ is a closed subgroup, then the homogeneous space $X = G/H$ is a Hausdorff space too.*

Proof: Recall that X is a Hausdorff space if and only if the diagonal $\Delta_X \subset X \times X$ is a closed subspace. Since G is locally compact and $H \subset G$ is closed, one easily checks that $X = G/H$ is locally compact. Hence, if $q : G \longrightarrow G/H = X$ is the quotient map, then $q \times q : G \times G \longrightarrow G/H \times G/H = X \times X$ is a quotient map too. The inverse image $(q \times q)^{-1}\Delta_X = \bigcup_{g \in G} gH \times gH$.

There is a continuous map $G \times H \longrightarrow G \times G$ given by $(g, h) \mapsto (gh, gh)$, whose image is $\bigcup_{g \in G} gH \times gH \subset G \times G$. Since $G \times H$ is a compact space and $G \times G$ is a Hausdorff space, $(q \times q)^{-1}\Delta_X = \bigcup_{g \in G} gH \times gH \subset G \times G$ is a closed subspace. Hence $\Delta_X \subset X \times X$ is closed, because $q \times q$ is an identification. $\square$

Chapter 2 Manifolds

That of a manifold is one of the central concepts in topology. Many of the relevant spaces in topology, as well as in other areas of mathematics are manifolds. For instance, manifolds play a role in algebraic and in differential geometry, also in analysis or in function theory. Among the most important examples of manifolds, we have the spheres $\mathbb{S}^n$, $n = 0, 1, 2, \ldots$, the real and complex projective spaces $\mathbb{RP}^n$ and $\mathbb{CP}^n$, and the real and complex Grassmann and Stiefel manifolds $G_k(\mathbb{R}^n)$, $V_k(\mathbb{R}^n)$, and $G_k(\mathbb{C}^n)$, $V_k(\mathbb{C}^n)$. In dimension two, one has the important example of the surfaces, in particular, the Riemann surfaces. In dimension one there are essentially only two mannifolds, namely the line $\mathbb{R}$ and the circle $\mathbb{S}^1$. This last is a very important space in many aspects. It will be carefully analyzed from the point of view of homotopy theory, but it will also be used to develop knot theory at the end of the book.

In this chapter we shall study topological manifolds, and we shall carefully discuss the construction of all closed surfaces and discuss their classification. We shall also see many other examples of manifolds.

2.1 Topological manifolds

In this section we shall study the general concept of topological manifold and we shall indicate the meaning of manifold with structure, e.g. differentiable, smooth, complex, holomorphic, et cetera.

2.1.1 Definition. A Hausdorff second-countable space X is a *topological manifold* or simply, a *manifold* of dimension n, also called an *n-manifold* if each point $x \in X$ has a neighborhood V which is homeomorphic to an open set U in the closed unit ball $\mathbb{B}^n$. We say that a point $x \in X$ is an *interior point*, resp. a *boundary point*, if for some homeomorphism $\varphi : V \longrightarrow U$, $\varphi(x) \in \overset{\circ}{\mathbb{B}}{}^n$, resp. $\varphi(x) \in \mathbb{S}^{n-1}$. We define the *interior* of X by $X^\circ = \{x \in X \mid x \text{ is an interior point of } X\}$ and the *boundary* of X by $\partial X = \{x \in X \mid x \text{ is a boundary point of } X\}$. The domain invariance theorem 1.1.8 and the boundary invariance theorem 1.1.10 guarantee that these concepts are well defined.

2.1.2 REMARK. In the definition of a manifold X it is enough to ask that each point $x \in X$ has a neighborhood V which is homeomorphic to $\mathbb{B}^n$ since, if $U \subset \mathbb{B}^n$ is open and $\varphi : V \longrightarrow U$ is a homeomorphism, then $\varphi(x) \in U$ has a neighborhood $U' \subset U$ (not necessarily open) which is homeomorphic to $\mathbb{B}^n$. Therefore $\varphi^{-1}(U')$ is a neighborhood of x which is homeomorphic (through φ) to $\mathbb{B}^n$ (see Figure 2.1). Thus, a point $x \in X$ is an interior point if and only if $\varphi(x) \in \overset{\circ}{\mathbb{B}}{}^n$, and it is a boundary point if $\varphi(x) \in \partial\mathbb{B}^n = \mathbb{S}^{n-1}$.

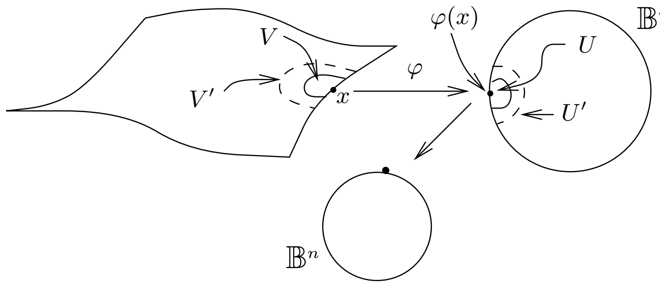


Figure 2.1 The unit n-ball models locally an n-manifold with boundary

2.1.3 **Proposition**. *The concept of boundary of a manifold is well defined, namely, it does not depend on the homeomorphism $\varphi : V \longrightarrow U$.*

Proof: By Remark 2.1.2 one may always assume that $U = \mathbb{B}^n$. Take a manifold X and a point $x \in X$. Let V be a neighborhood of x such that there are two homeomorphisms $\varphi, \psi : V \longrightarrow \mathbb{B}^n$. Then we have a homeomorphism $\varphi \circ \psi^{-1} : \mathbb{B}^n \longrightarrow \mathbb{B}^n$. If $\psi(x)$ lies on the boundary, then by the boundary invariance theorem 1.1.10, one has that $\varphi(x) = \varphi\psi^{-1}(\psi(x))$ lies also on the boundary. In other words, no matter what homeomorphism φ one takes, if a point x on the manifold is mapped to a boundary point of $\mathbb{B}^n$ by one of them, then it will be mapped to a boundary point by the other too. $\square$

2.1.4 NOTE. As it is the case with the boundary of a ball, the use of the expression "interior" of a manifold is once again an abuse of language. However, as in the case of the boundary, it is a generalization. Thus, if we want to be more precise, we should call the interior or the boundary of a subset A of a topological space X, interior or boundary of A *in* X, in order to distinguish them from the "intrinsic interior" or the "intrinsic boundary" of a manifold.

2.1.5 REMARK. In the case of a manifold without boundary X, namely, such that $\partial X = \emptyset$, we may redefine the concept by requiring that each point $x \in X$ has a neighborhood U which is homeomorphic to $\mathbb{R}^n$, since in this case, by definition, x would have a neighborhood which is homeomorphic to an open ball $\mathring{\mathbb{B}}^n$, which is homeomorphic to $\mathbb{R}^n$ (see the next exercise).

2.1.6 Exercise. Let $\mathring{\mathbb{B}}^n$ be the open unit ball (unit cell) in $\mathbb{R}^n$. Show that $\mathring{\mathbb{B}}^n$ is homeomorphic to $\mathbb{R}^n$. (*Hint:* The mapping $x \mapsto \frac{x}{1+x}$ determines a homeomorphism $\mathbb{R}^n \longrightarrow \mathring{\mathbb{B}}^n$.) Conclude that any open ball in $\mathbb{R}^n$, $B_r(y)$ with center y and radius r is homeomorphic to $\mathbb{R}^n$. (*Hint:* The mapping $x \mapsto \frac{x-y}{r}$ determines a homeomorphism $B_r(y) \longrightarrow \mathring{\mathbb{B}}^n$.)

From the exercise above, we obtain as a consequence the following result (cf. 1.1.4 and 1.1.6 (b)).

2.1.7 **Proposition**. *The finite product of cells is a cell and the finite product of balls is a ball. In particular, the product of an m-cell with a n-cell is an $(m+n)$-cell and the product of an m-ball and an n-ball is an $(m+n)$-ball.*

Proof: It is enough to show that one has homeomorphisms as depicted in the diagram

$$\begin{tikzcd} \mathring{\mathbb{B}}^m \times \mathring{\mathbb{B}}^n \arrow[r, "\approx"] \arrow[d, hook] & \mathring{\mathbb{B}}^{m+n} \arrow[d, hook] \\ \mathbb{B}^m \times \mathbb{B}^n \arrow[r, "\approx"] & \mathbb{B}^{m+n}\,. \end{tikzcd}$$

Since by 2.1.6, there are homeomorphisms $\mathring{\mathbb{B}}^k \xrightarrow{\approx} \mathbb{R}^k$ and $\mathbb{R}^m \times \mathbb{R}^n \xrightarrow{\approx} \mathbb{R}^{m+n}$, we obtain the first part of the statement.

From 1.1.6 (b) we obtain the second part. $\square$

2.1.8 **Corollary**. *The following are equal spaces: $\partial(\mathbb{B}^m \times \mathbb{B}^n)$ and $(\partial\mathbb{B}^m \times \mathbb{B}^n) \cup (\mathbb{B}^m \times \partial\mathbb{B}^n)$.*

Proof: One has equalities $\partial(\mathbb{B}^m \times \mathbb{B}^n) = \mathbb{B}^m \times \mathbb{B}^n - \mathring{\mathbb{B}}^m \times \mathring{\mathbb{B}}^n = ((\mathbb{B}^m - \mathring{\mathbb{B}}^m) \times \mathbb{B}^n) \cup (\mathbb{B}^m \times (\mathbb{B}^n - \mathring{\mathbb{B}}^n)) = (\partial\mathbb{B}^m \times \mathbb{B}^n) \cup (\mathbb{B}^m \times \partial\mathbb{B}^n)$. $\square$

The conditions that a manifold must be a Hausdorff and second-countable space are not consequences of the condition that the manifold is locally homeomorphic to $\mathbb{R}^n$. There are "pathological" spaces which are locally Euclidean, but are not Hausdorff or second-countable.

2.1.9 EXAMPLE. Consider the space X obtained after identifying in the topological sum of two copies of $\mathbb{R}$, the positive semilines. We do it as follows. Take two homeomorphic copies of $\mathbb{R}$, which will be denoted by $\mathbb{R}_1$ and $\mathbb{R}_2$, and take the quotient space

$$X = \mathbb{R}_1 \sqcup \mathbb{R}_2/\!\sim \ .$$

The equivalence relation is defined as follows. Take homeomorphisms $\varphi_\nu : \mathbb{R}_\nu \approx \mathbb{R}$, $\nu = 1, 2$, and declare $x_1 \sim x_2$ if and only if $\varphi_1(x_1) = \varphi_2(x_2) > 0$.

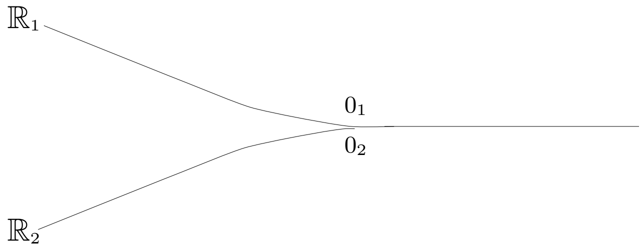


Figure 2.2 A space which is locally homeomorphic to $\mathbb{R}$ and is not Hausdorff

Obviously this quotient space, which is shown in Figure 2.2 (though not with high fidelity), is such that each of its points has a neighborhood which homeomorphic to an open interval in $\mathbb{R}^n$. However the points 0_1 and 0_2 that are such that $\varphi_\nu(0_\nu) = 0$, $\nu = 1, 2$, are different, but any neighborhood in X of one of them overlaps (on the positive side) with any other neighborhood of the other. Thus X is not Hausdorff.

2.1.10 EXAMPLE. The *closed long ray* L is defined as the cartesian product of the first uncountable ordinal Ω with the half-open interval $[0, 1)$, equipped with the order topology that arises from the lexicographical order on $\Omega \times [0, 1)$. The *open long ray* is obtained from the closed long ray by removing the smallest element $(0, 0)$.

The *long line* is obtained by putting together a long ray in each direction. In other words

$$L = \coprod_\alpha I_\alpha/\!\sim ,$$

where $I_\alpha \approx I$ by a homeomorphism ψ_α and α takes values in all the ordinals less than Ω and if $\psi_\alpha(0_\alpha) = 0$, $\psi_\alpha(1_\alpha) = 1$, then $1_\alpha \sim 0_{\alpha+1}$. Obviously each point in L has a neighborhood which is homeomorphic to $\mathbb{R}$. However L is not second-countable (*exercise*).

As in the case of 1.1.10 one may prove the following result (*exercise*).

2.1.11 **Theorem**. *Let $\varphi : X \longrightarrow Y$ be a homeomorphism of manifolds with boundary. Then φ determines a homeomorphism $\varphi|_{\partial X} : \partial X \longrightarrow \partial Y$ of the boundaries.* □

From 2.1.7 one obtains the following important result.

2.1.12 **Theorem**. *Let X be an m-manifold and Y an n-manifold with boundaries ∂X and ∂Y. Then $X \times Y$ is an $(m+n)$-manifold with boundary $\partial(X \times Y) = X \times \partial Y \cup \partial X \times Y$.* □

Given a manifold X and a point $x \in X$, there is a neighborhood V of x in X which is homeomorphic, via some homeomorphism $\varphi : V \longrightarrow U$, to an open set $U \subset \mathbb{B}^n$. If $x \in \partial X$, then $V \cap \partial X$ is a neighborhood of x in ∂X and the restriction $\varphi|_{V \cap \partial X}$ is a homeomorphism onto an open set U' in s^{n-1}. Without loss of generality we may assume that U' is not the whole sphere. Hence, composing $\varphi|_{V \cap \partial X}$ with a convenient stereographic projection, we have that $V \cap \partial X$ is homeomorphic to an open set in $\mathbb{R}^{n-1}$. Thus we have shown the following result.

2.1.13 **Theorem**. *Let X be an n-manifold. Then its boundary ∂X is an $(n-1)$-manifold with empty boundary.* □

In what follows, we shall only consider the case of manifolds without boundary, and we leave the study of the general case as an *exercise* to the reader.

Let X be an n-manifold. Thus X can be covered by a family of open sets V_λ such that for each λ, there is a homeomorphism $\varphi_\lambda : V_\lambda \longrightarrow U_\lambda$, where $U_\lambda \subset \mathbb{R}^n$ is an open set. Each pair $(V_\lambda, \varphi_\lambda)$ will be called a *chart* of the manifold X. The homeomorphism φ_λ allows to put coordinates on V_λ. Namely, if $p_k : \mathbb{R}^n \longrightarrow \mathbb{R}$ is the projection onto the kth coordinate and $\varphi_\lambda^k = p_k \circ \varphi_\lambda$, then for each point $x \in V_\lambda$ the n-tuple of real numbers $(\varphi_\lambda^1(x), \ldots, \varphi_\lambda^n(x))$ are coordinates for x. These are known as *local coordinates* of x with respect the chart $(V_\lambda, \varphi_\lambda)$.

A family $\mathcal{A}$ of charts $\{(V_\lambda, \varphi_\lambda)\}$ such that the family of open sets $\{V_\lambda\}$ is a cover of X, is called an *atlas* for the manifold X.

Let $\mathcal{A}$ be an atlas for the manifold X and let $(V_\lambda, \varphi_\lambda)$ and (V_μ, φ_μ) be two charts on $\mathcal{A}$. The homeomorphism

$$\gamma_\mu^\lambda = \varphi_\mu \circ \varphi_\lambda^{-1} : \varphi_\lambda(V_\mu \cap V_\lambda) \longrightarrow \varphi_\mu(V_\mu \cap V_\lambda)$$

is known as *change of coordinates.* or *transition map* These changes of coordinates are homeomorphisms between open sets of $\mathbb{R}^n$. Therefore they may satisfy additional conditions (see Figure 2.3).

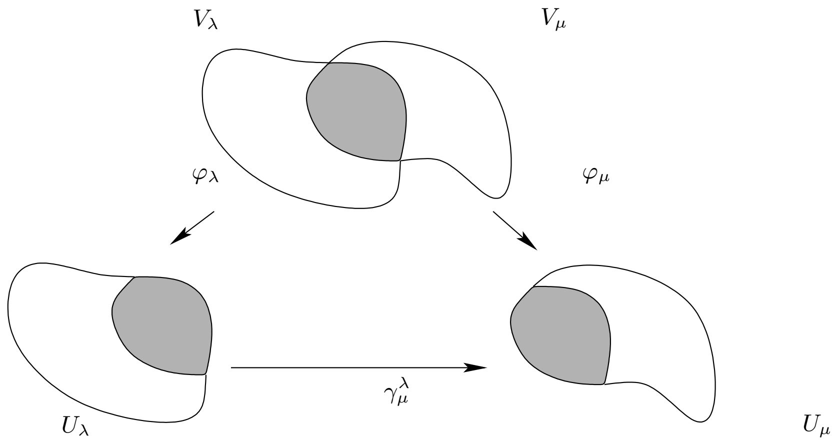


Figure 2.3 Change of coordinates

2.1.14 Definition. Let $\mathcal{A}$ be an atlas for a manifold X. If the homeomorphisms γ_μ^λ are differentiable of class $\mathcal{C}^r$ (resp. $\mathcal{C}^\infty$, analytic, holomorphic, et cetera) we say that the atlas $\mathcal{A}$ is a $\mathcal{C}^r$ (resp. $\mathcal{C}^\infty$, analytic, holomorphic, et cetera) *structure* on X. We also say that the manifold is *of class* $\mathcal{C}^r$ (resp. *of class* $\mathcal{C}^\infty$, *analytic*, *holomorphic*, et cetera). A manifold of class $\mathcal{C}^\infty$ is also called a *smooth* manifold.

2.1.15 Note. In the case of a holomorphic manifold, for instance, we require that n is even, namely $n = 2m$, and we take a fixed homeomorphism $\mathbb{R}^n \approx \mathbb{C}^m$. This way, the change of coordinates are homeomorphisms between open sets in the complex space $\mathbb{C}^m$ and it makes sense to ask these homeomorphisms to be holomorphic functions. If this is the case, then one considers X to be a *holomorphic* m*-manifold.*

2.1.16 Exercise. Give the corresponding charts and atlases for the more general cases of manifolds with boundary. Explain how to extend the structure concept to this case.

In this text we shall not study manifolds with structure and we shall restrict ourselves to *topological manifolds*, namely to manifolds with an atlas whose transition maps are only (topological) homeomorphisms.

2.1.17 Examples.

(a) A 0-manifold is nothing else but a discrete space.

(b) $\mathbb{R}^n$ is an n-manifold. If we choose as an atlas for $\mathbb{R}^n$ the family of all open sets, each together with the identity homeomorphisms, then this atlas determines an analytic structure on $\mathbb{R}^n$ (thus also of class $\mathcal{C}^r$, $r \leq \infty$). More generally, any open set in $\mathbb{R}^n$ is a manifold with any of the structures.

(c) $\mathbb{C}^m$ is a $2m$-manifold. If as before, we choose as an atlas for $\mathbb{C}^n$ the family of of all open sets, each together with the identity homeomorphisms, then this atlas determines an holomorphic structure on $\mathbb{C}^n$. Then $\mathbb{C}^n$ is a holomorphic n-manifold.

(d) The sphere $\mathbb{S}^n$ is a manifold. We have an atlas consisting of two charts, namely $V_1 = \mathbb{S}^n - N$ and $V_2 = \mathbb{S}^n - S$, where $N = (0, 0, \ldots, 1)$ is the *north pole* and $S = (0, 0, \ldots, -1)$ is the *south pole*. The homeomorphism $\varphi_1 : V_1 \longrightarrow \mathbb{R}^n$ is the stereographic projection p defined above 1.1.0 and $\varphi_2 : V_2 \longrightarrow \mathbb{R}^n$ is the other stereographic projection 1.1.1. Notice that $\varphi_2 = \varphi_1 \circ a$, where $a : V_2 \longrightarrow V_1$ is the antipodal map given by $a(x) = -x$. With this atlas, $\mathbb{S}^n$ is a smooth manifold.

(e) The cross $X = \{(x, 0) \mid x \in \mathbb{R}\} \cup \{(0, y) \mid y \in \mathbb{R}\} \subset \mathbb{R}^2$ is not a manifold, since the origin does not have any neighborhood homeomorphic to $\mathbb{R}$. (Indeed any connected neighborhood is a cross, and when one deletes the origin, it decomposes into four components. On the other hand, no matter what point we delete from $\mathbb{R}$, we always obtain two components. Thus the cross is not homeomorphic to the line.)

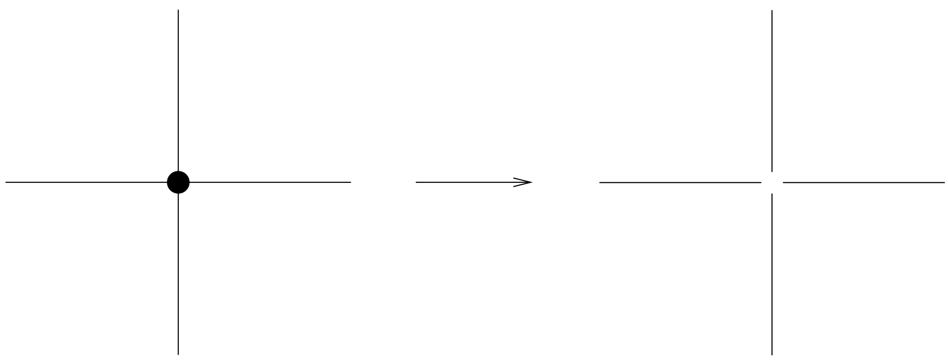


Figure 2.4 The cross is not homeomorphic to $\mathbb{R}$

(f) Take the upper halfspace $\mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_n \geq 0\}$. $\mathbb{R}^n_+$ is an n-manifold with boundary $\partial\mathbb{R}^n_+ = \mathbb{R}^{n-1}$.

(g) $\mathbb{B}^n$ is a manifold with boundary $\partial\mathbb{B}^n = \mathbb{S}^{n-1}$.

(h) If X is an n-manifold without boundary, then $X \times I$ is an $(n+1)$-manifold, whose boundary consists of two copies of X, namely $\partial(X \times I) = X \times \{0, 1\}$.

2.1.18 Exercise. Show that the interior of the manifold with boundary $\mathbb{R}^n_+$, i.e. $(\mathbb{R}^n_+)^\circ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_n > 0\}$, is homeomorphic to $\mathbb{R}^n$. (*Hint:* The mapping $(x_1, \ldots, x_n) \mapsto (x_1, x_2, \ldots, x_{n-1}, x_n - (1/x_n))$ determines a homeomorphism.)

2.1.19 EXERCISE. Take the point $N = (0, 0, \dots, 1) \in \mathbb{B}^n$. Show that one has a homeomorphism

$$\psi : \mathbb{B}^n - N \approx \mathbb{R}^n_+$$

such that $\psi|_{\mathbb{S}^{n-1}-N} : \mathbb{S}^{n-1} - N \longrightarrow \mathbb{R}^{n-1}$ is the stereographic projection. (*Hint:* See Figure 2.5.)

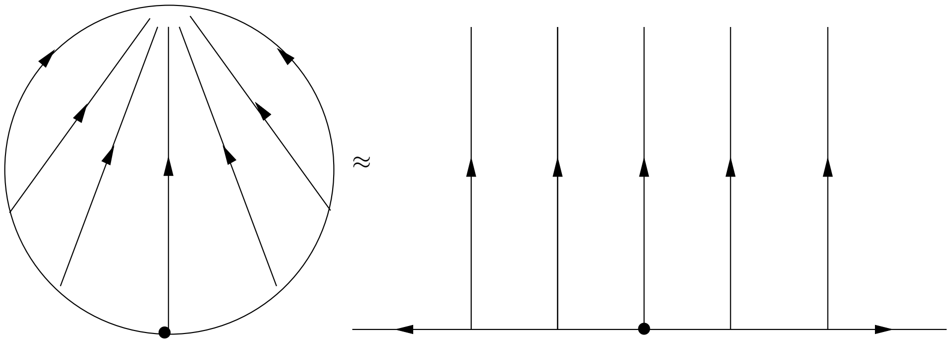


Figure 2.5 Homeomorphism of $\mathbb{B}^n - N$ with $\mathbb{R}^n_+$

2.1.20 DEFINITION. Let X be an n-manifold and take $Y \subset X$. We say that Y is a *submanifold* of dimension $m \leq n$ if the following hold:

(a) Y is an m-manifold.

(b) X admits an atlas $\mathcal{A}$ such that for each point $x \in Y$ and for each chart (V, φ) in $\mathcal{A}$ with $y \in V$, the local coordinates $\varphi^k|_{V\cap Y} : V \cap Y \longrightarrow \mathbb{R}$ are zero for $k = m + 1, \dots, n$.

(c) The pairs $(V \cap Y, p^m \circ \varphi|(V \cap Y))$, where $p^m : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is the projection onto the first m coordinates, form an atlas for Y.

Thus $\varphi(V \cap Y) = U \cap \mathbb{R}^m$, under the canonical inclusion of $\mathbb{R}^m$ into $\mathbb{R}^n$ in the first m coordinates.

2.1.21 EXAMPLES.

(a) $\mathbb{R}^m \subset \mathbb{R}^n$, together with the canonical inclusion, is a submanifold. More generally, if $V \subset \mathbb{R}^n$ is open, then $V \cap \mathbb{R}^m$ is a submanifold of V (and of $\mathbb{R}^n$).

(b) If V is open in $\mathbb{R}^n$, then V is a submanifold of $\mathbb{R}^n$. More generally, if $V \subset \mathbb{R}^n$ is open and $W \subset V$ is open too, then W is a submanifold of V.

(c) $\mathbb{S}^m \subset \mathbb{S}^n$, with the inclusion given by the canonical inclusion $\mathbb{R}^{m+1} \subset \mathbb{R}^{n+1}$, is a submanifold. Namely the charts $V_1 = \mathbb{S}^n - A_1$ and $V_2 = \mathbb{S}^n - A_2$, where $A_1 = (1, 0, 0, \dots, 0)$ and $A_2 = (-1, 0, 0, \dots, 0)$, with the corresponding stereographic projections onto $\mathbb{R}^n$, intersect $\mathbb{S}^m$ in charts W_1 and W_2 with stereographic projections onto $\mathbb{R}^m$.

(d) $\partial X \subset X$ is a submanifold of dimension $n-1$ if X is an n-manifold. Namely, if $x \in \partial X$, then a chart $\varphi : V \longrightarrow U \subset \mathbb{B}^n$ around x is such that $\varphi(x) \in \mathbb{S}^{n-1}$. Hence, $\varphi|_{V\cap\partial X} : V \cap \partial X \longrightarrow U \cap \mathbb{S}^{n-1}$ is a homeomorphism. Without loss of generality we may assume that $U \cap \mathbb{S}^{n-1}$ is an open proper subset of the sphere, thus by Exercise 2.1.19, this subset is homeomorphic to $\mathbb{R}^{n-1}$ by a homeomorphism as in the definition of a submanifold.

2.1.22 Definition. Let X and Y be topological manifolds and let $f : Y \longrightarrow X$ be continuous. We say that f is an *embedding of manifolds* if it is a topological embedding (i.e. a homeomorphism onto its image $f(Y)$) and $f(Y)$ is a submanifold of X.

2.1.23 Examples. The following are embeddings of manifolds.

(a) The canonical inclusion $\mathbb{R}^m \hookrightarrow \mathbb{R}^n$, $m \leq n$.

(b) The canonical inclusion $\mathbb{S}^m \hookrightarrow \mathbb{S}^n$, $m \leq n$.

(c) The canonical inclusion $\mathbb{S}^1 \hookrightarrow \mathbb{R}^2$.

(d) The mapping of the *torus* into $\mathbb{R}^3$, $f : \mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1 \hookrightarrow \mathbb{R}^3$, such that if $(x_1, x_2, y_1, y_2) \in \mathbb{S}^1 \times \mathbb{S}^1 \subset \mathbb{R}^2 \times \mathbb{R}^2$, then

$$f(x_1, x_2, y_1, y_2) = ((2 + y_1)x_1, (2 + y_1)x_2, y_2)\,.$$

2.1.24 Exercise. Consider the map $\varphi : I \times I \longrightarrow \mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ given by $\varphi(s,t) = (\mathrm{e}^{2\pi is}, \mathrm{e}^{2\pi it})$. Show that φ is surjective. If we consider the equivalence relation determined by $(s, 0) \sim (s, 1)$ and $(0, t) \sim (1, t)$, $s, t \in I$, show that φ is compatible with the identification. Furthermore, show that the induced map on the quotient $\psi : I \times I/{\sim} \longrightarrow \mathbb{T}^2$ is bijective. Since clearly $I \times I/{\sim}$ is compact and $\mathbb{T}^2 \subset \mathbb{C} \times \mathbb{C}$ is Hausdorff, show that ψ is a homeomorphism. This provides another construction of the torus. See Figure 2.15.

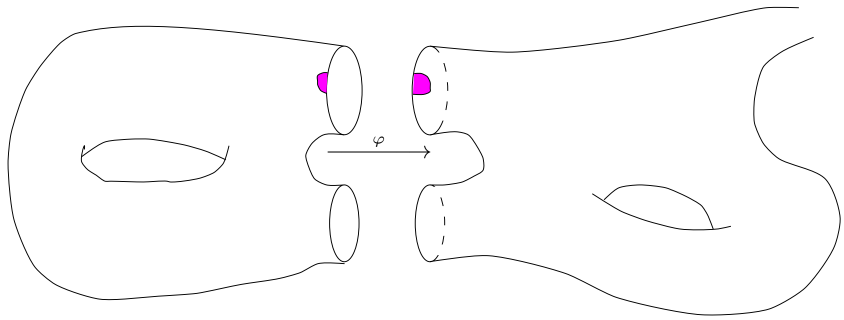


Figure 2.6 Union of two manifolds along their boundaries

2.1.25 DEFINITION. Let X and Y be two n-manifolds with boundary such that $\partial X \approx \partial Y$. If $\varphi : \partial X \longrightarrow \partial Y$ is a homeomorphism, we may construct $X \cup_\varphi Y$ by taking the disjoint union $X \sqcup Y$ and identifying each point $x \in \partial X$ with the point $\varphi(x) \in \partial Y$ (see Figure 2.6).

2.1.26 EXERCISE. Let $\varphi : \partial X \longrightarrow \partial Y$ be a homeomorphism. Show that $X \cup_\varphi Y$ is a manifold without boundary. (*Hint:* A point $x \in \partial X$ has a neighborhood, which is homeomorphic to $\mathbb{R}^n_+$ in such a way that x corresponds to the origin. An analogous situation happens with the point $\varphi(x) \in \partial Y$. It is now possible to take smaller neighborhoods which are homeomorphic to open semiballs, in such a way that after making the identification, both produce an open ball.)

2.1.27 DEFINITION. More generally, assume that the boundaries ∂X and ∂Y are not connected and A is a union of some of the connected components of ∂X and that B is also a union of some connected components of ∂Y. Assume also that there is a homeomorphism $\varphi : A \longrightarrow B$. As above, we can define $X \cup_\varphi Y$ as the result of taking the topological sum $X \sqcup Y$ and identifying $x \in A \subset \partial X$ with the point $\varphi(x) \in B \subset \partial Y$.

Analogously to 2.1.26, it is possible to solve the following.

2.1.28 EXERCISE. Let A and B be unions of connected components of ∂X and ∂Y, respectively, and let $\varphi : A \longrightarrow B$ be a homeomorphism. Show that $X \cup_\varphi Y$ is a manifold, whose boundary is the (disjoint) union of the connected components of ∂X and ∂Y that do not lie on A nor on B.

2.1.29 DEFINITION. A special case of the previous constructions is the *double* of a manifold (with boundary) X, which is defined by

$$2X = X \cup_{\mathrm{id}_{\partial X}} X\,,$$

namely as the result of identifying in the topological sum of two copies of X both boundaries through the identity. By 2.1.26 one knows that the one obtains a manifold without boundary. (If X has no boundary, then $2X = X \sqcup X$.)

2.1.30 EXERCISE. Show that the double of a manifold M is homeomorphic to the quotient space $M \times \{0,1\}/\sim$ where $(x,0) \sim (x,1)$ for all $x \in \partial M$.

2.1.31 EXERCISE.

(a) Show that the double of a (solid) ball is a sphere, namely, $2\mathbb{B}^n = \mathbb{S}^n$.

(b) Show that the double of $\mathbb{R}^n_+$ is $\mathbb{R}^n$.

(c) Show that the double of the cylinder $\mathbb{S}^1 \times I$ is the torus $\mathbb{S}^1 \times \mathbb{S}^1$.

Given two n-manifolds X and Y, their topological sum $X \sqcup Y$ is again an n-manifold, although it is disconnected. It is possible to make an operation between connected n-manifolds the produces another connected n-manifold.

2.1.32 DEFINITION. Let X and Y be connected n-manifolds and pierce a hole in each, namely, take n-balls B_X and B_Y embedded in X and Y, respectively, and consider $X' = X - B_X^\circ$ and $Y' = Y - B_Y^\circ$. Then X' and Y' are manifolds with boundaries $\partial X' = \partial B_X$ and $\partial Y' = \partial B_Y$. Since $\partial B_X \approx \mathbb{S}^{n-1} \approx \partial B_Y$, we may take a homeomorphism $\varphi : \partial X' = \partial B_X \longrightarrow \partial B_Y = \partial Y'$ and define

$$X \# Y = X' \cup_\varphi Y'\,.$$

This new n-manifold is called *connected sum* of X and Y. In the case $n = 2$, this construction is independent of the way one chooses the balls and the homeomorphism between their boundaries. In the general case, $n > 2$, the construction does depend on the choice of the homeomorphism.

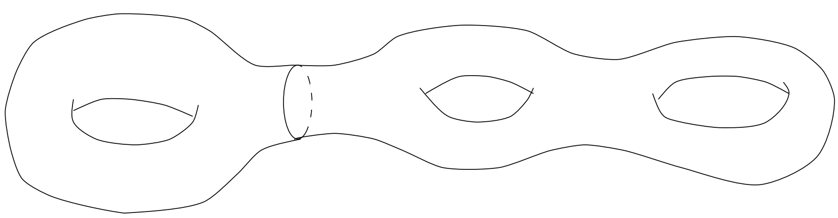


Figure 2.7 Connected sum of two manifolds

2.2 Surfaces

An important instance of manifolds are the surfaces. Their classification problem was already solved in the nineteenth century by Moebius (1861).

In this section we shall study the surfaces. We shall construct all the so-called closed surfaces and we shall state their classification theorem. In this section we agree to call a 2-ball simply a *disk*.

2.2.1 Definition. A 2-manifold S is a *surface*. If S is compact, connected and it has no boundary, then we say that the surface S is *closed*. (As it is the case with the interior and the boundary of a manifold, one should not confuse the concept of a *closed* surface with that of a *closed* set *in* a space.)

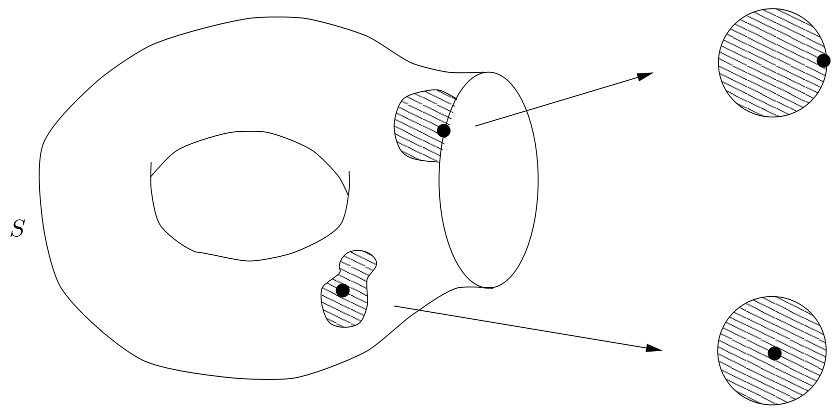


Figure 2.8 Surface with boundary

2.2.2 Examples. Besides $\mathbb{R}^2$, $\mathbb{R}^2_+$, or open subsets of them, there are the following examples.

(a) By 2.1.12, the torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ is a surface without boundary. Indeed, it is a closed surface.

(b) A closed disk D, which is homeomorphic to the unit 2-disk $\mathbb{D}^2$ (which is the same as the unit 2-ball $\mathbb{B}^2$), is a surface with boundary.

(c) Subsets of the plane, as depicted in Figure 2.10, are surfaces with boundary.

(d) Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be a continuous function. Then its *graph*

$$G(f) = \{(x, y, z) \in \mathbb{R}^3 \mid z = f(x, y)\}$$

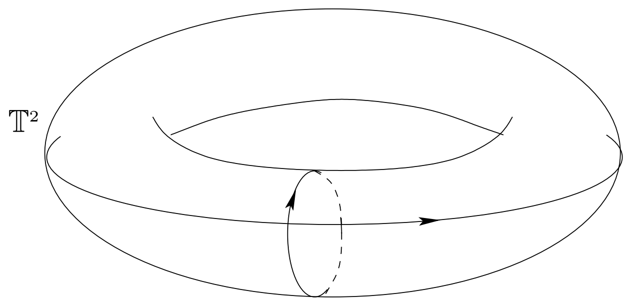


Figure 2.9 The torus is a surface without boundary

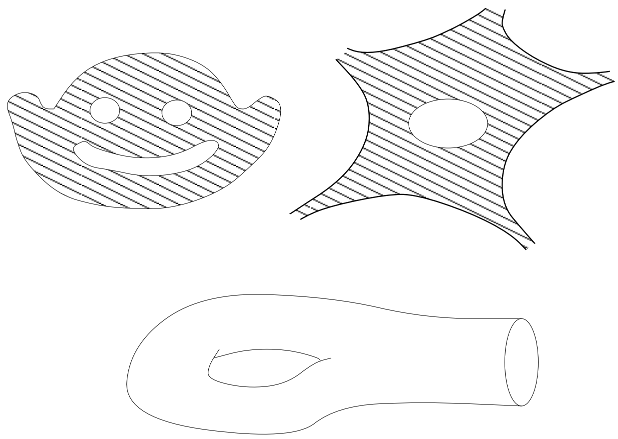


Figure 2.10 Surfaces with boundary

is a surface as one easily verifies. Thus a hyperboloid of one sheet, a hyperboloid of two sheets, a paraboloid, et cetera, are surfaces.

(e) The open cylinder $\mathbb{S}^1 \times \mathbb{R}$ and the pierced plane $\mathbb{R}^2 - S$, where S is some discrete set, are surfaces.

2.2.3 EXERCISE. Show that the surface $\mathbb{S}^1 \times \mathbb{R}$ is homeomorphic to $\mathbb{R}^2 - \{0\}$.

2.2.4 EXAMPLE. The *trivial strip* B is defined as the result of identifying in the square $I \times I$ a point of the form $(0, t)$ with the point $(1, t)$. Clearly B is homeomorphic to the closed cylinder $\mathbb{S}^1 \times I$. It is a manifold with boundary $\partial B \approx \mathbb{S}^1 \times \{0, 1\}$, namely the topological sum of two copies of $\mathbb{S}^1$. Therefore the boundary is disconnected: it has two connected components.

On the other hand, the *Moebius strip* M is defined as the result of identifying in $I \times I$ a point of the form $(0, t)$ with the point $(1, 1 - t)$. Clearly M is a surface with boundary $\partial M \approx \mathbb{S}^1$, namely one copy of $\mathbb{S}^1$, thus connected.

Hence $\partial B \not\approx \partial M$ and hence $M \not\approx B$.

The trivial strip B is an *orientable* surface, while the Moebius strip M is a *nonorientable* surface. Indeed, M is somehow the paradigm of the nonorientable surfaces. Its nonorientability can be explained by the following phenomenon, which does not take place in B. Take a coordinate system with the origin in some point on the *equator* $E = \{\overline{(s, \frac{1}{2})} \mid 0 \leq s \leq 1\} \subset M$, say $\overline{(\frac{1}{2}, \frac{1}{2})}$. One axis is horizontal, namely, it is parallel to E and points to the positive side, and the other axis is vertical, namely, it is parallel to $E' = \{\overline{(\frac{1}{2}, t)} \mid 0 \leq t \leq 1\}$ and points up.

Now we translate the system continuously along the equator around the Moebius strip. When we come back to the start point, the horizontal axis still points to the positive side, but the negative axis points down. In other words, if we look in the positive direction of the equator, at the start, the vertical axis points to the left, but when we come back to the start point, the vertical arrow points to the right: *it changed the orientation* of the coordinate system (see Figure 2.11).

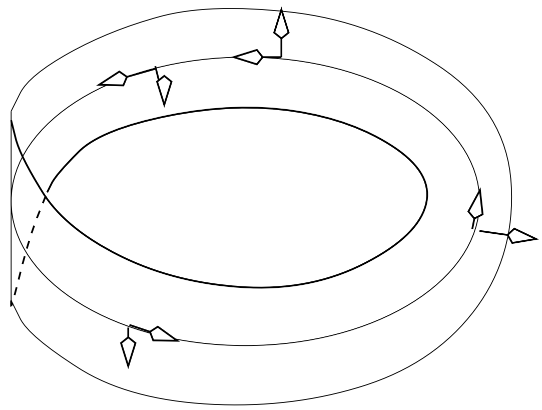


Figure 2.11 The Moebius strip is nonorientable

In general, and somehow informally, one says that a surface is *orientable* if for any *cycle*, namely for any subspace homeomorphic to the circle $\mathbb{S}^1$, if one translates a coordinate system along the cycle, the coordinate system comes back to the start point with the same orientation. The surface is *nonorientable* if it is not orientable, that is, if there is a cycle along which the coordinate system comes back with its orientation reversed. In the case of the Moebius strip, one of these cycles is E. It is not difficult to figure out that a surface S is nonorientable if and only if it admits an embedding of the Moebius strip, $M \hookrightarrow S$ (*exercise*).

2.2.5 NOTE. It is frequently comfortable to define the Moebius strip M as the quotient of $[-1,1] \times [-1,1]$ which results after identifying each point of the form $(1,t)$ with the point $(-1,-t)$.

2.2.6 EXAMPLE. The *Klein bottle*, is defined as the space K that results from the square $I \times I$ after identifying a each point $(0,t)$ with the point $(1,1-t)$ and each point $(s,0)$ with the point $(s,1)$. It is a compact surface without boundary, hence it is a closed surface. Furthermore, K is nonorientable, since a coordinate system which moves around the equator E, as defined for M in 2.2.4, comes back to the start point with the reversed orientation. Equivalently, if one takes the subspace M' of $I \times I$ consisting of the points (s,t) such that $\frac{1}{4} \leq t \leq \frac{3}{4}$, then the image of M' in K under the identification is homeomorphic to a Moebius strip M.

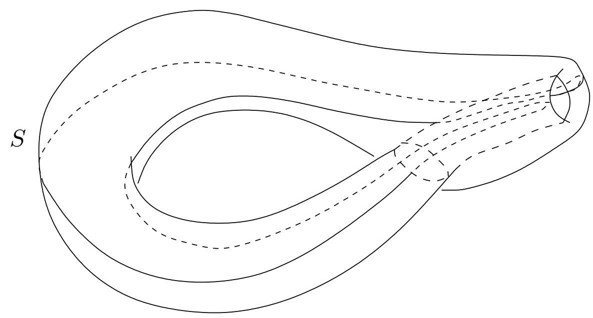


Figure 2.12 Klein bottle

2.2.7 EXERCISE. Show that the Klein bottle is homeomorphic to the double of the Moebius strip, namely $K \approx 2M$ (see 2.1.29).

2.2.8 EXAMPLE. Take a surface S and pierce in it two holes, namely, take two disjoint embeddings of the 2-ball into S, $e_0, e_1 : \mathbb{D}^2 \longrightarrow S$ and consider the surface $S' = S - e_1(\mathring{\mathbb{B}}^2) - e_2(\mathring{\mathbb{B}}^2)$. S' is a surface such that $\partial S' = \partial S \sqcup \mathbb{S}^1_0 \sqcup \mathbb{S}^1_1$, where $\mathbb{S}^1_0$ and $\mathbb{S}^1_1$ are circles in S. Let $\varphi_0 : \mathbb{S}^1 \times 0 \longrightarrow \mathbb{S}^1_0$ and $\varphi_1 : \mathbb{S}^1 \times 1 \longrightarrow \mathbb{S}^1_1$ be homeomorphisms and then glue the cylinder $\mathbb{S}^1 \times I$ onto S'. One says that the resulting surface S^+ is obtained by *gluing a handle* to S. Of course S^+ depends on how one glues the handle, i.e. from the homeomorphisms φ_0 and φ_1. In particular, as one can see in Figure 2.13, if $S = \mathbb{S}^2$ is the sphere, then one possibility is that S^+ is the torus, when S^+ is orientable (i.e. it does not admit an embedding of the Moebius strip). Otherwise, inverting the orientation of one of the homeomorphisms, what one obtains is the Klein bottle.

In what follows we shall describe several ways to construct surfaces.

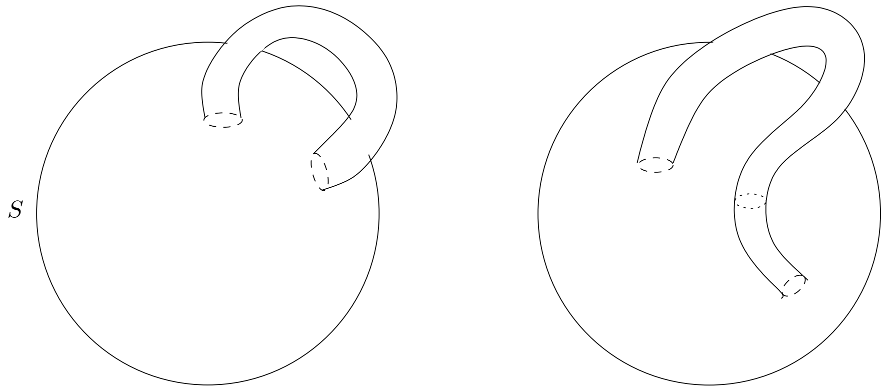


Figure 2.13 Spheres with handles

2.2.9 EXAMPLE. Let $g \geq 1$ be an integer.

(a) Let $\mathbb{D}_g \subset \mathbb{R}^2$ be the unit disk with g (disjoint) holes (see 2.2.8). If we take the manifold with boundary $S = \mathbb{D}_g \times I$, then its boundary is a surface A_g.

(b) Let $\mathbb{D}_g$ be as in (a). Its double $2\mathbb{D}_g$ is a surface B_g.

(c) Take the 2-torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$. The connected sum $C_g = \mathbb{T}^2 \# \mathbb{T}^2 \# \cdots \# \mathbb{T}^2$ of g copies of $\mathbb{T}^2$ is a surface.

(d) Gluing g handles to $\mathbb{S}^2$ in such a way that the resulting surface is orientable (i.e. it does not admit an embedding of the Moebius strip) we obtain a surface D_g.

All constructed surfaces in 2.2.9 are determined up to homeomorphism by the integer g. Moreover $A_g \approx B_g \approx C_g \approx D_g$. Figure 2.14 shows case $g = 3$.

2.2.10 EXERCISE. Show that there is a homeomorphism $A_g \approx B_g$. (*Hint:* $\partial S \cap (\mathbb{D}_g \times [0, 1/2]) \approx \mathbb{D}_g \approx \partial S \cap (\mathbb{D}_g \times [1/2, 1])$.)

There is another way to define these orientable surfaces without boundary. For $g \geq 1$ consider the regular polygon $E_{4g} \subset \mathbb{R}^2$ with $4g$ edges and whose vertices are the points $p_n = \mathrm{e}^{2\pi i n/4g}$, where $n = 1, 2, \ldots, 4g$. Since E_{4g} is a convex compact subset of $\mathbb{R}^2$ with nonempty interior, it is homeomorphic to the closed unit disk $\mathbb{D}^2$ and hence it is a surface with boundary.

We define an equivalence relation on the boundary of E_{4g} declaring equivalent the following points:

$$(1-t)p_{4i-3} + tp_{4i-2} \sim (1-t)p_{4i} + tp_{4i-1}$$
$$(1-t)p_{4i-2} + tp_{4i-1} \sim (1-t)p_{4i+1} + tp_{4i}$$

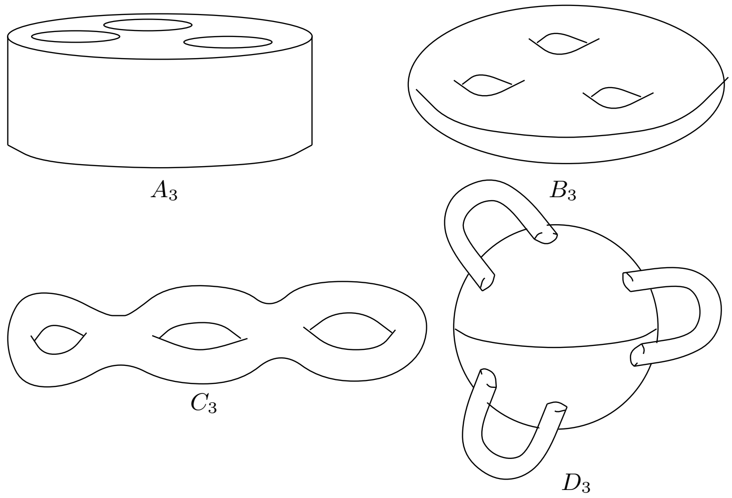


Figure 2.14 Surfaces of genus 3

Namely, the edges of the polygon are identified as shown in Figure 2.15.

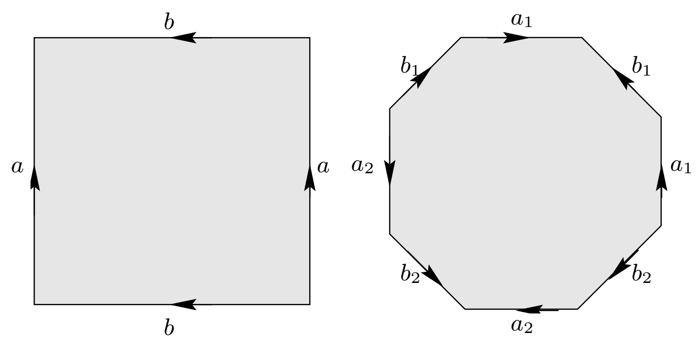


Figure 2.15 $4g$-gons for $g = 1$ and $g = 2$

2.2.11 **Proposition**. *The resulting space after identifying the edges of the polygon E_{4g} according to the relation defined above, is a closed orientable surface S_g.*

The surface S_g and any other surface S homeomorphic to S_g is called *orientable surface of genus g*. We denote by S_0 the orientable surface of genus 0, which is nothing else but the unit sphere de dimension 2.

Proof: Before we start with the proof, it is convenient to denote in E_{4g} every two edges which will be identified with each other, using the same letter. We also give an orientation to each edge, according to how they identify. Namely, the first edge a_1 has the positive (counterclockwise) orientation, while the third edge a_1 has the negative (clockwise) orientation. Furthermore, the second edge b_1 has the positive orientation, while the fourth edge b_1 has the negative orientation. The same happens for the next block of four edges a_2, b_2, a_2, b_2. Figure 2.15 shows this.

One must show that each point in S_g has a neighborhood which is homeomorphic to an open disk. There are three cases:

(i) If $x \in S_g$ is a point which comes from an interior point $y \in E_{4g}$, then it is possible to take a neighborhood V of y in E_{4g} which is an open disk and does not intersect the boundary. Then the image of V under the identification is again a neighborhood of x which is homeomorphic to an open disk.

(ii) If $x \in S_g$ comes from a point $y_1 \in \partial E_{4g}$ which is not a vertex, then it is possible to take a neighborhood V_1 of y_1 in E_{4g} which is the intersection with E_{4g} of a disk D_1 in $\mathbb{R}^2$ centered at y_1, and which does not contain any vertex. Then $y_1 \sim y_2$, where y_2 lies also on the boundary and is not a vertex. We can now take a neighborhood V_2 of y_2 in E_{4g} which, again, is the intersection with E_{4g} of another open disk D_2 in $\mathbb{R}^2$ centered at y_2 and has the same radius as D_1. Then, as one can see in Figure 2.16, the image of the union $V_1 \cup V_2$ is a neighborhood V of x which is homeomorphic to a disk.

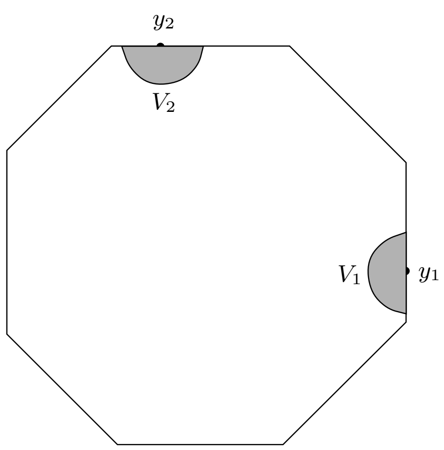


Figure 2.16 Two halfdisks yield a disk

(iii) Notice first that all vertces of E_{4g} identify to yield one point $x_0 \in S_g$. In order to construct a neighborhood of x_0 which is an open disk, for each vertex p_i we take a neighborhood V_i which is the intersection of some disk D_i in $\mathbb{R}^2$ centered at p_i with E_{4g} and small radius, which is equal for all vertices of the polygon. After the identification, the sides of the $4g$ circular sectors are glued in such a way that, as shown in Figure 2.17, the result is a neighborhood of x_0 homeomorphic to an open disk. $\square$

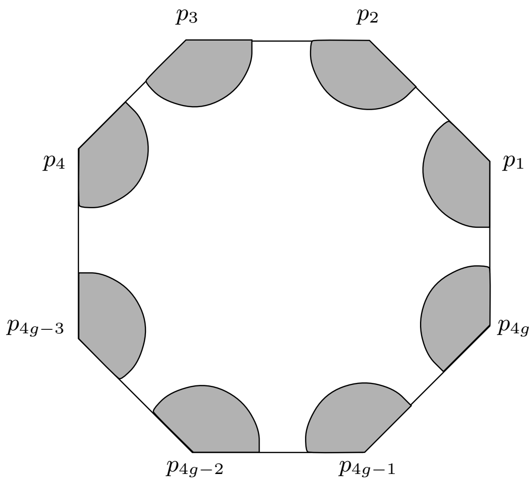


Figure 2.17 $4g$ sectors yield a disk

2.2.12 REMARK. Instead of the polygon E_{4g} one can take the unit disk $\mathbb{D}^2$ and mark the same vertices $p_n = \mathrm{e}^{2\pi \mathrm{i} n/4g}$, $n = 0, 1, \ldots, 4g-1$. Then one takes the arcs determined by two consecutive vertices p_i and p_{i+1} and makes the identifications accordingly. More precisely, if $0 \leq t \leq \frac{1}{4g}$, then

$$\mathrm{e}^{2\pi \mathrm{i} t/4g} \sim \mathrm{e}^{2\pi \mathrm{i}(3-t)/4g}$$

$$\mathrm{e}^{2\pi \mathrm{i}(1+t)/4g} \sim \mathrm{e}^{2\pi \mathrm{i}(4-t)/4g}$$

and in general, for $k = 0, \ldots, g-1$,

$$\mathrm{e}^{2\pi \mathrm{i}(4k+t)/4g} \sim \mathrm{e}^{2\pi \mathrm{i}(4k+3-t)/4g}$$

$$\mathrm{e}^{2\pi \mathrm{i}(4k+1+t)/4g} \sim \mathrm{e}^{2\pi \mathrm{i}(4k+4-t)/4g}\,.$$

One easily shows that the result of this identification is homeomorphic to S_g.

Let us now analyze the construction with which we obtain S_g. For the case $g = 1$, the polygon E_4 is a square in which opposite sides are identified as shown in Figure 2.15. Therefore what we obtain is a torus, namely, we have $S_1 \approx \mathbb{T}^2$ (see Exercise 2.1.24.

For $g > 1$ we proceed by induction on g. The surface S_g is a quotient E_{4g}. Take in E_{4g} the line segment joining p_1 and p_5. Then cut E_{4g} along this line segment and identify first p_1 and p_5 (recall that in all vertices S_g are identified to one point) in each of both pieces. So we obtain one copy of E_4 with a hole and a copy of E_{4g-4} with a hole as shown in Figure 2.18.

Hence we can obtain the identification $q : E_{4g} \longrightarrow S_g$ in three steps. First we identify in the pierced copy of E_4 the edges according to q and so we obtain a copy of a pierced S_1. Then we do the same with the pierced of E_{4g-4} and so we obtain

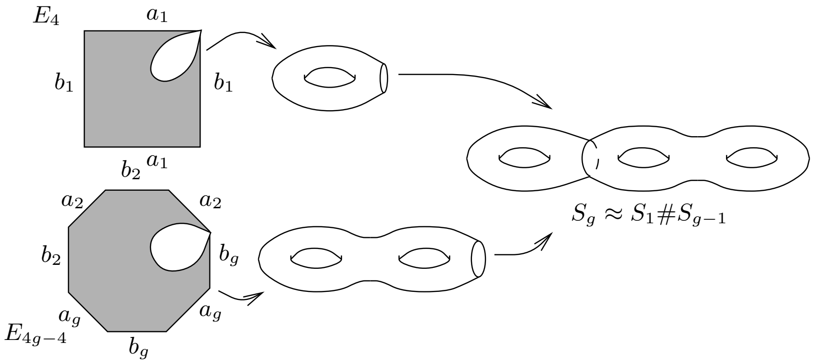


Figure 2.18 The connected sum of S_1 and S_g

a copy of a pierced S_{g-1}. The third step consists of identifying both pierced copies along the line segment, along which we cut E_{4g}, and so we obtain the connected sum of S_1 and S_{g-1}. In other words, we have established a recurrence formula

$$S_g \approx S_1 \# S_{g-1} .$$

Since S_1 is the torus, we may assume inductively that S_{g-1} is a connected sum of $g-1$ copies of the torus. Hence we have that S_g is the connected sum of g copies of the torus. Hence we have the following.

2.2.13 **Theorem**. *The closed oriented surface of genus g is the connected sum of g copies of the torus.* □

We shall now analyze the case of the closed nonorientable surfaces, namely of those surfaces which admit an embedding of the Moebius strip.

We start with the following example.

2.2.14 EXAMPLE. Two points x, $y \in \mathbb{R}^3 - \{0\}$ will be declared as equivalent if there exists a number $\lambda \in \mathbb{R}$ such that $y = \lambda x$. In other words, two (nonzero) points will be equivalent if and only if they lie on the same straight line through the origin. Hence the equivalence classes for this relation are in one to one correspondence with the straight lines in $\mathbb{R}^3$ which contain the origin. Let $\mathbb{P}^2$ be the quotient space of $\mathbb{R}^3 - \{0\}$ under this equivalence relation. This space is the *projective plane* or the *projective space* of dimension 2. We can say that $\mathbb{P}^2$ is the space of straight lines through the origin in $\mathbb{R}^3$. Take $x \in \mathbb{R}^3 - \{0\}$. If we denote by $\langle x \rangle$ the straight line in $\mathbb{R}^3$ through x, then a neighborhood of $\langle x \rangle$ always contains a double circular cone, which has the given line as axis. This shows clearly that $\mathbb{P}^2$ is a Hausdorff space. See Figure 2.19.

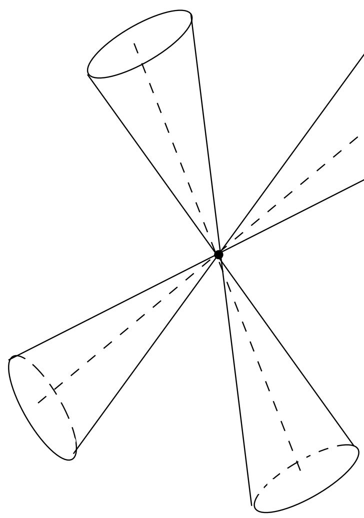


Figure 2.19 Two disjoint cones show that $\mathbb{P}^2$ is Hausdorff

The following result establishes properties of the projective plane and shows that it is the same space as one obtained from the 2-sphere by identifying antipodal points.

2.2.15 **Theorem**. *The projective plane $\mathbb{P}^2$ is homeomorphic to the quotient space of $\mathbb{S}^2$ obtain as*

$$\mathbb{RP}^2 = \mathbb{S}^2/\!\sim, \quad \textit{where} \quad x \sim y \quad \textit{if and only if} \quad x = y \quad \textit{or} \quad x = -y\,.$$

Proof: Let $\varphi : \mathbb{S}^2 \longrightarrow \mathbb{P}^2$ be given by $\varphi(x) = \langle x \rangle$. Then φ is a continuous, surjective map from a compact space to a Hausdorff space. A well-known result in general topology implies that φ is an identification. Furthermore $\varphi(x) = \varphi(y)$ if and only if $x \sim y$, since each straight line through the origin in $\mathbb{R}^3$ intersects the 2-sphere $\mathbb{S}^2$ in exactly two antipodal points. Therefore φ induces a homeomorphism $\overline{\varphi} : \mathbb{RP}^2 \xrightarrow{\approx} \mathbb{P}^2$. □

The next result gives another construction of the projective plane.

2.2.16 **Theorem**. *Let Q be obtained from the disk $\mathbb{D}^2$ after identifying two points of its boundary $\mathbb{S}^1$ if and only if they are antipodes, namely, if $x,\, y \in \mathbb{D}^2$, then $x \sim y$ if and only if $x = y$ or $x,\, y \in \mathbb{S}^1$ and $x = -y$. Then there is a homeomorphism $\overline{\psi} : Q \approx \mathbb{P}^2$.*

Proof: The map $\psi : \mathbb{D}^2 \longrightarrow \mathbb{P}^2$ given by $\psi(x) = \langle (x_1, x_2, \sqrt{1 - |x|^2}) \rangle$ if $x = (x_1, x_2)$, is continuous and surjective from a compact space to a Hausdorff space.

Hence, by the general topology result mentioned in the previous proof, ψ is an identification. Furthermore $\psi(x) = \psi(y)$ if and only if $x \sim y$. Thus ψ determines a homeomorphism $\overline{\psi} : Q \xrightarrow{\approx} \mathbb{P}^2$. □

Making use of 2.2.15 we shall prove that $\mathbb{P}^2$ is a surface, namely that it is a 2-manifold. We have the following result.

2.2.17 **Theorem**. *The projective plane $\mathbb{P}^2$ is homeomorphic to the space obtained by gluing a disk D to the Moebius strip M along their homeomorphic boundaries.*

Proof: Take the identification $\psi : \mathbb{D}^2 \longrightarrow \mathbb{P}^2$ constructed in the proof of 2.2.16, given by mapping antipodal points in the boundary of $\mathbb{D}^2$ to the same point in $\mathbb{P}^2$.

We can decompose $\mathbb{D}^2$ into two pieces as follows. Take

$$A = \{(x_1, x_2) \in \mathbb{D}^2 \mid |x_2| \leq 1/2\} \quad \text{and} \quad B = \{(x_1, x_2) \in \mathbb{D}^2 \mid |x_2| \geq 1/2\}\,.$$

If we restrict the identification ψ to each piece, we obtain from A a Moebius strip M and from B simply a disk D, since no two points of B are identified by ψ. Hence $\mathbb{P}^2$ is the result of gluing $M = \psi(A)$ with $D = \psi(B)$ along their boundaries, which are the images in M and D under the restrictions of ψ of the subset $\{(x_1, x_2) \in \mathbb{D}^2 \mid |x_2| = 1/2\}$ of A and of B. See Figure 2.20. □

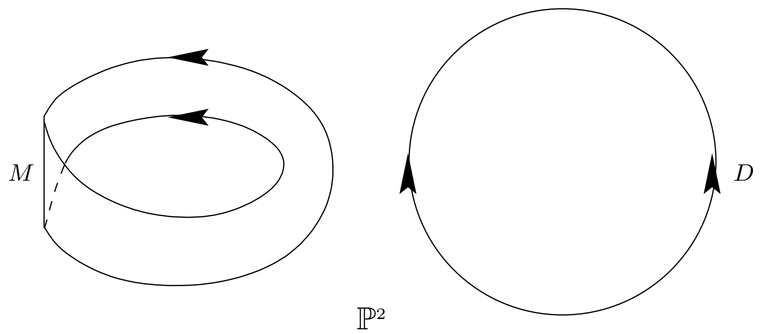


Figure 2.20 The identification of a Moebius strip with a disk along their boundaries yields a projective plane

Since both M and D are manifolds with homeomorphic boundaries, if we identify the boundaries via a homeomorphism, then we obtain $\mathbb{P}^2$. Hence by 2.1.26 we have the following.

2.2.18 **Corollary**. *$\mathbb{P}^2$ is a closed surface.* □

In the case of the orientable closed surfaces, we proved in 2.2.13 that the fundamental piece to construct them is the torus. This means that the closed orientable surface of genus g, S_g, is homeomorphic to the connected sum of g copies of the torus. Anther consequence of 2.2.17 is the next result, which is the first step towards showing that for the nonorientable case the fundamental piece to construct the nonorientable closed surfaces is the projective plane.

2.2.19 **Corollary**. *The connected sum of two copies of the projective plane $\mathbb{P}^2$ is the Klein bottle K.*

Proof: By 2.2.17, if we pierce a hole in $\mathbb{P}^2$ we obtain a Moebius strip. Therefore the connected sum of two projective planes is homeomorphic to the space obtained by gluing two Moebius strips along their boundaries. Namely, $\mathbb{P}^2$ is the double $2M$ of the Moebius strip. Consequently, by 2.2.7, $\mathbb{P}^2\#\mathbb{P}^2 \approx K$. □

2.2.20 EXERCISE. Decompose $I \times I$ into two pieces

$$A = \{(s,t) \mid 1/4 \le t \le 3/4\} \quad \text{and} \quad B = \{(s,t) \mid 0 \le t \le 1/4 \text{ or } 3/4 \le t \le 1\}$$

and show that by restricting the identifications of 2.2.6 to A and to B, respectively, one obtains two Moebius strips M_A and M_B. Conclude the Klein bottle is the union of M_A and M_B along their boundaries.

Analogously to 2.2.9, consider the next example.

2.2.21 EXAMPLE. Let $g \ge 1$ be an integer.

(a) The connected connected sum $C'_g = \mathbb{P}^2\#\mathbb{P}^2\#\cdots\#\mathbb{P}^2$ of g copies of the projective plane is a closed nonorientable surface.

(b) Piercing g holes in $\mathbb{S}^2$ and gluing a Moebius strip along the boundary of each hole through a homeomorphism of their boundaries, one obtains a closed nonorientable surface D'_g.

The surfaces described in 2.2.21 are determined up to homeomorphism by g and one has $C'_g \approx D'_g$. As in the orientable case, there is another way of constructing these nonorientable closed surfaces.

Let $E_{2g} \subset \mathbb{R}^2$, $g \ge 2$, be the regular polygon with $2g$ edges and vertices $p_n = e^{2\pi i n/2g}$, where $n = 1, 2, \ldots, 2g$. Since E_{2g} is a convex compact subset of $\mathbb{R}^2$ with nonempty interior, it is homeomorphic to $\mathbb{D}^2$ and hence it is a surface with boundary.

Define an equivalence relation on the boundary of E_{2g} by declaring as equivalent the following points:

$$(1-t)p_{2i-1} + tp_{2i} \sim (1-t)p_{2i} + tp_{2i+1}\,.$$

Namely, the edges of the polygon are identified as illustrated for $g = 2$ and for $g = 3$ in Figure 2.21.

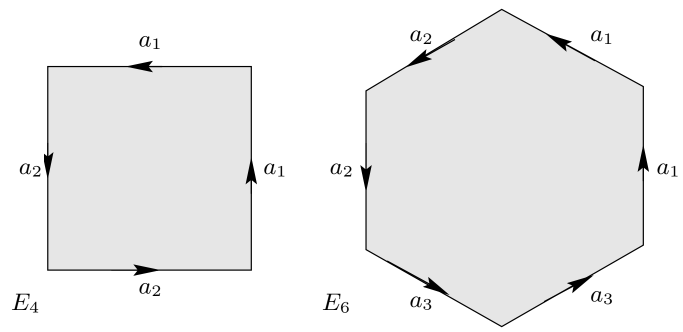


Figure 2.21 $2g$-gons, $g = 2$, $g = 3$

2.2.22 **Proposition**. *The space obtained by identifying the edges of the polygon E_{2g} according to the equivalence relation defined above, is a nonorientable closed surface denoted by N_g.*

The surface N_g and any other surface S homeomorphic to N_g is called *nonorientable surface of genus g*. We can extend the definition of N_g to the case $g = 1$ taking as E_2 the 2-gon or "digon" the unit 2-disk, one of whose edges is the left half-circle and the other edge, the right half-circle (see Figure 2.22). The relation described above for $g > 1$ corresponds in this case to the relation $x \sim y$ if and only if either $x = y$ or $x, y \in \mathbb{S}^1$ and $x = -y$, in other words, what we do in order to define N_1 is to identify two boundary points in the disk if they are antipodal. We call N_1 the nonorientable surface of genus 1.. By Theorem 2.2.16, N_1 is the projective plane.

The *proof* of 2.2.22 is virtually the same as that of 2.2.11, so we leave it to the reader as an *exercise.*

We do not count with enough tools to make a full study of the surfaces S_g and N_g. However we shall state the classification theorem of the closed surfaces, whose full proof we shall omit. Corollary 4.4.7 below proves the first half of the classification. We refer the reader to the books of Massey [19] or Armstrong [4] for the full proof.

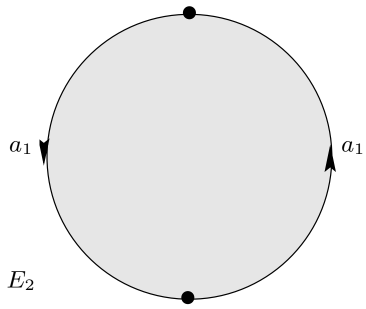


Figure 2.22 2-gon

2.2.23 **Theorem**. *The surfaces in the list*

$$S_1,\, S_2,\, \ldots \qquad \textit{and} \qquad N_1,\, N_2,\, \ldots$$

are not homeomorphic to each other. Furthermore, any closed surface is homeomorphic to one (and only one) in the list.

2.2.24 Exercise. Show that the quotient space $M/\partial M$ of the Moebius strip M that collapses the boundary ∂M to one point is homeomorphic to the projective plane $\mathbb{P}^2$. (*Hint:* The map $f : [-1,1] \times [-1,1] \longrightarrow \mathbb{D}^2$ given by $f(s,t) = (s\sqrt{1-t^2}, t)$ is an identification which maps the horizontal edges of the square to the poles of the disk and determines the desired homeomorphism between the quotient spaces $M/\partial M$ and $\mathbb{P}^2$. See 2.2.5.)

2.2.25 Exercise. In a similar way to the Klein bottle, the projective plane cannot be embedded into $\mathbb{R}^3$. Show that $\mathbb{P}^2$ can be embedded into $\mathbb{R}^4$. (*Hint:* The mapping $e : \mathbb{S}^2 \longrightarrow \mathbb{R}^4$ given by $e(x_1,x_2,x_3) = (x_1^2 - x_2^2, x_1x_2, x_1x_3, x_2x_3)$ determines an embedding $\widehat{e} : \mathbb{P}^2 \longrightarrow \mathbb{R}^4$.)

2.2.26 Exercise. A topological space X is called a *homogeneous space* if given any two points $x, y \in X$, a homeomorphism $\varphi : X \longrightarrow X$ exists, such that $\varphi(x) = y$. Show that the torus $\mathbb{S}^1 \times \mathbb{S}^1$ is a homogeneous space. (*Hint*: Given $x = (1,1)$, $y = (\zeta_1, \zeta_2) \in \mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1 \subset \mathbb{C} \times \mathbb{C}$, the map $\varphi : \mathbb{T}^2 \longrightarrow \mathbb{T}^2$, given by $\varphi(\zeta, \zeta') = (\zeta_1\zeta, \zeta_2\zeta')$ is a homeomorphism, such that $\varphi(x) = y$.)

2.2.27 Exercise. Show that given any two points $x,\, y \in \overset{\circ}{\mathbb{B}}{}^n$, there is a homeomorphism $\varphi : \mathbb{B}^n \longrightarrow \mathbb{B}^n$, such that for any point $z \in \mathbb{S}^{n-1}$, $\varphi(z) = z$, and $\varphi(x) = \varphi(y)$.

2.2.28 Exercise. Show that any connected surface is a homogeneous space. (*Hint*: Use the previous exercise.)

2.3 More low-dimensional manifolds

In this section, we shall study the manifolds of dimension 1. First of all we shall prove that they are not too many. Indeed, we shall see that the only connected 1-manifolds without boundary are homeomorphic either to $\mathbb{R}$ or to $\mathbb{S}^1$. Notwithstanding, the closed 1-manifolds become more interesting if they are embedded in the Euclidean space $\mathbb{R}^3$ or, equivalently, in the sphere $\mathbb{S}^3$. In other words, the richness of the 1-manifolds is to be found in the knots and links that we shall study in the last chapter.

Below, in this chapter, we shall study succinctly some constructions that give rise to 3-manifolds, namely the so-called Heegaard decomposition. We end the chapter stating one of the most important results in topology of the last decade, namely the classification of the simply connected 4-manifolds, due to M. Freedman, and we shall compare this classification with the classical classification of the 2-manifolds (surfaces), which was stated in Section 2.2, where we slightly change the statement.

We start by recalling that a connected 1-manifold is a connected Hausdorff second-countable space V, which has an open cover, whose elements are called U_i, and are such that for every i there is a homeomorphism of an open or half-open interval onto it, say $\varphi_i : I_i \longrightarrow U_i$.

The following lemma will be basic for the classification of manifolds of dimension 1.

2.3.1 **Lemma**. *Let X be a connected Hausdorff space such that $X = X_1 \cup X_2$, where $X_1, X_2 \subset X$ are open sets with the property that $X_1 \approx X_2 \approx \mathbb{R}$. Then $X \approx \mathbb{R}$ or $X \approx \mathbb{S}^1$.*

Proof: Assume that $X_1 \not\subseteq X_2$ and $X_2 \not\subseteq X_1$, since otherwise the result would be trivial.

Let $\varphi_1 : X_1 \longrightarrow \mathbb{R}$ and $\varphi_2 : X_2 \longrightarrow \mathbb{R}$ be homeomorphisms. Since the intersection $X_1 \cap X_2$ is open, then in X_1 as well as in X_2, the sets $\varphi_1(X_1 \cap X_2)$ and $\varphi_2(X_1 \cap X_2)$ are open in $\mathbb{R}$. Hence their components are open intervals.

None of these intervals can be bounded, since if an interval (a, b) were a component of, say, $\varphi_1(X_1 \cap X_2)$, then the set $\varphi_1^{-1}(a, b)$ would be closed in X_2 (since it is the intersection of the compact set $\varphi_1^{-1}[a, b]$ with X_2), as well as open in X_2. Then we would have $X_2 = \varphi_1^{-1}(a, b) \subset X_1$, which contradicts the initial assumption.

Furthermore, $\varphi_1(X_1 \cap X_2) \neq \mathbb{R}$, since otherwise $X_1 \subset X_2$. Similarly, $\varphi_2(X_1 \cap X_2) \neq \mathbb{R}$. We still have two more more possible cases:

(i) The set $\varphi_1(X_1 \cap X_2)$, as well as the set $\varphi_2(X_1 \cap X_2)$ are open half-lines (rays).

(ii) The set $\varphi_1(X_1 \cap X_2)$ as well as the set $\varphi_2(X_1 \cap X_2)$ are, each, the union of two disjoint open half-lines.

Since we can multiply either φ_1 or φ_2 by -1, we may assume that case in (i), $\varphi_1(X_1 \cap X_2)$ has the form $(-\infty, a)$ and $\varphi_2(X_1 \cap X_2)$ has the form (b, ∞). The composite

$$(-\infty, a) = \varphi_1(X_1 \cap X_2) \xrightarrow{\varphi_1^{-1}|} X_1 \cap X_2 \xrightarrow{\varphi_2|} \varphi_2(X_1 \cap X_2) = (b, \infty)$$

is a continuous injective map, hence it is monotonous and obviously increasing (otherwise the points $\varphi_1^{-1}(a)$ and $\varphi_2^{-1}(b)$ would not have disjoint neighborhoods in X and thus X would not be Hausdorff). Hence

$$X = \varphi_2^{-1}((-\infty, \varphi_2(x_0)]) \cup \varphi_1^{-1}([\varphi_1(x_0), \infty)) \,,$$

for some point $x_0 \in X_1 \cap X_2$, so that in case (i) X is homeomorphic to $\mathbb{R}$.

In case (ii), $\varphi_1(X_1 \cap X_2) = (-\infty, a) \cup (a', \infty)$ and $\varphi_2(X_1 \cap X_2) = (-\infty, b) \cup (b', \infty)$, for some elements a, a', b, b', $(a < a',\ b < b')$. Thus we may assume that the composed homeomorphism

$$\varphi_1(X_1 \cap X_2) \xrightarrow{\varphi_1^{-1}|} X_1 \cap X_2 \xrightarrow{\varphi_2|} \varphi_2(X_1 \cap X_2)$$

maps $(-\infty, a)$ homeomorphically onto (b', ∞), and maps (a', ∞) homeomorphically on $(-\infty, b)$.

Both homeomorphisms $(-\infty, a) \longrightarrow (b', \infty)$ and $(a', \infty) \longrightarrow (-\infty, b)$ induced by the composed homeomorphism given above, are increasing (since if, for instance, the first were not increasing, then the points $\varphi_1^{-1}(a)$ and $\varphi_2^{-1}(b')$ would not have disjoint neighborhoods in X). Thus we may write

$$X = \varphi_2^{-1}([\varphi_2(y), \varphi_2(x)]) \cup \varphi_1^{-1}([\varphi_1(x), \varphi_1(y)]) \,,$$

for certain points $x \in \varphi_1^{-1}(-\infty, a) = \varphi_2^{-1}(b', \infty)$, $y \in \varphi_1^{-1}(a', \infty) = \varphi_2^{-1}(-\infty, b)$. Therefore, in case (ii), X is homeomorphic to $\mathbb{S}^1$. □

The first part of our classification theorem is the following.

2.3.2 **Proposition**. *Any compact connected* 1*-manifold* V *is homeomorphic to* $\mathbb{S}^1$ *or to the interval* $I = [0, 1]$.

Proof: If the manifold V closed, namely, it has no boundary, then it can be covered by a finite number of open sets, each homeomorphic to $\mathbb{R}$. We can order them in a sequence $U_1, \ldots, U_k$, so that each union $V_l = U_1 \cup \cdots \cup U_l$ is connected. According to Lemma 2.3.1, the first of the sets V_l which is not homeomorphic to $\mathbb{R}$, must be homeomorphic to $\mathbb{S}^1$, and since it both open and closed, it must be the whole manifold V. Therefore, V must be homeomorphic to $\mathbb{S}^1$.

If we now assume that V has a boundary, then its double $2V$ is a closed connected 1-manifold. Hence, by the first part of this proof, $2V$ is homeomorphic to $\mathbb{S}^1$. Consequently, the original manifold V must be homeomorphic to a proper open subset of $\mathbb{S}^1$, which is closed, connected, nonempty, and it is not a point. Then it must be homeomorphic to the closed interval $[0, 1]$. □

The last result we need is the next.

2.3.3 **Lemma**. *If a topological space X can be represented as the union of a non-decreasing sequence of open subsets, all of which are homeomorphic to $\mathbb{R}$, then X itself is homeomorphic to $\mathbb{R}$.*

Proof: Let $X = \cup V_i$ be the given representation. It is clear that any homeomorphism of V_i with some interval (a, b) can be extended to a homeomorphism of V_{i+1} with one of the intervals (a, b), $(a-1, b)$, $(a, b+1)$, or $(a-1, b+1)$. This way, it is possible to construct inductively a sequence of intervals $I_1 \subset I_2 \subset \cdots \subset I_i \subset \cdots$ and a sequence of homeomorphisms $\varphi_i : V_i \longrightarrow I_i$, $i = 1, 2, \ldots$, such that $\varphi_i|_{V_{i-1}} = \varphi_{i-1} : V_{i-1} \longrightarrow I_{i-1}$. It is obvious that the map $\varphi : X \longrightarrow \cup I_i$ given by $\varphi|_{V_i} = \varphi_i$, is a homeomorphism. □

Using the previous lemmas, we may prove the second part of the classification theorem for manifolds of dimension 1.

2.3.4 **Proposition**. *Any connected, noncompact 1-manifold V is homeomorphic to $\mathbb{R}$ or to $\mathbb{R}^+$.*

Proof: Suppose first that V has no boundary. Then it can be covered with countably many open sets homeomorphic to $\mathbb{R}$. One can denote them as a sequence $U_1, U_2, \ldots$, such that all unions $U_1 \cup \cdots \cup U_k$ are connected. Hence all these unions are homeomorphic to $\mathbb{R}$, since otherwise the first of them which is not homeomorphic to $\mathbb{R}$ must be, by Lemma 2.3.1, homeomorphic to $\mathbb{S}^1$ and since it is closed and open, and thus must coincide with V, which is a contradiction. Hence we apply Lemma 2.3.3 to the manifold V to conclude that it is homeomorphic to $\mathbb{R}$.

If we now assume that V has a boundary, then its double $2V$ no has none, and therefore it is a connected noncompact manifold and thus it must be homeomorphic to $\mathbb{R}$. Hence we deduce that V is homeomorphic to a closed connected noncompact subset of $\mathbb{R}$, different from $\mathbb{R}$. Therefore it must be homeomorphic to $\mathbb{R}^+ = [0, \infty)$. $\square$

We may summarize Propositions 2.3.2 and 2.3.4 in the following theorem, which provides the full classification of the 1-manifolds.

2.3.5 **Theorem**. *Let V be a connected 1-manifold. Then the following hold:*

1. *If V has no boundary, then there are two possibilities:*
   (a) *if V is compact, then $V \approx \mathbb{S}^1$;*
   (b) *if V is not compacta, then $V \approx \mathbb{R}$;*
2. *if V has boundary, then there are two possibilities:*
   (a) *if ∂V is connected, then $V \approx [0, \infty)$;*
   (b) *if ∂V is disconnected, then $V \approx [0, 1] = I$.* $\square$

In what follows, we shall explain some features about 3-dimensional manifolds. We start by considering two solid tori $T_1 = \mathbb{S}^1 \times \mathbb{D}^2$ and $T_2 = \mathbb{D}^2 \times \mathbb{S}^1$. These are 3-dimensional manifolds whose boundaries are equal, namely $\partial T_1 = \partial T_2 = \mathbb{S}^1 \times \mathbb{S}^1$. If we now identify T_1 with T_2 along their common boundaries, using the identity, in other words, if we take the attaching space corresponding to the attaching situation

$$T_1 \supset \mathbb{S}^1 \times \mathbb{S}^1 \hookrightarrow T_2 \,,$$

then the corresponding identification space is precisely

$$T_1 \cup T_2 = (\mathbb{S}^1 \times \mathbb{D}^2) \cup (\mathbb{D}^2 \times \mathbb{S}^1) = (\partial\mathbb{D}^2 \times \mathbb{D}^2) \cup (\mathbb{D}^2 \times \partial\mathbb{D}^2) = \partial(\mathbb{D}^2 \times \mathbb{D}^2) \approx \partial\mathbb{B}^4 = \mathbb{S}^3 \,,$$

namely the boundary of a 4-ball (see 2.1.8). We have thus shown the following.

2.3.6 **Proposition**. *The union of two copies of the solid torus $T = \mathbb{S}^1 \times \mathbb{D}^2$ along their boundary, through the homeomorphism $\partial T \longrightarrow \partial T$ given by $(x, y) \mapsto (y, x)$, is homeomorphic to the unit 3-sphere $\mathbb{S}^3$.* $\square$

This situation is shown in Figure 2.23, where we can appreciate how it is possible that the union of two (solid) balls along their boundary is the same thing as the union of two solid tori along their boundary.

In general, there are two homeomorphisms of the torus onto itself $\varphi : \partial T \longrightarrow \partial T$, which are essentially different to the one described in the previous proposition. In general, they determine 3-manifolds which are different to the 3-sphere.

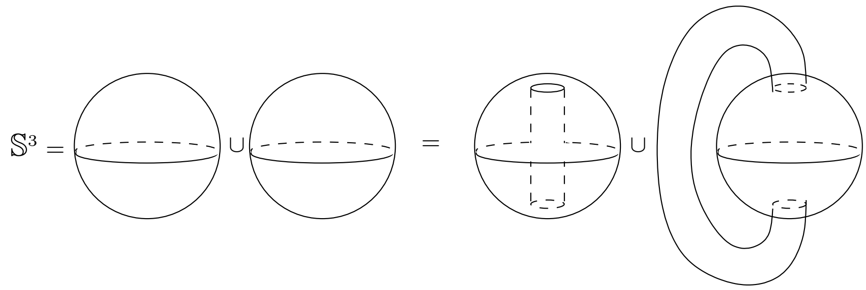


Figure 2.23 A 3-sphere is the union of two 3-balls or of two solid tori

Important members of this family of automorphisms of the torus are the homeomorphisms $f^{ac}_{bd} : \mathbb{S}^1 \times \mathbb{S}^1 \longrightarrow \mathbb{S}^1 \times \mathbb{S}^1$, given by $f^{ac}_{bd}(\zeta, \eta) = (\zeta^a \cdot \eta^b, \zeta^c \cdot \eta^d)$, $a, b, c, d \in \mathbb{Z}$, $ad - bc = \pm 1$, which will be described in more detail below in 4.2.15.

We may define the corresponding attaching spaces M^{ac}_{bd} resulting from the attaching situations

$$T_1 \supset \mathbb{S}^1 \times \mathbb{S}^1 \xrightarrow{f^{ac}_{bd}} \mathbb{S}^1 \times \mathbb{S}^1 \subset T_2 \,.$$

2.3.7 Exercise. For $a, b, c, d \in \mathbb{Z}$ such that $ad - bc = \pm 1$, show that the map $f^{ac}_{bd} : \mathbb{S}^1 \times \mathbb{S}^1 \longrightarrow \mathbb{S}^1 \times \mathbb{S}^1$ is a homeomorphism whose inverse is of the same type, and give the inverse explicitly.

2.3.8 Exercise. Show that if $f : \mathbb{S}^1 \times \mathbb{S}^1 \longrightarrow \mathbb{S}^1 \times \mathbb{S}^1$ is a homeomorphism, then the attaching space M corresponding to the attaching situation

$$T_1 \supset \mathbb{S}^1 \times \mathbb{S}^1 \xrightarrow{f} \mathbb{S}^1 \times \mathbb{S}^1 \subset T_2$$

is a 3-dimensional compact connected manifold without boundary.

From the previous exercises, we obtain immediately the following result.

2.3.9 **Theorem**. *The attaching spaces M^{ac}_{bd}, $a, b, c, d \in \mathbb{Z}$, $ad - bc = \pm 1$, are manifolds of dimension* 3*, connected, compact and without boundary, namely they are closed* 3*-manifolds.* □

Take $M = M^{ac}_{bd}$ with attaching map $f = f^{ac}_{bd}$. Take another collection $a', b', c', d' \in \mathbb{Z}$ such that $a'd' - b'c' = \pm 1$, then we have $M' = M^{a'c'}_{b'd'}$ with attaching map $f' = f^{a'c'}_{b'd'}$. Consider $\alpha, \beta, \gamma, \delta \in \{-1, 1\}$ and $m, n \in \mathbb{Z}$. Then we have homeomorphisms $F : \mathbb{D}^2 \times \mathbb{S}^1 \longrightarrow \mathbb{D}^2 \times \mathbb{S}^1$, given by $F(\zeta, \eta) = (\zeta^\alpha \eta^m, \eta^\beta)$, and

$G : \mathbb{S}^1 \times \mathbb{D}^2 \longrightarrow \mathbb{S}^1 \times \mathbb{D}^2$, $G(\zeta, \eta) = (\zeta^\gamma \eta^n, \eta^\delta)$, where $\xi^{-1} = \overline{\xi}$. Assume that the following equations hold:

$$\gamma a = \alpha a', \ \ \gamma b = ma' + \beta b', \ \ \alpha c' = na + \delta c, \ \ mc' + \delta c' = nb + \delta d\,. \tag{2.3.10}$$

Then $(G|\mathbb{S}^1 \times \mathbb{S}^1) \circ f = f' \circ (F|\mathbb{S}^1 \times \mathbb{S}^1)$, that is, the following diagram commutes:

$$\begin{tikzcd}
\mathbb{S}^1 \times \mathbb{S}^1 \arrow[r, "F|\mathbb{S}^1\times\mathbb{S}^1"] \arrow[d, "f"'] & \mathbb{S}^1 \times \mathbb{S}^1 \arrow[d, "f'"] \\
\mathbb{S}^1 \times \mathbb{S}^1 \arrow[r, "G|\mathbb{S}^1\times\mathbb{S}^1"'] & \mathbb{S}^1 \times \mathbb{S}^1\,,
\end{tikzcd}$$

so that clearly $M \approx M'$.

In the case $a = 0$, one has that $bc = \pm 1$, and we can take $\alpha = \gamma = -1$, $\beta = -b$, $\delta = -c$, $n = 0$ and $m = -cd$. Then $M \approx M^{1\,0}_{0\,1} \approx (\mathbb{D}^2 \cup_{\mathrm{id}} \mathbb{D}^2) \times \mathbb{S}^1 \approx \mathbb{S}^2 \times \mathbb{S}^1$.

In the case $a = \pm 1$, we take $\alpha = \delta = a$, $\beta = ad - bc$, $\gamma = 1$, $m = b$ and $n = -c$. Then $M \approx M^{0\,1}_{1\,0} \approx \mathbb{S}^3$ (*exercise*, see Figure 2.23).

In what follows, we assume $|a| \geq 2$. If $a < 0$, we take $\alpha = -1$, $\beta = \gamma = \delta = 1$ and $m = n = 0$ and we obtain $a' > 0$, so that we may assume from the beginning that $a > 0$. If $ad - bc = -1$, then we assume $\alpha = \beta = \gamma = 1$, $\delta = -1$ and $m = n = 0$, so that we obtain $a'd' - b'c' = 1$, where $a' = a > 0$. Therefore, we consider $ad - bc = 1$. The numbers c and d are determined by a and b in the following sense. If one has $ad'' - bc'' = 1$, then $c'' = c + na$ and $d'' = d + nb$ for some $n \in \mathbb{Z}$. If in (2.3.10) we take these n and $m = 0$, $\alpha = \beta = \gamma = \delta = 1$, then $M \approx M^{ac''}_{bd''}$. Hence M, up to homeomorphism, is determined by a and b, and it is possible to simply write M^a_b instead of M^{ac}_{bd}. Taking now $n = 0$ and $\alpha = \beta = \gamma = \delta = -1$ in (2.3.10), $M^a_b \approx M^a_{b+ma}$, so one may reduce b modulo a and obtain $1 \leq b < a$. The manifold M^a_b is known as the *lens space* associated to a, b, usually denoted by $L(a, b)$. We have shown the following.

2.3.11 **Theorem**. *The 3-manifold M^{ac}_{bd} is homeomorphic to the sphere $\mathbb{S}^3$, to $\mathbb{S}^2 \times \mathbb{S}^1$, or to the lens space $L(a, b)$, where a and b are relatively prime and $1 \leq b < a$.* □

2.3.12 Note. The given definition of the lens spaces is *ad hoc*. For instance, in [4] the traditional definition of a lens space is given. It is thus interesting to show that our lens spaces are homeomorphic to the traditional one.

2.3.13 Example. Let $g \geq 1$ be an integer and let $\mathbb{D}_g$ be the disc with g holes, as in 2.2.9(a). Then $H_g = \mathbb{D}_g \times I$ is a 3-manifold whose boundary $A_g = \partial(\mathbb{D}_g \times I)$ is an orientable surface of genus g. The space H_g is called *handle-body* of genus

g. If $\varphi : A_g \longrightarrow A_g$ is some homeomorphism, then $M = H_g \cup_\varphi H_g$ is a closed 3-manifold. The pair (H_g, φ) is known as *Heegaard decomposition* of M of genus g.

The following result is due to Dyck and Heegaard. Its proof goes outside of the scope of this text, so we omit it (see [14]).

2.3.14 **Theorem**. *Every closed orientable 3-manifold M admits a Heegaard decomposition of genus $g \geq 0$, namely $M \approx H_g \cup_\varphi H_g$ for some g and some homemorphism $\varphi : A_g \longrightarrow A_g$, $A_g = \partial H_g$.*

One dimension higher, that is, in dimension 4, the situation turns to be particularly interesting. As we already noticed in 2.2.23, Moebius gave the classification of all closed orientable manifolds of dimension 2. Using homological techniques, it is possible to assign a bilinear form to every even-dimensional manifold –say of dimension $2n$–, which is called its *intersection form*, and in a sense counts the (finite) number of points (with a sign) in which its submanifolds of dimension n intersect. This is registered in the homology in dimension n (the *semidimension* of M). To this bilinear form one can canonically associate a symmetric *unimodular* matrix (namely with determinant ± 1) with integral coefficients. In the case of the surfaces, to each symmetric unimodular matrix, put in canonical form, corresponds a surface. The classification theorem of surfaces 2.2.23 can be reformulated in terms of these symmetric unimodular matrices as follows.

2.3.15 **Theorem**. *Let S be a path-connected (0-connected) surface. If S is orientable, then its canonical matrix can be 0, in whose case $S \approx \mathbb{S}^2$, or it has the form*

$$\begin{pmatrix} 0 & 1 & & 0 & 0 \\ 1 & 0 & & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & & 0 & 1 \\ 0 & 0 & & 1 & 0 \end{pmatrix},$$

in whose case $S \approx (\mathbb{S}^1 \times \mathbb{S}^1)\# \cdots \#(\mathbb{S}^1 \times \mathbb{S}^1)$, with as many summands as there are 2×2 blocks in the matrix. If S is nonorientable, then its canonical matrix has the form

$$\begin{pmatrix} 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & & 0 & 1 \end{pmatrix},$$

in whose case $S \approx \mathbb{RP}^2 \# \cdots \# \mathbb{RP}^2$, with as many summands as there are ones in the diagonal.

The theorem assures that the matrices are invariants which classify surfaces: In the first case one deals with block matrices of dimension $2g \times 2g$. In the second case one deals with identity matrices of dimension $g \times g$, where in both cases, g is the genus of S.

In dimension 4 the situation is particularly interesting. In the first place, it does not seem possible to give a full classification theorem, that is a theorem that classifies *all* 4-manifolds. There are algebraic reasons for that, as we comment in 4.4.18. However, back in 1982 a fundamental step in the classification was given. Freedman [9] gave the classification of all simply connected 4-manifolds (see 4.1.24 below). Freedman shows that given a unitary, unimodular canonical matrix, there is a 4-manifold that corresponds to the matrix, and that these matrices classify the manifolds. There is an exceptional symmetric, unimodular 8×8-matrix, namely

$$E_8 = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 \end{pmatrix},$$

to which, according to Freedman's results, a certain simply connected manifold of dimension 4 corresponds. We denote this manifold by V_8. In a similar form as one way to endow a 4-manifold with an orientation or not, one can endow it with a structure called *spin*. The classification theorem is the following.

2.3.16 **Theorem**. *Let V be a simply connected (1-connected) 4-manifold. If V is spin, then its canonical matrix can be 0, in whose case $V \approx \mathbb{S}^4$, or it has the form*

$$\begin{pmatrix} 0 & 1 & & & & & & 0 & 0 \\ 1 & 0 & & & & & & 0 & 0 \\ & & \ddots & & & & & & \\ & & & 0 & 1 & & & & \\ & & & 1 & 0 & & & & \\ & & & & & E_8 & & & \\ 0 & 0 & & & & & \ddots & & \\ 0 & 0 & & & & & & & E_8 \end{pmatrix},$$

in whose case $V \approx (\mathbb{S}^2 \times \mathbb{S}^2)\# \cdots \#(\mathbb{S}^2 \times \mathbb{S}^2)\#V_8\# \cdots \#V_8$, with as many summands of the form $\mathbb{S}^2 \times \mathbb{S}^2$ as there are 2×2-blocks in the matrix, and as many summands V_8 as there are E_8-blocks in the matrix. If V is not spin, then its canon-

ical matrix has the form

$$\begin{pmatrix} 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & & -1 & 0 \\ 0 & 0 & & 0 & -1 \end{pmatrix},$$

in whose case $V \approx \mathbb{CP}^2 \# \cdots \# \mathbb{CP}^2 \# \overline{\mathbb{CP}}^2 \# \cdots \# \overline{\mathbb{CP}}^2$*, with as many summands* $\mathbb{CP}^2$ *as there are ones, and as many summands* $\overline{\mathbb{CP}}^2$ *as there are minus ones in the diagonal of the matrix. The space* $\overline{\mathbb{CP}}^2$ *denotes the usual complex projective plane with the opposite orientation to that of* $\mathbb{CP}^2$*.*

The parallelism between this statement on the classification of 4-manifolds and that of the classification of 2-manifolds 2.3.15 is astonishing. The two main differences are the block E_8 in the matrix, as well as the summand V_8 in the 4-manifold. The reader may take a look at [11] for details on this result.

2.4 Classical groups

Starting with groups of matrices, known as *classical groups* which happen to be manifolds, in this section we shall define a variety of other manifolds, which are useful in several areas of mathematics. Examples are the Grassmann and the Stiefel manifolds. We shall see what their relationships are and how the Grassman manifolds generalize the projective spaces. Furthermore, we shall briefly study the classical groups.

Let us begin by considering the space $\mathrm{M}_n(\mathbb{R})$ (resp. $\mathrm{M}_n(\mathbb{C})$) of all $n \times n$ matrices with real (resp. complex) entries. Putting the entries one line after the other, one may give homeomorphisms

$$\mathrm{M}_n(\mathbb{R}) \approx \mathbb{R}^{n^2}, \quad \mathrm{M}_n(\mathbb{C}) \approx \mathbb{C}^{n^2} \approx \mathbb{R}^{2n^2}.$$

The determinant functions

$$\det_{\mathbb{R}} : \mathrm{M}_n(\mathbb{R}) \longrightarrow \mathbb{R}, \quad \det_{\mathbb{C}} : \mathrm{M}_n(\mathbb{C}) \longrightarrow \mathbb{C}$$

are continuous. Therefore they are continuous functions. Hence the subsets

$$\mathrm{GL}_n(\mathbb{R}) = \det_{\mathbb{R}}^{-1}(\mathbb{R} - 0), \quad \mathrm{GL}_n(\mathbb{C}) = \det_{\mathbb{C}}^{-1}(\mathbb{C} - 0)$$

are open and, therefore they are open sets and thus submanifolds of $\mathbb{R}^{n^2}$ and of $\mathbb{C}^{n^2} = \mathbb{R}^{2n^2}$, so that they have dimensions n^2 and $2n^2$, respectively. These sets consist of invertible matrices and, with respect to matrix multiplication, they are groups. Since this multiplication, as well as the map that sends a matrix to its inverse matrix (using Cramer's rule), are continuous, then the matrix groups are *topological groups*

2.4.1 DEFINITION. The manifolds $\mathrm{GL}_n(\mathbb{R})$ and $\mathrm{GL}_n(\mathbb{C})$ are known as the *real general linear group of $n \times n$ matrices* and the *complex general linear group of $n \times n$ matrices* (or shorter, the *nth real general linear group* and the *nth complex general linear group*). Furthermore, $\dim \mathrm{GL}_n(\mathbb{R}) = n^2$ and $\dim \mathrm{GL}_n(\mathbb{C}) = 2n^2$.

Given a real (resp. complex) $n \times n$ matrix A, we may interpret its columns $A_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix}$ as vectors in $\mathbb{R}^n$ (resp. $\mathbb{C}^n$). The vectors $A_1, \ldots, A_n$ are linearly independent if and only if $A \in \mathrm{GL}_n(\mathbb{R})$ (resp. $A \in \mathrm{GL}_n(\mathbb{C})$). Let us assume a little more, namely that these vectors build an orthonormal basis, that is they satisfy the $\frac{n(n+1)}{2}$ equations $\langle A_i, A_j \rangle = \delta_{ij}$, $1 \leq i \leq j \leq n$, where $\langle -,- \rangle$ represents the usual scalar product in $\mathbb{R}^n$ (resp. the usual hermitian product in $\mathbb{C}^n$). Putting these equations (of which, in the complex case, n of them have real values and $\frac{n(n+1)}{2} - n = \frac{n(n-1)}{2}$, complex values) in an adequate order, one gets mappings

$$\varphi : \mathrm{GL}_n(\mathbb{R}) \longrightarrow \mathbb{R}^{\frac{n(n+1)}{2}}, \quad (\text{resp. } \varphi : \mathrm{GL}_n(\mathbb{C}) \longrightarrow \mathbb{R}^{n^2}),$$

which are even differentiable. For each point $A \in \mathrm{GL}_n(\mathbb{R})$ (resp. $A \in \mathrm{GL}_n(\mathbb{C})$), such that $\varphi(A) = \delta$, where $\delta \in \mathbb{R}^{\frac{n(n+1)}{2}}$ (resp. $\delta \in \mathbb{R}^{n^2}$) is the point given by the Kronecker delta δ_{ij}, $i \leq j = 1, \ldots, n$, then its derivative has maximal rank. The implicit function theorem states that there is a neighborhood U_A of A in $\mathrm{GL}_n(\mathbb{R})$ (resp. in $\mathrm{GL}_n(\mathbb{C})$), a neighborhood V_A of 0 in $\mathbb{R}^{n^2}$ (resp. $\mathbb{R}^{2n^2}$), and homeomorphisms $\psi_A : V_A \longrightarrow U_A$, which map 0 to A. These maps are such that the composites $\varphi' \circ \psi_A : V_A \longrightarrow \mathbb{R}^{\frac{n(n+1)}{2}}$ (resp. $\varphi' \circ \psi_A : V_A \longrightarrow \mathbb{R}^{n^2}$), where $\varphi' = \varphi - \delta$, correspond simply to the projection onto the first $\frac{n(n+1)}{2}$ real coordinates (resp. onto the first n^2 real coordinates). This way, the set of solutions of the $\frac{n(n+1)}{2}$ equations $M = \varphi^{-1}(\delta) = \varphi'^{-1}(0)$ is such that the restriction of ψ_A to the last $\frac{n(n-1)}{2}$ (resp. n^2) real coordinates induces a homeomorphism, which we denote again by $\psi_A : U_A \cap (0 \times \mathbb{R}^{\frac{n(n-1)}{2}}) \longrightarrow V_A \cap M$ (resp. $\psi_A : U_A \cap (0 \times \mathbb{R}^{n^2}) \longrightarrow V_A \cap M$). Figure 2.24 shows what we have done.

We have shown that each point $A \in M$ has a neighborhood $W_A = V_A \cap M$ and there is a homeomorphism $\psi_A : U'_A \longrightarrow W_A$, where U'_A is open in $\mathbb{R}^{\frac{n(n-1)}{2}}$ (resp en $\mathbb{R}^{n^2}$). Therefore M is a manifold of dimension $\frac{n(n-1)}{2}$ (resp. n^2). Furthermore, since M is a set of solutions, it must be closed and since each matrix $A \in M$ it is built up by unit vectors, clearly M is bounded. Hence M is a closed manifold (namely, it is compact and has no boundary).

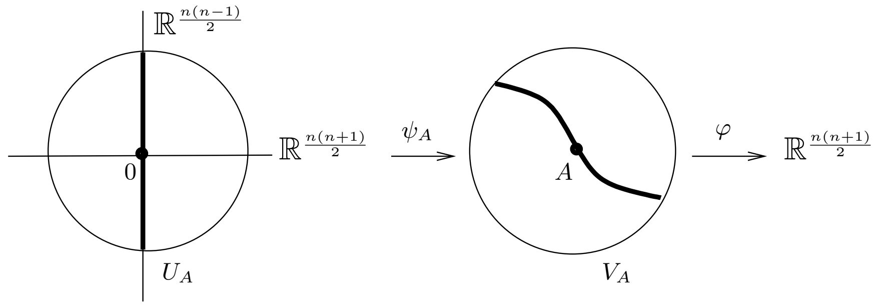


Figure 2.24 The implicit function theorem

The subset $M \subset \mathrm{GL}_n(\mathbb{R})$ (resp. $M \subset \mathrm{GL}_n(\mathbb{C})$) is indeed a subgroup. It is characterized by the following: A matrix $A \in \mathrm{GL}_n(\mathbb{R})$ (resp. $A \in \mathrm{GL}_n(\mathbb{C})$) is such that $A \in M$ if and only if $AA^* = 1$, where A^* is the transposed (resp. conjugate transposed) matrix of A and 1 is the identity matrix.

2.4.2 Definition. The group $\mathrm{O}_n = \{A \in \mathrm{GL}_n(\mathbb{R}) \mid AA^* = 1\}$ is called *nth orthogonal group* or *orthogonal group of $n \times n$ matrices* and the group $\mathrm{U}_n = \{A \in \mathrm{GL}_n(\mathbb{C}) \mid AA^* = 1\}$, *nth unitary group* or *unitary group of $n \times n$ matrices*. The first is a closed manifold of dimension $\frac{n(n-1)}{2}$, and the second is a closed manifold of dimension n^2. O_n and U_n are topological groups with respect to the relative topology induced by $\mathrm{GL}_n(\mathbb{R})$ and $\mathrm{GL}_n(\mathbb{C})$, respectively.

We have a more general construction. One considers the so-called *frames*, namely collections of k orthonormal vectors in $\mathbb{R}^n$, $k \leq n$. This is equivalent to taking matrices with k columns and n rows, i.e., $n \times k$ matrices, $A = (A_1, \ldots, A_k)$, where $A_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix}$ is a vector in $\mathbb{R}^n$ (resp. $\mathbb{C}^n$) and the equality $\langle A_i, A_j \rangle = \delta_{ij}$, $1 \leq i \leq j \leq k$, where again $\langle -, - \rangle$ represents the usual scalar product in $\mathbb{R}^n$ (resp. the usual hermitian product in $\mathbb{C}^n$). Take the sets

$$\mathrm{V}_k(\mathbb{R}^n) = \{A = (A_1, \ldots, A_k) \in \mathbb{R}^n \times \cdots \times \mathbb{R}^n \mid \langle A_i, A_j \rangle = \delta_{ij}\},$$

$$\mathrm{V}_k(\mathbb{C}^n) = \{A = (A_1, \ldots, A_k) \in \mathbb{C}^n \times \cdots \times \mathbb{C}^n \mid \langle A_i, A_j \rangle = \delta_{ij}\},$$

$1 \leq i \leq j \leq k$. We may consider $\mathrm{V}_k(\mathbb{R}^n)$ (resp. $\mathrm{V}_k(\mathbb{C}^n)$) as the subset of $\mathbb{R}^{nk}$, (resp. $\mathbb{C}^{nk}$) of solutions of the $\frac{k(k+1)}{2}$ (resp. $k(k+1) - k = k^2$) equations

$$\langle A_i, A_j \rangle = \delta_{ij}, \quad 1 \leq i \leq j \leq k.$$

The same arguments used above for the orthogonal groups (resp. the unitary groups), show that $\mathrm{V}_k(\mathbb{R}^n)$ (resp. $\mathrm{V}_k(\mathbb{C}^n)$) is a closed manifold of dimension

$nk - \frac{k(k+1)}{2} = \frac{k(2n-k-1)}{2}$ (resp. $2nk - k^2 = k(2n-k)$). Hence we have the following.

2.4.3 Definition. The set $\mathrm{V}_k(\mathbb{R}^n) = \{A = (A_1, \dots, A_k) \in \mathbb{R}^n \times \cdots \times \mathbb{R}^n \mid \langle A_i, A_j \rangle = \delta_{ij},\ 1 \leq i \leq j \leq k\}$ is called the *real Stiefel manifold of k-frames in* $\mathbb{R}^n$ and the set $\mathrm{V}_k(\mathbb{C}^n) = \{A = (A_1, \dots, A_k) \in \mathbb{C}^n \times \cdots \times \mathbb{C}^n \mid \langle A_i, A_j \rangle = \delta_{ij}, 1 \leq i \leq j \leq k\}$ is called the *real Stiefel manifold of k-frames in* $\mathbb{C}^n$. The first is a closed manifold of dimension $\frac{k(2n-k-1)}{2}$, while the second is a closed manifold of dimension $k(2n-k)$.

2.4.4 Note. If in the previous definition we take $k = n$, we clearly have that $\mathrm{V}_n(\mathbb{R}^n) = \mathrm{O}_n$ and that $\mathrm{V}_n(\mathbb{C}^n) = \mathrm{U}_n$. On the other hand, if we take $k = 1$, we have that $\mathrm{V}_1(\mathbb{R}^n) = \mathbb{S}^{n-1}$ and that $\mathrm{V}_1(\mathbb{C}^n) = \mathbb{S}^{2n-1}$.

Take an element $M \in \mathrm{O}_n$ (resp. $M \in \mathrm{U}_n$) and an element $A \in \mathrm{V}_k(\mathbb{R}^n)$ (resp. $A \in \mathrm{V}_k(\mathbb{C}^n)$). The matrix M acts on each vector A_i in the k-frame A. Since M an orthogonal (resp. unitary) matrix, the image vectors MA_i, $i = 0, 1, \dots, k$, build up again an orthogonal frame in $\mathbb{R}^n$ (resp. in $\mathbb{C}^n$). We denote this new k-frame by MA. This determines an *action* of the orthogonal (unitary) group on the Stiefel manifold (see 5.1.22), since it is simple to check that the mapping $\mathrm{O}_n \times \mathrm{V}_k(\mathbb{R}^n) \longrightarrow \mathrm{V}_k(\mathbb{R}^n)$ (resp. $\mathrm{U}_n \times \mathrm{V}_k(\mathbb{C}^n) \longrightarrow \mathrm{V}_k(\mathbb{C}^n)$) given by $(M, A) \mapsto MA$ is continuous and the equations $IA = A$ and $(MN)A = M(NA)$ hold (see 1.4.4).

2.4.5 Exercise. Show that given any two frames $A, B \in \mathrm{V}_k(\mathbb{R}^n)$ (resp. $\mathrm{V}_k(\mathbb{C}^n)$), then there is a matrix $M \in \mathrm{O}_n$ (resp. $A \in \mathrm{U}_n$) such that $MA = B$. In other words, show that the action of O_n (resp. of U_n) on $\mathrm{V}_k(\mathbb{R}^n)$ (resp. $\mathrm{V}_k(\mathbb{C}^n)$) is *transitive* (see 1.4.12).

The former exercise shows that the Stiefel manifold $\mathrm{V}_k(\mathbb{R}^n)$ (resp. $\mathrm{V}_k(\mathbb{C}^n)$), at least as a set, is homogeneous (see 1.4.13), that is, it is a quotient of the group O_n by a (closed) subgroup. Namely, if we take the *canonical* k-frame$E$ defined by $E_i = e_i$, where e_i denotes the ith canonical unit vector in $\mathbb{R}^n$, $i = 1, \dots, k$, then for each $A \in \mathrm{V}_k(\mathbb{R}^n)$ (resp. $A \in \mathrm{V}_k(\mathbb{C}^n)$), there is a matrix $M_A \in \mathrm{O}_n$ (resp. $M_A \in \mathrm{U}_n$) such that $M_A E = A$. Thus we have a surjective function

$$\mathrm{O}_n \longrightarrow \mathrm{V}_k(\mathbb{R}^n) \quad (\text{resp.} \quad \mathrm{U}_n \longrightarrow \mathrm{V}_k(\mathbb{C}^n))\,,$$

given by $M \mapsto ME$. One easily verifies that this is a continuous map. But O_n (resp. U_n) is a compact space and $\mathrm{V}_k(\mathbb{R}^n)$ (resp. $\mathrm{V}_k(\mathbb{C}^n)$) is a Hausdorff space. Thus we have the following.

2.4.6 **Proposition**. *The maps* $\mathrm{O}_n \longrightarrow \mathrm{V}_k(\mathbb{R}^n)$ *and* $\mathrm{U}_n \longrightarrow \mathrm{V}_k(\mathbb{C}^n)$ *given by* $M \mapsto ME$ *are identifications.* □

2.4.7 EXERCISE. Show that $ME = NE$, $M, N \in \mathrm{O}_n$ (resp. $M, N \in \mathrm{U}_n$), where $E \in \mathrm{V}_k(\mathbb{R}^n)$ (resp. $E \in \mathrm{V}_k(\mathbb{C}^n)$) is the canonical k-frame, if and only if there is a matrix $K \in \mathrm{O}_{n-k} \subset \mathrm{O}_n$ ($\mathrm{U}_{n-k} \subset \mathrm{U}_n$) such that $MK = N$, where O_{n-k} is seen as a closed subgroup of O_n (resp. U_{n-k} as a closed subgroup of U_n) via

$$K \longmapsto \begin{pmatrix} 1_k & 0 \\ 0 & K \end{pmatrix},$$

where $1_k \in \mathrm{O}_k$ (resp. $1_k \in \mathrm{U}_k$) is the identity matrix.

The previous exercise shows that the Stiefel manifold of k-frames in $\mathbb{R}^n$ (resp. in $\mathbb{C}^n$) is the set of cosets of the subgroup O_{n-k} of O_n (resp. of the subgroup U_{n-k} of U_n) with the identification topology. Namely

$$\mathrm{O}_n/\mathrm{O}_{n-k} \approx \mathrm{V}_k(\mathbb{R}^n) \quad (\text{resp.} \quad \mathrm{U}_n/\mathrm{U}_{n-k} \approx \mathrm{V}_k(\mathbb{C}^n))\,. \tag{2.4.8}$$

In other words, the Stiefel manifolds are homogeneous spaces (see 1.4.13).

An alternative way for defining the real and complex projective spaces $\mathbb{RP}^{n-1}$ and $\mathbb{CP}^{n-1}$ is as certain spaces of real or complex straight lines (namely of subspaces of real or complex dimension 1) of $\mathbb{R}^n$ or of $\mathbb{C}^n$. In other words, consider in $\mathbb{R}^n - \{0\}$ (resp. in in $\mathbb{C}^n - \{0\}$) the equivalence relation $x \sim y$ if and only if there is a real (resp. complex) number ξ such that $y = \xi x$. Then $\mathbb{RP}^{n-1} \approx \mathbb{R}^n - \{0\}/\sim$ (resp. $\mathbb{CP}^{n-1} \approx \mathbb{C}^n - \{0\}/\sim$).

2.4.9 EXERCISE. Consider the topological group $G = \mathbb{R} - \{0\}$ (resp. $G = \mathbb{C} - \{0\}$ described in 1.4.2). Then G acts on $\mathbb{R}^n - \{0\}$ (resp. on $\mathbb{C}^n - \{0\}$) by scalar multiplication. Show that the inclusion $\mathbb{S}^{n-1} \hookrightarrow \mathbb{R}^n - \{0\}$ (resp. $\mathbb{S}^{2n-1} \hookrightarrow \mathbb{C}^n - \{0\}$) induces a homeomorphism $\mathbb{RP}^n \longrightarrow \mathbb{R}^n - \{0\}/\sim$ (resp. $\mathbb{CP}^n \longrightarrow \mathbb{C}^n - \{0\}/\sim$).

More generally we may consider the sets of subspaces of (real) dimension k of $\mathbb{R}^n$ (resp. of complex dimension k of $\mathbb{C}^n$). Take a frame $A \in \mathrm{V}_k(\mathbb{R}^n)$ (resp. $A \in \mathrm{V}_k(\mathbb{C}^n)$). Since this frame consists of k linearly independent vectors in $\mathbb{R}^n$ (resp. in $\mathbb{C}^n$), then A determines a subspace of real dimension k, $L_A \subset \mathbb{R}^n$ (resp. of complex dimension k, $L_A \subset \mathbb{C}^n$). Furthermore, any subspace L of dimension k of $\mathbb{R}^n$ (resp. of $\mathbb{C}^n$) has an orthonormal basis $A \in \mathrm{V}_k(\mathbb{R}^n)$ (resp. $A \in \mathrm{V}_k(\mathbb{C}^n)$). If we call $\mathrm{G}_k(\mathbb{R}^n)$ (resp. $\mathrm{G}_k(\mathbb{C}^n)$) the set of subspaces of real dimension k of $\mathbb{R}^n$ (resp. of complex dimension k of $\mathbb{C}^n$), then there are surjective functions

$$q : \mathrm{V}_k(\mathbb{R}^n) \twoheadrightarrow \mathrm{G}_k(\mathbb{R}^n) \quad (\text{resp.} \quad q : \mathrm{V}_k(\mathbb{C}^n) \twoheadrightarrow \mathrm{G}_k(\mathbb{C}^n))\,.$$

Giving the codomain the identification topology, then the set $\mathrm{G}_k(\mathbb{R}^n)$ (resp. $\mathrm{G}_k(\mathbb{C}^n)$) becomes a compact topological space.

2.4.10 Exercise. Show that two frames $A, B \in \mathrm{V}_k(\mathbb{R}^n)$ (resp. $A, B \in \mathrm{V}_k(\mathbb{C}^n)$) determine the same k-dimensional subspace $L_A = L_B$ of $\mathbb{R}^n$ (resp. of $\mathbb{C}^n$) if and only if there is a matrix $K \in \mathrm{O}_k \subset \mathrm{O}_n$ ($K \in \mathrm{U}_k \subset \mathrm{U}_n$)such that $KA = B$, where O_k (resp. U_k) is seen as a subgroup of O_n (resp. of U_n) via

$$K \longmapsto \begin{pmatrix} K & 0 \\ 0 & 1_{n-k} \end{pmatrix},$$

where $1_{n-k} \in \mathrm{O}_{n-k}$ (resp. $1_{n-k} \in \mathrm{U}_{n-k}$) is the identity matrix. Conclude that taking the composite

$$\mathrm{O}_n \longrightarrow \mathrm{V}_k(\mathbb{R}^n) \longrightarrow \mathrm{G}_k(\mathbb{R}^n) \quad (\text{resp.} \quad \mathrm{U}_n \longrightarrow \mathrm{V}_k(\mathbb{C}^n) \longrightarrow \mathrm{G}_k(\mathbb{C}^n)\,),$$

then $L_{ME} = L_{NE}$ if and only if there is a matrix $K \in \mathrm{O}_k \times \mathrm{O}_{n-k} \subset \mathrm{O}_n$ (resp. $K \in \mathrm{U}_k \times \mathrm{U}_{n-k} \subset \mathrm{U}_n$), where we include $\mathrm{O}_k \times \mathrm{O}_{n-k}$ in O_n (resp. $\mathrm{U}_k \times \mathrm{U}_{n-k}$ in U_n) via

$$(K_1, K_2) \longmapsto \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix}.$$

What must be shown in the previous exercise is that the space $\mathrm{G}_k(\mathbb{R}^n)$ (resp. $\mathrm{G}_k(\mathbb{C}^n)$) of k-dimensional subspaces of $\mathbb{R}^n$ (resp. of $\mathbb{C}^n$) is the set of cosets of the subgroup $\mathrm{O}_k \times \mathrm{O}_{n-k}$ of O_n (resp. $\mathrm{U}_k \times \mathrm{U}_{n-k}$ of U_n) with the identification topology. Namely they are the homogeneous spaces

$$\mathrm{O}_n/\mathrm{O}_k \times \mathrm{O}_{n-k} \approx \mathrm{G}_k(\mathbb{R}^n) \quad (\text{resp.} \quad \mathrm{U}_n/\mathrm{U}_k \times \mathrm{U}_{n-k} \approx \mathrm{G}_k(\mathbb{C}^n))\,. \tag{2.4.10}$$

2.4.11 Note. If in the definition of $\mathrm{G}_k(\mathbb{R}^n)$ (resp. of $\mathrm{G}_k(\mathbb{C}^n)$) we take $k = n$, then we have that $\mathrm{G}_n(\mathbb{R}^n) = *$ (resp. $\mathrm{G}_n(\mathbb{C}^n) = *$). On the other hand, if we take $k = 1$, then we have that $\mathrm{G}_1(\mathbb{R}^n) = \mathbb{RP}^{n-1}$ (resp. $\mathrm{G}_1(\mathbb{C}^n) = \mathbb{CP}^{2n-1}$). Furthermore, from the equation (2.4.10) we conclude that

$$\mathrm{O}_n/\mathrm{O}_1 \times \mathrm{O}_{n-1} \approx \mathbb{RP}^{n-1} \quad (\text{resp.} \quad \mathrm{U}_n/\mathrm{U}_1 \times \mathrm{U}_{n-1} \approx \mathbb{CP}^{n-1})\,.$$

Seen from another point of view, we have

$$\mathrm{O}_n/\mathrm{O}_{n-1} \approx \mathbb{S}^{n-1} \quad (\text{resp.} \quad \mathrm{U}_n/\mathrm{U}_{n-1} \approx \mathbb{S}^{2n-1})$$

and since $\mathrm{O}_1 = \mathbb{Z}_2$ (resp. $\mathrm{U}_1 = \mathbb{S}^1$), we recover the original definition of $\mathbb{RP}^{n-1}$ (resp. of $\mathbb{CP}^{n-1}$) as the orbit space of $\mathbb{S}^{n-1}$ with respect to the antipodal action of $\mathbb{Z}_2 = \mathrm{O}_1$ (resp. the orbit space of $\mathbb{S}^{2n-1}$ with respect to the action of $\mathbb{S}^1 = \mathrm{U}_1$ given by complex multiplication).

We have shown above that spaces such as $\mathrm{V}_k(\mathbb{R}^n)$, $\mathrm{V}_k(\mathbb{C}^n)$, $\mathrm{G}_k(\mathbb{R}^n)$, or $\mathrm{G}_k(\mathbb{C}^n)$ are obtained as quotients of some topological groups modulo certain closed subgroups. Thus they are homogeneous spaces.

More precisely, by the implicit function theorem, the topological groups O_n and U_n are (smooth) manifolds and the mapping $(A, B) \mapsto AB^{-1}$ is a (smooth) map. Hence they are *Lie groups* (see 2.1.14). The considered subgroups are closed submanifolds. The following result is the smooth case of Proposition 1.4.14, whose proof uses specific techniques of the theory of Lie groups, and is beyond the scope of this book. A proof can be found in [6].

2.4.12 **Theorem**. *If G is a compact Lie group and $H \subset G$ is a closed subgroup, then the homogeneous space G/H is a closed (smooth) manifold such that* $\dim(G/H) = \dim G - \dim H$. □

Consequently for the cases analyzed above, we can write the following definition.

2.4.13 DEFINITION. The homogeneous space $\mathrm{G}_k(\mathbb{R}^n) = \{L \subset \mathbb{R}^n \mid \dim_{\mathbb{R}} L = k\}$ is called the *real Grassmann manifold of k-frames in* $\mathbb{R}^n$ and the homogeneous space $\mathrm{G}_k(\mathbb{C}^n) = \{L \subset \mathbb{C}^n \mid \dim_{\mathbb{C}} L = k\}$ is called the *complex Grassmann manifold of k-frames in* $\mathbb{C}^n$. $\mathrm{G}_k(\mathbb{R}^n)$ is a closed manifold of dimension $k(n-k)$, while $\mathrm{G}_k(\mathbb{C}^n)$ is a closed manifold of dimension $2k(n-k)$.

There are other Lie groups which play an important role in several branches of mathematics. Two of them are $\mathrm{SO}_n = \{A \in \mathrm{O}_n \mid \det A = 1\}$ and its complex counterpart $\mathrm{SU}_n = \{A \in \mathrm{U}_n \mid \det A = 1\}$. They are called the *special orthogonal group* of real $n \times n$ matrices, and the *special unitary group* of complex $n \times n$ matrices. Indeed we have that the determinant functions $\det_{\mathbb{R}} : \mathrm{O}_n \longrightarrow \mathbb{Z}_2 = \mathrm{O}_1$ and $\det_{\mathbb{C}} : \mathrm{U}_n \longrightarrow \mathbb{S}^1 = \mathrm{U}_1$ are group epimorphisms, whose kernels are the subgroups SO_n and SU_n defined above. Namely $\det_{\mathbb{R}}^{-1}(1) = \mathrm{SO}_n$ and $\det_{\mathbb{C}}^{-1}(1) = \mathrm{SU}_n$. In the real case, O_n has exactly two connected components, namely $\det_{\mathbb{R}}^{-1}(1)$ and $\det_{\mathbb{R}}^{-1}(-1)$, each of which is a (smooth) submanifold of O_n with the same dimension. The first corresponds to the subgroup of orthogonal matrices that preserve the canonical orientation of $\mathbb{R}^n$. The second consists of the matrices that reverse the orientation. They do not form a group, since the product of two of them preserves the orientation. Indeed, topologically $\mathrm{O}_n = \mathrm{SO}_n \sqcup \mathrm{SO}_n^-$, where SO_n^- is the path-component of O_n in which the matrix $1_n^- = \begin{pmatrix} -1 & 0 \\ 0 & 1_{n-1} \end{pmatrix}$ lies.

In the complex case, the group U_n is path connected and the subgroup SU_n has a smaller dimension than U_n.

2.4.14 Exercise.

(a) Show that O_n is not path connected. Show that SO_n is path connected. Conclude that SO_n^- is also path connected. (*Hint:* Use the Gram-Schmidt orthonormalization process to produce a path from any orthogonal matrix of determinant 1 to the identity matrix.)

(b) Show that U_n is path connected. (*Hint:* Use the Gram-Schmidt orthonormalization process, as well as the fact that $\mathbb{S}^1$ is path connected, to produce a path from any unitary matrix to the identity matrix.)

2.4.15 Note. The subgroup SU_n of U_n is a submanifold of codimension 1, namely $\dim \mathrm{SU}_n = \dim \mathrm{U}_n - 1 = n^2 - 1$. This follows from the fact that $\dim \mathbb{S}^1 = 1$ and thus the submanifold $\{1\}$ has codimension 1 in $\mathbb{S}^1$, and any inverse image of a regular value under a smooth surjection (submersion) preserves the codimension. Indeed there is an epimorphism $\det : \mathrm{U}_n \longrightarrow \mathbb{S}^1$ given by the determinant, such that $\mathrm{SU}_n = \det^{-1}(1)$.

More generally we have determinants

$$\det_{\mathbb{R}} : \mathrm{GL}_n(\mathbb{R}) \longrightarrow \mathbb{R}-\{0\} = \mathrm{GL}_1(\mathbb{R}) \text{ and } \det_{\mathbb{C}} : \mathrm{GL}_n(\mathbb{C}) \longrightarrow \mathbb{C}-\{0\} = \mathrm{GL}_1(\mathbb{C})\,,$$

which are epimorphisms, whose kernels are the topological groups

$$\mathrm{SL}_n(\mathbb{R}) = \det{}_{\mathbb{R}}^{-1}(1) \subset \mathrm{GL}_n(\mathbb{R}) \quad \text{ and } \quad \mathrm{SL}_n(\mathbb{C}) = \det{}_{\mathbb{C}}^{-1}(1) \subset \mathrm{GL}_n(\mathbb{C})\,.$$

These groups, which are submanifolds of $\mathrm{GL}_n(\mathbb{R})$ and $\mathrm{GL}_n(\mathbb{C})$, respectively, are known as the *real special linear group of $n \times n$ matrices* and the *complex special linear group of $n \times n$ matrices* (or shorter, the *nth real special linear group* and the *nth complex special linear group*). These groups are noncompact (up to the case $n = 1$) and their codimensions are 1 and 2, respectively, namely $\dim \mathrm{SL}_n(\mathbb{R}) = n^2 - 1$ and $\dim \mathrm{SL}_n(\mathbb{C}) = 2n^2 - 2$.

Furthermore, the following space equalities are immediate:

$$\mathrm{SL}_n(\mathbb{R}) \cap \mathrm{O}_n = \mathrm{SO}_n \quad \text{y} \quad \mathrm{SL}_n(\mathbb{C}) \cap \mathrm{U}_n = \mathrm{SU}_n\,.$$

Let us get back once more to the general linear group $\mathrm{GL}_n(\mathbb{R})$ and we consider the subgroups

$$B_n = \{A = (a_{ij}) \mid a_{ij} = 0 \text{ si } i > j\} \subset \mathrm{GL}_n(\mathbb{R})\,,$$

$$B_n^+ = \{A = (a_{ij}) \mid a_{ii} > 0\} \subset B_n\,,$$

$$B_n^{(1)} = \{A = (a_{ij}) \mid a_{ii} = 1\} \subset B_n^+\,.$$

The first is the group of *real upper triangular* $n \times n$ *matrices*, also called *Borel subgroup* of $\mathrm{GL}_n(\mathbb{R})$. The last is the group of *real upper triangular* $n \times n$ *unipotent matrices*. Consider the group of *diagonal* $n \times n$ *matrices with positive elements*

$$D_n^+ = \{ \begin{pmatrix} l_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \mid \lambda_i \in \mathbb{R}, \lambda_i > 0 \} .$$

2.4.16 EXERCISE. Show that one has a homeomorphism $D_n^+ \times B_n^{(1)} \approx B_n^+$. Notice that it is not a group isomorphism.

These subgroups of $\mathrm{GL}_n(\mathbb{R})$ play a role in the following theorem, which is called *Iwasawa decomposition theorem.*

2.4.17 **Theorem**. *The matrix multiplication defines a homeomorphism*

$$\mathrm{O}_n \times B_n^+ \approx \mathrm{GL}_n(\mathbb{R}) .$$

Proof: Take a matrix $B \in \mathrm{GL}_n(\mathbb{R})$ and let $B_j = Be_j$ be its jth column, where e_j is the canonical j-vector. By the Gram-Schmidt orthonormalization process, one obtains an orthonormal basis $A_1, \dots, A_n$ such that

$$B_j = \mu_{jj} A_j + \sum_{i<j} \mu_{ij} A_i ,$$

where $\mu_{jj} > 0$. That is, $B = AM$, where $M = (\mu_{ij}) \in B_n^+$ and $A = (A_1 \cdots A_n) \in \mathrm{O}_n$.

If $AM = A'M'$, then $A'^{-1}A = M'M^{-1} \in \mathrm{O}_n \cap B_n^+$. But $\mathrm{O}_n \cap B_n^+$ consists only of the identity matrix 1_n, since an orthogonal $n \times n$ matrix with n positive eigenvalues must be the matrix 1_n. Hence $A'^{-1}A = M'M^{-1} = 1_n$ and therefore $A = A'$ and $M = M'$. Thus we have shown that the Gram-Schmidt process defines a map $\mathrm{GL}_n(\mathbb{R}) \longrightarrow \mathrm{O}_n \times B_n^+$, $B \mapsto (A, M)$, which is clearly continuous and is the inverse of the map $\mathrm{O}_n \times B_n^+ \longrightarrow \mathrm{GL}_n(\mathbb{R})$ given by $(A, M) \mapsto AM$. □

The Iwasawa decomposition homeomorphism is not a homomorphism (isomorphism) (if $n > 1$), since the matrices in O_n do not commute with those of B_n^+.

Combining 2.4.16 and 2.4.17 one obtains the following decomposition.

2.4.18 **Corollary**. *There is a homeomorphism*

$$\mathrm{GL}_n(\mathbb{R}) \approx \mathrm{O}_n \times D_n^+ \times B_n^{(1)} .$$

□

The following result shows the role played by O_n in the Iwasawa decomposition.

2.4.19 **Theorem**. *The subgroup* $\mathrm{O}_n \subset \mathrm{GL}_n(\mathbb{R})$ *is a maximal compact subgroup, namely if* $K \subset \mathrm{GL}_n(\mathbb{R})$ *is a compact subgroup such that* $\mathrm{O}_n \subset K$*, then* $\mathrm{O}_n = K$*.*

Proof: If $K \subset \mathrm{GL}_n(\mathbb{R})$ is a compact subgroup such that $\mathrm{O}_n \subset K$, then $H = K \cap B_n^+$ is also a compact subgroup. By the Iwasawa decomposition 2.4.17, one has that $K = \mathrm{O}_n H$. We shall prove that every compact subgroup H of B_n^+ is the trivial group.

Let us thus take a compact subgroup $H \subset B_n^+$ and take $A \in H$. By the compactness of H, any eigenvalue of any element in H must lie in certain compact subset of $\mathbb{R}^* = \mathbb{R} - 0$, so that the eigenvalues of A and of all its powers must be in the same set. Therefore these eigenvalues must be 1 and hence $A \in B_n^{(1)}$.

Let us take now a minimal $i \in \{1, \dots, n\}$ such that if $i \leq j < k$, then $A_{jk} = 0$. We must show that $i = 1$. If we had $i > 1$, then there would be an index $k \geq i$ such that $A_{i-1\,k} \neq 0$. Hence for $C = A^m$ one would have that $C_{i-1\,k} = mA_{i-1\,k}$, which would contradict the compactness of H, because the entries of the matrices in H would build up an unbounded set.

2.4.20 EXERCISE. Show that any connected compact subgroup of $\mathrm{GL}_n(\mathbb{R})$ is a conjugate in $\mathrm{GL}_n(\mathbb{R})$ to a subgroup of O_n.

2.4.21 NOTE. According to Theorem 2.4.12, starting with any of the many (Lie) groups that we have defined, we can produce many different (smooth) manifolds by taking the quotients of the groups by their closed subgroups, the way we did to obtain the Stiefel and the Grassmann manifolds.

Chapter 3 Homotopy

Homotopy is one of the fundamental concepts in topology and is the base of algebraic topology. In this chapter, we shall introduce the elements of homotopy theory that we shall need to understand the next chapters.

3.1 The homotopy concept

In this section we introduce the basic concept of homotopy of maps between topological spaces. As usual, I will denote the unit interval $[0,1] \subset \mathbb{R}$.

3.1.1 Definition. Let X and Y be topological spaces. A *homotopy* from X to Y is a map

$$H : X \times I \longrightarrow Y \,.$$

If $f, g : X \longrightarrow Y$ are continuous maps, we shall say that they are *homotopic* if there exists a homotopy H from X to Y such that

$$\begin{aligned} H(x,0) &= f(x) \\ H(x,1) &= g(x) \,. \end{aligned}$$

We say that the homotopy H *starts* at f and *ends* at g, or that it is a homotopy *between* f and g. If we define $H_t : X \longrightarrow Y$ by $H_t(x) = H(x,t)$, $t \in I$, then H_t is continuous for each t and we may identify a homotopy H with the family $\{H_t\}$ of maps from X to Y parametrized by t. Indeed, the mapping $t \mapsto H_t$ determines a (continuous) path in the function space $\mathbf{M}(X,Y)$, of maps from X to Y endowed with the compact open topology.

If there is a homotopy H between f and g, we usually denotes this fact by $H : f \simeq g$.

3.1.2 Exercise. Prove the last assertion in Definition 3.1.1, namely, show that to have a homotopy $H : f \simeq g$ is equivalent to have a path, i.e., a continuous map $\omega : I \longrightarrow Y^X$ (cf. 4.1.1 below), where $\mathbf{M}(X,Y)$ is the space of continuous maps from X to Y with the compact-open topology (cf. [21]) and $\omega(t) = H_t$. This statement is equally true if instead of $\mathbf{M}(X,Y)$ we write Y^X, namely the space of maps from X to Y with the compactly generated topology associated to the compact-open topology.

3.1.3 **Proposition**. *The relation $\simeq$ is an equivalence relation in $\mathbf{M}(X,Y)$.*

Proof: If $f : X \longrightarrow Y$ is continuous, then the homotopy $H : X \times I \longrightarrow Y$ given by

$$H(x,t) = f(x)$$

shows that $f \simeq f$.

If $H : f \simeq g$, then $\overline{H} : g \simeq f$, where the homotopy $\overline{H}$ is given by

$$\overline{H}(x,t) = H(x, 1-t)\,.$$

Finally, if $H : f \simeq g$ and $K : g \simeq h$, then $H * K : f \simeq h$, where the homotopy $H * K$ is given by

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

□

The homotopy relation is compatible with compositions; namely, one has the following.

3.1.4 **Proposition**. *If $f \simeq g : X \longrightarrow Y$ and $f' \simeq g' : Y \longrightarrow Z$, then $f' \circ f \simeq g' \circ g : X \longrightarrow Z$.*

Proof: If $H : f \simeq g$ and $H' : f' \simeq g'$ are homotopies, then $K : f' \circ f \simeq g' \circ g$, where

$$K(x,t) = H'(H(x,t),t)\,.$$

□

3.1.5 NOTE. Alternatively, one may define $K : f' \circ f \simeq g' \circ g$ in the previous proof by

$$K(x,t) = \begin{cases} f'H(x,2t) & \text{if } 0 \leq t \leq \frac{1}{2}, \\ H'(g(x),2t-1) & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

That is, one proves first $f' \circ f \simeq f' \circ g$ and then $f' \circ g \simeq g' \circ g$, and one uses the transitivity of the relation $\simeq$.

The problem of deciding whether two maps are homotopic or not is, in general, a very difficult problem. To be able to provide a positive answer requires a good knowledge of the given spaces and of the given maps in order to give an explicit homotopy, or else, have other elements to answer the question. In general, the negative case is not so difficult whenever one counts with adequate tools for doing it. This is one of the fundamental aspects of algebraic topology. In the following chapters we shall develop methods to analyze this sort of questions.

3.1.6 Definition. We say that a map $f : X \longrightarrow Y$ is *nullhomotopic* if it is homotopic to the *constant map* $c_{y_0} : X \longrightarrow Y$ given by $c(x) = y_0 \in Y$ for every y. A homotopy $H : f \simeq c_{y_0}$ is called *nullhomotopy.* We say that a topological space X is *contractible* if the identity map $\mathrm{id}_X : X \longrightarrow X$ is nullhomotopic. A nullhomotopy $H : X \simeq c_{x_0}$ of id_X is called a *contraction* of X. If $H(x_0, t) = x_0$ for $t \in I$, namely, if the point x_0 remains fixed throughout the contraction, we say that X es *strongly contractible* to x_0.

Contractibility is a topological invariant, that is, if X is contractible, then any other space Y homeomorphic to X is contractible. We shall see below that in a certain sense, contractible spaces behave like points. Before doing that, we consider the following notation.

3.1.7 Definition. As we proved in 3.1.3, the homotopy relation is an equivalence relation. Given $f : X \longrightarrow Y$, we call the equivalence class $[f] = \{g : X \longrightarrow Y \mid g \simeq f\}$, *homotopy class* of $f : X \longrightarrow Y$. Moreover, we denote by $[X, Y]$ the set of homotopy classes of maps $f : X \longrightarrow Y$.

3.1.8 Remark. According to Exercise 3.1.2, $f \simeq g : X \longrightarrow Y$ if and only if f and g are connected by a path in Y^X. Therefore, the set of path components of this space coincides with that of homotopy classes, namely

$$\pi_0(Y^X) = [X, Y]\,.$$

3.1.9 Remark. If Y is path connected, then any two nullhomotopic maps $f, g : X \longrightarrow Y$ are homotopic. Namely, take nullhomotopies $F : f \simeq c_{y_0}$ and $G : g \simeq c_{y_1}$, and take a path $\sigma : y_0 \simeq y_1$ en Y. Then the homotopy $H : X \times I \longrightarrow Y$ given by

$$H(x,t) = \begin{cases} F(x, 3t) & \text{if } 0 \le t \le \frac{1}{3}, \\ \sigma(3t-1) & \text{if } \frac{1}{3} \le t \le \frac{2}{3}, \\ G(x, 3-3t) & \text{if } \frac{2}{3} \le t \le 1, \end{cases}$$

is a homotopy from f to g.

If Y is path connected, then the homotopy class of all nullhomotopic maps from X en Y is well defined and will be denoted by $0 \in [X, Y]$. In this case, we write the fact that a map $f : X \longrightarrow Y$ is nullhomotopic simply as $f \simeq 0$.

3.1.10 Examples.

(a) Any two maps $f, g : X \longrightarrow \mathbb{R}^n$ are homotopic; for example, through the *linear homotopy* $H(x, t) = (1-t)f(x) + tg(x)$ determined by the *line segments* that join the point $f(x)$ with the point $g(x)$ in $\mathbb{R}^n$ for each $x \in X$. Thus, $[X, \mathbb{R}^n] = 0$.

(b) Every map $f : \mathbb{R}^n \longrightarrow Y$ es nullhomotopic, for example, through the homotopy $H(x,t) = f((1-t)x)$, determined by the image in Y of the line segments from x to 0. Thus, if Y is path connected, $[\mathbb{R}^n, Y] = 0$.

(c) If $*$ is a singular space, then for any space Y, there is exactly one homotopy class in $[*, Y]$ for each path component of Y, since a homotopy of maps from $*$ to Y is the same thing as a path in Y, that is, $[*, Y]$ is in one-to-one correspondence with the path components of Y; in other words, $[*, Y] = \pi_0(Y)$.

(d) If $f : W \longrightarrow X$ and $g : Y \longrightarrow Z$ are continuous, and $H : X \times I \longrightarrow Y$ is a nullhomotopy, then $H \circ (f \times \mathrm{id}_I) : W \times I \longrightarrow Y$ and $g \circ H : X \times I \longrightarrow Z$ are nullhomotopies. Hence the composite (in any order) of an arbitrary map and a nullhomotopic map is always nullhomotopic.

(e) If X is contractible, then any maps $f : W \longrightarrow X$ and $g : X \longrightarrow Y$ are nullhomotopic, since $f = \mathrm{id}_X \circ f$ and $g = g \circ \mathrm{id}_X$ and id_X is nullhomotopic. Since, in particular, a contractible space is path connected (*exercise*), if X is contractible, then $[W, X] = 0$ and $[X, Y] = 0$ for any nonempty space W and any path-connected space Y.

(f) $\mathbb{B}^n$ and $\mathbb{R}^n$ are contractible through the *linear contractions* given by $(x,t) \mapsto (1-t)x$. Along with these spaces, all (closed) balls (disks) and all (open) cells are contractible. This, together with (a), puts (a) and (b) in a more general context.

(g) A subset $A \subset \mathbb{R}^n$ is said to be *star-like* if it contains a point x_0 such that for any other point $x \in A$ the line segment from x to x_0 lies inside A, that is, if $(1-t)x + tx_0 \in A$ for every $t \in I$. In particular, convex sets are star-like. Any star-like set A is contractible via the contraction $H : A \times I \longrightarrow A$ given by $H(x,t) = (1-t)x + tx_0$.

(h) Since we already know by the stereographic projection that for any point $x_0 \in \mathbb{S}^n$, $\mathbb{S}^n - x_0$ is homeomorphic to $\mathbb{R}^n$, or in other words, that $\mathbb{S}^n - x_0$ is a cell, we have that any continuous nonsurjective map $f : X \longrightarrow \mathbb{S}^n$ such that $x_0 \notin f(X)$, f factors through the inclusion $i : \mathbb{S}^n - x_0 \hookrightarrow \mathbb{S}^n$, that is nullhomotopic. Consequently, f is nullhomotopic. See 3.1.11 below for another proof of this fact.

(i) Consider the *infinite comb*, defined as the subspace of $\mathbb{R}^2$ given by

$$P = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1, 0 \le y \le 1, y > 0 \Leftrightarrow x = 0,\ \frac{1}{n},\ n \in \mathbb{N}\}\,.$$

(See Figure 3.1).

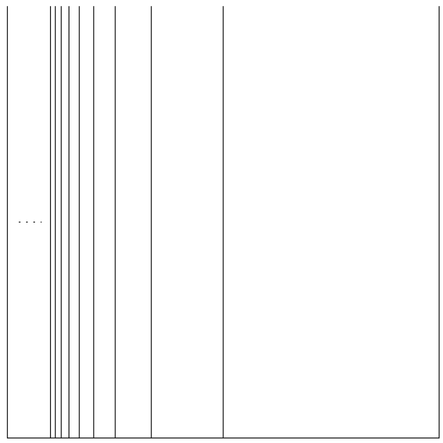


Figure 3.1 The infinite comb

The homotopy

$$H(x,y,t) = \begin{cases} (x,(1-2t)y) & \text{if } 0 \le t \le \frac{1}{2}, \\ ((2-2t)x,0) & \text{if } \frac{1}{2} \le t \le 1, \end{cases}$$

shows that P is strongly contractible to the point $(0,0)$. However, P is not strongly contractible to any point of the form $(0,y)$, $0 < y \le 1$; in particular, it is not strongly contractible to $(0,1)$, because it is not locally connected at this point, nor at any point of the form $(0,t)$, $t > 0$, (see Figure 3.2).

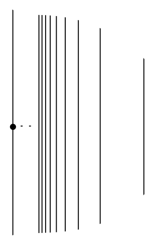


Figure 3.2 A neighborhood in the infinite comb

3.1.11 Remark. An alternative proof of statement (h) above is the following. Since the line segment that joins $-x_0$ with $f(x)$ in $\mathbb{R}^{n+1}$ does not contain 0, the homotopy $H : X \times I \longrightarrow \mathbb{S}^n$ given by

$$H(x,t) = \frac{(1-t)f(x) - tx_0}{|(1-t)f(x) - tx_0|}$$

is well defined, starts at f and ends at the constant map c_{-x_0}; namely, it is a nullhomotopy of f.

3.1.12 EXERCISE. Check that, indeed, the comb space P is not strongly contractible to the point $(0,1) \in P$.

3.1.13 EXERCISE. Prove that if $f, g : X \longrightarrow \mathbb{S}^n$ are maps such that for every $x \in X$, $f(x) \neq -g(x)$, then $f \simeq g$.

It is frequently convenient to have certain restriction on the homotopy concept. In what follows we analyze some cases.

3.1.14 DEFINITION. Take $f, g : X \longrightarrow Y$ and let $A \subset X$ be such that $f|_A = g|_A$. We say that f and g are *homotopic relative* to A if there exists a *relative homotopy* $H : X \times I \longrightarrow Y$, namely a homotopy such that $H(x,0) = f(x)$, $H(x,1) = g(x)$ and $H(a,t) = f(a) = g(a)$ for every $a \in A$. In other words, a homotopy between f and g which is *stationary* on A. This fact is denoted by $H : f \simeq g$ rel A. Relative homotopy, such as it is the case with the *free* homotopy, is an equivalence relation.

3.1.15 REMARK. If $f, g : X \longrightarrow Y$ are continuous and $A \subset X$ is such that $f|_A = g|_A$, then $f \simeq g$ rel $A \Longrightarrow f \simeq g$. However, the converse of this statement is false, as we easily infer from 3.1.10(i).

Another frequent concept which is also useful in homotopy theory is the following.

3.1.16 DEFINITION. Recall that a topological space X together with a subspace A is called a *pair of spaces* and is denoted by (X, A). A *map of pairs*, denoted by $f : (X, A) \longrightarrow (Y, B)$, is a map $f : X \longrightarrow Y$ such that $f(A) \subset B$. A *homotopy of pairs* is a map of pairs such that $H(a,t) \in B$ for every $a \in A$, $t \in I$. If $f(x) = H(x,0)$ and $g(x) = H(x,1)$, $f, g : (X, A) \longrightarrow (Y, B)$, then we say that H is a homotopy of pairs between f and g; in symbols, $H : f \simeq g : (X, A) \longrightarrow (Y, B)$. So as the free homotopy, homotopy of pairs is an equivalence relation, and similarly to 3.1.7, we denote the class $\{g \mid f \simeq g : (X, A) \longrightarrow (Y, B)\}$ by $[f]$ and the set of these classes by $[X, A; Y, B]$. If $A = \emptyset = B$, then this *homotopy set* coincides with $[X, Y]$, but in general, if $f : (X, A) \longrightarrow (Y, B)$, then its class $[f] \in [X, A; Y, B]$ differs from its class $[f] \in [X, Y]$, as we shall see.

3.1.17 EXERCISE. If (X_1, A_1) and (X_2, A_2) are pairs of spaces, then their *product* $(X_1, A_1) \times (X_2, A_2)$ is the pair $(X_1 \times X_2, A_1 \times X_2 \cup X_1 \times A_2)$. Prove the following statements.

(a) The product of pairs is compatible with the identification of a space X with the pair $(X, \emptyset)$.

(b) If $f_1 : (X_1, A_1) \longrightarrow (Y_1, B_1)$ and $f_2 : (X_2, A_2) \longrightarrow (Y_2, B_2)$ are maps of pairs, then $f_1 \times f_2 : (X_1, A_1) \times (X_2, A_2) \longrightarrow (Y_1, B_1) \times (Y_2, B_2)$ is a map of pairs.

3.1.18 Exercise. A *homeomorphism of pairs* $\varphi : (X, A) \longrightarrow (Y, B)$ is a homeomorphism $\varphi : X \longrightarrow Y$ such that $\varphi(A) = B$.

(a) Prove that there is a homeomorphism of pairs

$$\alpha : (I \times I, (\{0\} \times I) \cup (I \times \{1\}) \cup (\{1\} \times I)) \longrightarrow (I \times I, \{1\} \times I)\,.$$

(*Hint:* Take

$$\alpha(s,t) = \begin{cases} (1-3s, \frac{t}{3}) & \text{if } 0 \leq s \leq \frac{1-t}{3}, \\ (t, \frac{3s-2st+2t-1}{2t+1}) & \text{if } \frac{1-t}{3} \leq s \leq \frac{2+t}{3}, \\ (3s-2, \frac{3-t}{3}) & \text{if } \frac{2+t}{3} \leq s \leq 1.) \end{cases}$$

(b) Let (Z, X) be a pair of spaces and take $W = (Z \times \{0\}) \cup (X \times I) \cup (Z \times \{1\}) \subset Z \times I$. Prove that there exists a homeomorphism of pairs

$$\begin{aligned} \varphi : (Z \times I \times I, (Z \times I \times \{1\}) \cup (W \times I)) \\ \longrightarrow (Z \times I \times I, (Z \times \{1\} \times I) \cup (X \times I \times I)) \\ = (Z \times I, (Z \times \{1\}) \cup (X \times I)) \times I\,. \end{aligned}$$

(*Hint:* The pair on the left hand side is the product of pairs $(Z, X) \times (I \times I, (\{0\} \times I) \cup (I \times \{1\}) \cup (\{1\} \times I))$, while that on the right hand side is the product $(Z, X) \times (I \times I, \{1\} \times I)$. The desired homeomorphism φ is then $\mathrm{id}_Z \times \alpha$, where α is as in (a).)

The following exercise will be used below.

3.1.19 Exercise. Prove that if there exists a retraction $\sigma : Z \times I \longrightarrow (Z \times \{1\}) \cup (X \times I)$, then there also exists a retraction $\rho : Z \times I \times I \longrightarrow (Z \times I \times \{1\}) \cup (W \times I)$, where $W = (Z \times \{0\}) \cup (X \times I) \cup (Z \times \{1\}) \subset Z \times I$. (*Hint:* Let $\rho = \varphi^{-1} \circ (\sigma \times \mathrm{id}_I) \circ \varphi$, where φ is as in the previous exercise 3.1.18.)

What the previous exercise asserts, in more technical terms, is that if the inclusion $X \hookrightarrow Z$ is a *cofibration*, then the inclusion $W \hookrightarrow Z \times I$ is also a cofibration (see 3.4.8 below or [2, 4.1]).

3.1.20 Example. Since the unit interval I is contractible, $[I, I] = 0$. However, the set $[I, \partial I; I, \partial I]$ consists of four elements; namely the homotopy class of the identity map id_I, that of the constant map c_0 with value 0, that of the constant map c_1 with value 1, and that of the map $\overline{\mathrm{id}_I}$ given by $\overline{\mathrm{id}_I}(t) = 1 - t$. Indeed, if $f \simeq g : (I, \partial I) \longrightarrow (I, \partial I)$, then $f|_{\partial I} = g|_{\partial I}$, that is, $f(0) = g(0)(= 0 \text{ or } 1)$ and $f(1) = g(1)(= 0 \text{ or } 1)$. Given $f : (I, \partial I) \longrightarrow (I, \partial I)$, the homotopy $H(s,t) = (1-t)f(s) + t[(1-s)f(0) + sf(1)]$ is of pairs; it starts at f and ends at c_0, id_I, c_1 or $\overline{\mathrm{id}_I}$, according to whether $f|_{\partial I} = c_0$, $\mathrm{id}_{\partial I}$, c_1, or $\overline{\mathrm{id}_{\partial I}}$. Therefore, $[f] = [g]$ if and only if $f|_{\partial I} = g|_{\partial I}$.

3.1.21 Remark. If $f \simeq g : (X, A) \longrightarrow (Y, B)$, then $f \simeq g : X \longrightarrow Y$ and $f|_A \simeq g|_A : A \longrightarrow B$. However, the converse statement is not true (see 3.3.8).

3.1.22 Exercise. Prove that X and Y are contractible spaces if and only if the topological product $X \times Y$ is a contractible space.

3.1.23 Exercise. Recall that the diagonal map $d_X : X \longrightarrow X \times X$ is given by $d_X(x) = (x, x)$. Prove that d_X is nullhomotopic if and only if X is contractible. More generally, take $f : X \longrightarrow Y$ and let $d_f : X \longrightarrow X \times Y$ be the *graph map* of f, that is, $d_f(x) = (x, f(x))$; prove that d_f is nullhomotopic if and only if X is contractible.

3.1.24 Exercise. Prove that for every space X, its cone $CX = X \times I/X \times \{1\}$ is contractible.

A special case of pairs of spaces is the following, where A and B are singular subspaces.

3.1.25 Definition. Given a space X and a point $x_0 \in X$, the pair (X, x_0) is called *pointed space* and the point x_0, *base point* of the pointed space. A map of pairs $f : (X, x_0) \longrightarrow (Y, y_0)$, that is, a map such that $f(x_0) = y_0$, is called *pointed map*, and so also a homotopy of pairs $H : (X, x_0) \times I \longrightarrow (Y, y_0)$, namely a homotopy such that $H(x_0, t) = y_0$, $t \in I$, is called a *pointed homotopy*.

In this case we may alternatively denote the homotopy set $[X, x_0; Y, y_0]$ by $[X, Y]_*$ if the base points are obvious or their names are not important.

3.1.26 Exercise. If we consider $\mathbb{S}^0 = \{-1, 1\}$ as a pointed space with base point 1, prove that there is a bijection

$$\pi_0(X) \cong [\mathbb{S}^0, X]_*\,,$$

if X is considered as a pointed space with any point $x_0 \in X$ as base point. (*Hint:* By 3.1.10, $\pi_0(X) \cong [P, X]$, where P is a singular space; prove hence, that $[P, X] \cong [\mathbb{S}^0, 1; X, x_0]$.)

3.1.27 Note. The pairs of spaces together with the maps of pairs build a *category* in the sense that id_X determines a map of pairs $\mathrm{id}_{(X,A)} : (X, A) \longrightarrow (X, A)$ for every $A \subset X$, and if $f : (X, A) \longrightarrow (Y, B)$, $g : (Y, B) \longrightarrow (Z, C)$ are maps of pairs, then the composite $g \circ f : (X, A) \longrightarrow (Z, C)$ is a map of pairs. In this case we refer to the *category of pairs of spaces.* In particular, if we restrict the pairs of spaces to pointed spaces, we shall be dealing with the *category of pointed spaces.*

Moreover, we can associate to every pair of spaces (X, A) a pointed space $(X/A, \{A\})$, where $\{A\}$ represents the point to which A collapses in the quotient space X/A. Thus, a map of pairs $f : (X, A) \longrightarrow (Y, B)$ determines a map of pointed spaces $\overline{f} : (X/A, \{A\}) \longrightarrow (Y/B, \{B\})$ in a *functorial* way, that is, such that if $f = \mathrm{id}_{(X,A)}$, then $\overline{f} = \mathrm{id}_{(X/A,\{A\})}$, and if $f : (X, A) \longrightarrow (Y, B)$ and $g : (Y, B) \longrightarrow (Z, C)$ are maps of pairs, then

$$\overline{g \circ f} = \overline{g} \circ \overline{f} : (X/A, \{A\}) \longrightarrow (Z/C, \{C\})\,.$$

In other words, the assignment

$$\begin{tikzcd} (X,A) \arrow[d, "f"'] & (X/A,\{A\}) \arrow[d, "\overline{f}"] \\ (Y,B) \arrow[r, phantom, "\longmapsto"] & (Y/B,\{B\}) \end{tikzcd}$$

is a *functor* from the category $\mathcal{T}op_2$ of pairs of spaces and continuous maps of pairs to the category $\mathcal{T}op_*$ of pointed spaces and continuous pointed maps.

A space X can also be seen as a pair if we take the pair of spaces $(X, \emptyset)$, so that the assignment

$$\begin{tikzcd} X \arrow[d, "f"'] & (X,\emptyset) \arrow[d, "f"] \\ Y \arrow[r, phantom, "\longmapsto"] & (Y,\emptyset) \end{tikzcd}$$

is also a functor, now from the category $\mathcal{T}op$ of topological spaces and continuous maps to $\mathcal{T}op_2$.

If we recall that $X/\emptyset = X^+ = X \sqcup \{x_0\}$, then we may compose the functors described above to obtain a functorial way of associating to any topological space X a pointed space; namely (X^+, x_0). Thus, we have that

$$\begin{tikzcd} X \arrow[d, "f"'] & (X^+,x_0) \arrow[d, "f^+"] \\ Y \arrow[r, phantom, "\longmapsto"] & (Y^+,y_0), \end{tikzcd}$$

where $f^+|_X = f$ and $f^+(x_0) = y_0$, is a functor from $\mathcal{T}op$ to $\mathcal{T}op_*$.

Similarly, given any pair of spaces (X, A), we may associate to it either the space X or the space A and so we have that

$$\begin{tikzcd} (X,A) \arrow[d, "f"'] & \longmapsto & X \arrow[d, "f"] \\ (Y,B) & & Y \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} (X,A) \arrow[d, "f"'] & \longmapsto & A \arrow[d, "f|_A"] \\ (Y,B) & & B \end{tikzcd}$$

are functors from $\mathcal{T}op_2$ to $\mathcal{T}op$.

Since we frequently define new spaces through identifications, it is quite convenient to know that if the homotopies are compatible with the identifications, then they determine homotopies in the quotient spaces.

3.1.28 **Proposition**. *Let $q : X \longrightarrow \overline{X}$ be an identification and let $H : X \times I \longrightarrow Y$ be a homotopy compatible with q, that is, such that if $q(x_1) = q(x_2)$, then also $H(x_1, t) = H(x_2, t)$, for all $t \in I$. Then H determines a homotopy $\overline{H} : \overline{X} \times I \longrightarrow Y$ such that $\overline{H}(q(x), t) = H(x, t)$, $x \in X$, $t \in I$.*

Proof: Since I is (locally) compact, the map $q \times \mathrm{id}_I : X \times I \longrightarrow \overline{X} \times I$ is an identification and, since H is compatible with it, H determines $\overline{H}$ as desired. $\square$

An immediate consequence of this fact is the following corollary.

3.1.29 **Corollary**. *Given equivalence relations in the spaces X and Y, both denoted, for simplicity, by the same symbol $\sim$, take a homotopy $H : X \times I \longrightarrow Y$ such that if $x_1 \sim x_2$ in X, then $H(x_1, t) \sim H(x_2, t)$ in Y, $t \in I$. Then H determines a homotopy $\overline{H} : (X/\sim) \times I \longrightarrow Y/\sim$ such that if $\overline{x} \in X/\sim$ denotes the class of $x \in X$ and $\overline{y} \in Y/\sim$ denotes the class of $y \in Y$, then $\overline{H}(\overline{x}, t) = \overline{H(x, t)}$, $x \in X$, $t \in I$.* $\square$

There are other interesting consequences of 3.1.28. For instance, we have the following.

3.1.30 **Proposition**. *Given an attaching situation; that is, a diagram $X \supset A \xrightarrow{f} Y$, and homotopies $H : X \times I \longrightarrow Z$, $K : Y \times I \longrightarrow Z$ such that if $a \in A$, $H(a, t) = K(f(a), t) \in Z$, $t \in I$, then the map $(H, K) : (X \sqcup Y) \times I \longrightarrow Z$ such that $(H, K)|_{X \times I} = H$, $(H, K)|_{Y \times I} = K$, induces a homotopy $\langle H, K \rangle : (Y \cup_f X) \times I \longrightarrow Z$.* $\square$

3.1.31 Note. Congruently with 3.1.27, we have that the construction of the homotopy sets $[X,Y]$ (resp. $[X,A;Y,B]$, $[X,x_0;Y,y_0]$) are functorial in the following sense. Given a map $\varphi : Y \longrightarrow Z$ (resp. $\varphi : (Y,B) \longrightarrow (Z,C)$, $\varphi : (Y,y_0) \longrightarrow (Z,z_0)$), there is a function

$$\varphi_* : [X,Y] \longrightarrow [X,Z]$$

(resp. $\varphi_* : [X,A;Y,B] \longrightarrow [X,A;Z,C]$, $\varphi_* : [X,x_0;Y,y_0] \longrightarrow [X,x_0;Z,z_0]$) defined by $\varphi_*([f]) = [\varphi \circ f]$.

Analogously, given a map $\psi : Z \longrightarrow X$ (resp. $\psi : (Z,C) \longrightarrow (X,A)$, $\psi : (Z,z_0) \longrightarrow (X,x_0)$), we obtain a function

$$\psi^* : [X,Y] \longrightarrow [Z,Y]$$

(resp.

$$\psi^* : [X,A;Y,B] \longrightarrow [Z,C;Y,B],$$
$$\psi^* : [X,x_0;Y,y_0] \longrightarrow [Z,z_0;Y,y_0])$$

defined by $\psi^*([f]) = [f \circ \psi]$.

These assignments are functorial, since $(\mathrm{id}_Y)_* = 1_{[X,Y]}$ (resp. $(\mathrm{id}_{(Y,B)})_* = 1_{[X,A;Y,B]}$, $(\mathrm{id}_{(Y,y_0)})_* = 1_{[X,x_0;Y,y_0]}$) and if $\varphi : Y \longrightarrow Z$ and $\gamma : Z \longrightarrow W$ (resp. $\varphi : (Y,B) \longrightarrow (Z,C)$, and $\gamma : (Z,C) \longrightarrow (W,D)$, $\varphi : (Y,y_0) \longrightarrow (Z,z_0)$ and $\gamma : (Z,z_0) \longrightarrow (W,w_0)$), then $(\gamma \circ \varphi)_* = \gamma_* \circ \varphi_* : [X,Y] \longrightarrow [X,W]$ (resp. $(\gamma \circ \varphi)_* = \gamma_* \circ \varphi_* : [X,A;Y,B] \longrightarrow [X,A;W,D]$, $(\gamma \circ \varphi)_* = \gamma_* \circ \varphi_* : [X,x_0;Y,y_0] \longrightarrow [X,x_0;W,w_0]$); thus,

$$\begin{tikzcd} Y \arrow[d, "\varphi"'] & {[X,Y]} \arrow[d, "\varphi_*"] \\ Z \arrow[r, phantom, "\longmapsto", shift left=4ex] & {[X,Z]} \end{tikzcd}$$

is a *covariant* functor, namely a functor such that the induced arrows point in the same direction, from the category $\mathcal{T}op$, (resp. $\mathcal{T}op_2$), to the category $\mathcal{S}et$ of sets and functions. On the other hand, $(\mathrm{id}_X)^* = 1_{[X,Y]}$ (resp. $(\mathrm{id}_{(X,A)})^* = 1_{[X,A;Y,B]}$, $(\mathrm{id}_{(X,x_0)})^* = 1_{[X,x_0;Y,y_0]}$), and if $\psi : Z \longrightarrow X$ and $\lambda : W \longrightarrow Z$ (resp. $\psi : (Z,C) \longrightarrow (X,A)$ and $\lambda : (W,D) \longrightarrow (Z,C)$, $\psi : (Z,z_0) \longrightarrow (X,x_0)$ and $\lambda : (W,w_0) \longrightarrow (Z,z_0)$), then $(\psi \circ \lambda)^* = \lambda^* \circ \psi^* : [X,Y] \longrightarrow [W,Y]$ (resp. $(\psi \circ \lambda)^* = \lambda^* \circ \psi^* : [X,A;Y,B] \longrightarrow [W,D;Y,B]$, $(\psi \circ \lambda)^* = \lambda^* \circ \psi^* : [X,x_0;Y,y_0] \longrightarrow [W,w_0;Y,y_0]$); thus,

$$\begin{tikzcd} Z \arrow[d, "\pi"'] & {[Z,Y]} \\ X \arrow[r, phantom, "\longmapsto", shift left=4ex] & {[X,Y]} \arrow[u, "\psi^*"'] \end{tikzcd}$$

is a *contravariant* functor; that is, a functor such that the induced arrows point in the opposite direction, from $\mathcal{T}op$, (resp. $\mathcal{T}op_2$, $\mathcal{T}op_*$), to $\mathcal{S}et$.

3.1.32 EXERCISE. Recall Exercise 3.1.26, where one has to show that

$$\pi_0(X) \cong [\mathbb{S}^0, X]_* .$$

Prove the following:

(a) The correspondences

$$X \longmapsto \pi_0(X) \quad \text{and} \quad X \longmapsto [\mathbb{S}^0, X]_*$$

are both functors from the category $\mathcal{T}op_*$ to the category $\mathcal{S}et$.

(b) The bijection is *natural*; namely, for any pointed map $f : X \longrightarrow Y$ the diagram

$$\begin{tikzcd} \pi_0(X) \arrow[r, "\cong"] \arrow[d, "f_*"'] & {[\mathbb{S}^0, X]_*} \arrow[d, "f_*"] \\ \pi_0(Y) \arrow[r, "\cong"'] & {[\mathbb{S}^0, Y]_*}, \end{tikzcd}$$

is commutative, where the vertical arrow f_* on the left-hand side sends the path component of X corresponding to a point x to the path component of Y corresponding to the point $f(x)$, while f_* on the right-hand side sends the class $[\gamma]$ to the class $[f \circ \gamma]$, for any pointed map $\gamma : \mathbb{S}^0 \longrightarrow X$.

Thanks to Proposition 3.1.28, we see that a homotopy of pairs $H : (X, A) \times I \longrightarrow (Y, B)$ defines a homotopy of pointed spaces $\overline{H} : (X/A, \{A\}) \times I \longrightarrow (Y/B, \{B\})$, so that if $f \simeq g : (X, A) \longrightarrow (Y, B)$, then $\overline{f} \simeq \overline{g} : (X/A, \{A\}) \longrightarrow (Y/B, \{B\})$. In other words, we have the following.

3.1.33 **Proposition**. *There is a natural function*

$$[X, A; Y, B] \longrightarrow [X/A, \{A\}; Y/B, \{B\}]$$

given by $[f] \mapsto [\overline{f}]$. □

3.1.34 EXERCISE. Prove that, in fact, the function in the preceding proposition is *natural* in both (X, A) and $Y, B)$; namely, show that if $\varphi : (X, A) \longrightarrow (X', A')$ and $\psi : (Y, B) \longrightarrow (Y', B')$ are maps of pairs, then the diagrams

$$\begin{tikzcd} {[X, A; Y, B]} \arrow[r] & {[X/A, \{A\}; Y/B, \{B\}]} \\ {[X', A'; Y, B]} \arrow[r] \arrow[u, "\varphi^*"] & {[X'/A', \{A'\}; Y/B, \{B\}]} \arrow[u, "\overline{\varphi}^*"'] \end{tikzcd}$$

and

$$\begin{tikzcd}
{[X,A;Y,B]} \arrow[r] \arrow[d, "\psi_*"'] & {[X/A,\{A\};Y/B,\{B\}]} \arrow[d, "\overline{\psi}_*"] \\
{[X,A;Y',B']} \arrow[r] & {[X/A,\{A\};Y'/B',\{B'\}]}
\end{tikzcd}$$

are commutative.

3.1.35 Exercise. Given two homeomorphisms $f, g : X \longrightarrow Y$, we say that they are *isotopic* if there exists an *isotopy* $H : f \simeq g$; that is, a homotopy such that for every $t \in I$, $H_t : X \longrightarrow Y$ is a homeomorphism. Prove that every *rotation* $r : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$, that is, a map r given by $r(\mathrm{e}^{2\pi \mathrm{i}t}) = \mathrm{e}^{2\pi\mathrm{i}(t_0+t)}$, is isotopic to $\mathrm{id}_{\mathbb{S}^1}$.

3.1.36 Exercise. Given continuous $f, g : X \longrightarrow \mathbb{S}^1$, we define $f \cdot g : X \longrightarrow \mathbb{S}^1$ simply by $(f \cdot g)(x) = f(x)g(x)$, with the usual multiplication of complex numbers. Prove that the multiplication $[f] \cdot [g] = [f \cdot g]$ is well defined and equips $[X, \mathbb{S}^1]$ with the structure of an abelian group. For "nice" spaces X, this abelian group is the so-called *first cohomology group* of X and is usually denoted by $H^1(X)$ (see [2, §7.1]).

3.2 Homotopy of mappings of the circle into itself

Up to this point we have not seen any explicit maps that are not nullhomotopic. In this section we shall study from the homotopical viewpoint the first example of nontrivial maps. The mappings of the circle into itself will not only be the first example, but, in a sense they provide us with a fundamental example of this. We shall follow here the very convenient approach of Stöcker–Zieschang [23].

Recall that the points of the circle $\mathbb{S}^1 \subset \mathbb{C}$ have the form $\mathrm{e}^{2\pi\mathrm{i}t}$. Let $q : I \longrightarrow \mathbb{S}^1$ be the identification such that $q(t) = \mathrm{e}^{2\pi\mathrm{i}t}$.

Let $\varphi : I \longrightarrow \mathbb{R}$ be a continuous pointed function, that is, a function such that $\varphi(0) = 0$, that also satisfies $\varphi(1) = n \in \mathbb{Z}$. The map $I \longrightarrow \mathbb{S}^1$ given by $t \mapsto \mathrm{e}^{2\pi\mathrm{i}\varphi(t)}$, is compatible with the identification q. Hence it determines a pointed map

$$\widehat{\varphi} : \mathbb{S}^1 \longrightarrow \mathbb{S}^1 ,$$

that is, a map such that $\widehat{\varphi}(1) = 1$, given by $\widehat{\varphi}(\mathrm{e}^{2\pi\mathrm{i}t}) = \mathrm{e}^{2\pi\mathrm{i}\varphi(t)}$. Therefore, one has a commutative diagram

$$\begin{tikzcd}
I \arrow[r, "\varphi"] \arrow[d, "q"'] & \mathbb{R} \arrow[d] \\
\mathbb{S}^1 \arrow[r, dashed, "\widehat{\varphi}"'] & \mathbb{S}^1 .
\end{tikzcd}$$

We might say, in plain words, that the values of the map φ run along the interval $[0, n]$ (since we start from 0 and arrive at n) in one time unit, that is, while letting the argument of the function run along the interval $[0, 1]$. Therefore, the map $\widehat{\varphi}$ is such that while its argument runs around $\mathbb{S}^1$ once, starting at 1 and returning to 1, its value runs around $\mathbb{S}^1$ n times, also starting at 1 and returning to 1. In other words, after one turn of the argument, there are n turns of the value of $\widehat{\varphi}$. More precisely, this number n counts n counterclockwise turns if $n > 0$, and $-n$ clockwise turns if $n < 0$. We shall prove in what follows that any mapping $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ coincides with $\widehat{\varphi}$ for some $\varphi : I \longrightarrow \mathbb{R}$, that is, that one can "unwind" the mapping.

3.2.1 **Proposition**. *Given any pointed map $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$, that is, a map such that $f(1) = 1$, there exists a unique pointed function $\varphi : I \longrightarrow \mathbb{R}$, that is, with $\varphi(0) = 0$, such that $f(\zeta) = \widehat{\varphi}(\zeta)$, $\zeta \in \mathbb{S}^1$.*

Proof: The function is unique, since if $\varphi, \psi : (I, 0) \longrightarrow (\mathbb{R}, 0)$ are such that $\widehat{\varphi} = \widehat{\psi} : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$, that is, if they are such that $e^{2\pi i\varphi(t)} = e^{2\pi i\psi(t)}$, then $\psi(t) - \varphi(t) \in \mathbb{Z}$ for all $t \in I$. Therefore, since the function $I \longrightarrow \mathbb{Z}$ given by $t \mapsto \psi(t) - \varphi(t)$ is continuous, and since I is connnected and $\mathbb{Z}$ discrete, it follows that this function has to be constant. Moreover, since $\psi(0) - \varphi(0) = 0 - 0 = 0$, then $\psi = \varphi$.

Let us see now that φ exists. We need a mapping φ such that $\varphi(0) = 0$ and such that $f(e^{2\pi it}) = e^{2\pi i\varphi(t)}$. To that end, let us take the main branch, log, of the complex logarithm; namely, if $z = re^{i\alpha} \in \mathbb{C}$, $r > 0$, $-\pi < \alpha < \pi$, then $\log(z) = \ln(r) + i\alpha$, where ln is the natural logarithm function. Let $h : I \longrightarrow \mathbb{S}^1$ be such that $h(t) = f(e^{2\pi it})$. Since I is compact, h is uniformly continuous, and so there exists a partition $0 = t_0 < t_1 < \cdots < t_k = 1$ of I such that

$$|h(t) - h(t_j)| < 2 \quad \text{if} \quad t \in [t_j, t_{j+1}] \quad \text{and} \quad j = 0, 1, \ldots, k-1\,.$$

Hence $h(t) \neq -h(t_j)$, that is, $h(t) \cdot h(t_j)^{-1} \neq -1$. Therefore, $\log(h(t) \cdot h(t_j)^{-1})$ is well defined. The desired function is thus the following. If $t \in [t_j, t_{j+1}]$, take

$$\varphi(t) = \frac{1}{2\pi i}\left(\log\left(\frac{h(t_1)}{h(t_0)}\right) + \cdots + \log\left(\frac{h(t_j)}{h(t_{j-1})}\right) + \log\left(\frac{h(t)}{h(t_j)}\right)\right).$$

Then φ is well defined, continuous, and has real values. Using the exponential law $e^{a+b} = e^a e^b$, and $e^{\log(z)} = z$, since $h(t_0) = h(0) = 1$, one gets

$$e^{2\pi i\varphi(t)} = \frac{h(t)}{h(t_0)} = h(t) = f(e^{2\pi it})\,.$$

□

As a consequence of this last proposition, we obtain the fundamental result of this section.

3.2.2 **Theorem**. *Given any mapping $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$, there exists a unique pointed function $\varphi : I \longrightarrow \mathbb{R}$ such that $f(\zeta) = f(1) \cdot \widehat{\varphi}(\zeta)$, $\zeta \in \mathbb{S}^1$ (where the dot here means the complex product).*

Proof: Let $g : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ be given by $g(\zeta) = f(1)^{-1} \cdot f(\zeta)$. Then $g(1) = 1$, and therefore by 3.2.1, there exists a unique pointed function $\varphi : I \longrightarrow \mathbb{R}$ such that $g(\zeta) = \widehat{\varphi}(\zeta)$. Hence, $f(\zeta) = f(1) \cdot \widehat{\varphi}(\zeta)$. □

Given a function $\varphi : I \longrightarrow \mathbb{R}$ such that $\varphi(0) = 0$ and $\varphi(1) = n \in \mathbb{Z}$, then $\varphi \simeq \varphi_n$ rel $\{0, 1\}$ for $\varphi_n : I \longrightarrow \mathbb{R}$ given by $\varphi_n(s) = ns$, since $H : I \times I \longrightarrow \mathbb{R}$ defined by

$$H(s,t) = (1-t)\varphi(s) + nst$$

is a homotopy relative to $\{0, 1\}$ between φ and φ_n. Applying the exponential mapping to both φ and φ_n, we obtain the following result.

3.2.3 **Lemma**. *Let $\varphi : I \longrightarrow \mathbb{R}$ satisfy $\varphi(0) = 0$ and $\varphi(1) = n \in \mathbb{Z}$. Then $\widehat{\varphi} \simeq \widehat{\varphi_n}$ rel $\{1\}$.* □

Given a mapping $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$, we have by Theorem 3.2.2 that $f = f(1) \cdot \widehat{\varphi}$, that is, f is the result of composing a map of type $\widehat{\varphi}$ with a rotation as was defined in Exercise 3.1.35. By this exercise, we know that a rotation is homotopic to the identity map $\mathrm{id}_{\mathbb{S}^1}$; therefore, $f \simeq \widehat{\varphi}$, for some $\varphi : I \longrightarrow \mathbb{R}$ such that $\varphi(0) = 0$ and $\varphi(1) = n \in \mathbb{Z}$. By 3.2.3, we have the following.

3.2.4 **Proposition**. *Given $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$, then there exists a unique $n \in \mathbb{Z}$ such that $f \simeq \widehat{\varphi_n} : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$.*

We have the following definition.

3.2.5 Definition. Let $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ be continuous and let $\varphi : I \longrightarrow \mathbb{R}$ be the unique function that by 3.2.2 exists and is such that $f(\zeta) = f(1) \cdot \widehat{\varphi}(\zeta)$. Since the integer $\varphi(1) = n$ is well defined, we define the *degree* of f as this integer n and denote it by $\deg(f)$.

It is geometrically clear what is meant by $\deg(f)$, since by 3.2.2 this integer indicates how many times $f(\zeta)$ turns around $\mathbb{S}^1$ when ζ turns once around $\mathbb{S}^1$. This motion of $f(\zeta)$ is counterclockwise if $n > 0$ and clockwise if $n < 0$, while if $n = 0$, it means that $f \simeq c_0$, that is, the total number of turns is 0.

We observe that $\deg(f)$ depends only on the homotopy class of f; namely, we have the following.

3.2.6 **Lemma**. *If* $f \simeq g : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$*, then* $\deg(f) = \deg(g)$.

Proof: Let $H : \mathbb{S}^1 \times I \longrightarrow \mathbb{S}^1$ be a homotopy such that $H(\zeta, 0) = f(\zeta)$ and $H(\zeta, 1) = g(\zeta)$, and let $f_s : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ be given by $f_s(\zeta) = H(\zeta, s)$. By 3.2.2, there exists a unique continuous function $\varphi_s : I \longrightarrow \mathbb{R}$ such that $\varphi_s(0) = 0$, $\varphi_s(1) \in \mathbb{Z}$, and $f_s(\zeta) = f_s(1) \cdot \widehat{\varphi_s}(\zeta)$. We shall see that the mapping $I \times I \longrightarrow \mathbb{R}$ given by $(t, s) \mapsto \varphi_s(t)$ is a homotopy; that is, it is continuous. As in the proof of Proposition 3.2.2, the map $h : I \times I \longrightarrow \mathbb{S}^1$ given by $(s, t) \mapsto h(s, t) = f_s(\mathrm{e}^{2\pi\mathrm{i}t})$ is uniformly continuous, and hence one can choose the partition of I in the proof of that proposition in such a way that

$$|h(s,t) - h(s,t_j)| < 2 \quad \text{if} \quad s \in I,\ t \in [t_j, t_{j+1}] \quad \text{and} \quad j = 0, 1, \ldots, k-1\,.$$

As before, one can now define φ_s with the same formula, but inserting in it the map $h_s : t \longrightarrow h(s, t)$ instead of h; that is, if $s \in I$ and $t \in [t_j, t_{j+1}]$, then

$$\varphi_s(t) = \frac{1}{2\pi\mathrm{i}} \left(\log\left(\frac{h(s,t_1)}{h(s,t_0)}\right) + \cdots + \log\left(\frac{h(s,t_j)}{h(s,t_{j-1})}\right) + \log\left(\frac{h(s,t)}{h(s,t_j)}\right) \right).$$

Therefore, $\varphi_s(t)$ is continuous as a function of s and of t; in particular, the function $s \mapsto \varphi_s(1)$ is continuous, and since $\varphi_s(1) \in \mathbb{Z}$, it has to be constant. Since $f(\zeta) = f(1) \cdot \widehat{\varphi_0}(\zeta)$ and $g(\zeta) = g(1) \cdot \widehat{\varphi_1}(\zeta)$, we obtain that $\deg(f) = \varphi_0(1) = \varphi_1(1) = \deg(g)$. ☐

Thus, the degree determines a function $[\mathbb{S}^1, \mathbb{S}^1] \longrightarrow \mathbb{Z}$. The fundamental result in this section, which shows us how an invariant is used for classification problems, is the following.

3.2.7 **Theorem**. *The function*

$$[\mathbb{S}^1, \mathbb{S}^1] \longrightarrow \mathbb{Z} \quad \textit{given by} \quad [f] \mapsto \deg(f)\,,$$

is well defined and bijective. More precisely, one has the following

(a) *If* $n \in \mathbb{Z}$*, then the map* $g_n : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ *given by* $g_n(\zeta) = \zeta^n$ *is such that* $\deg(g_n) = n$.

(b) *Take* $f, g : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$*. Then* $f \simeq g$ *if and only if* $\deg(f) = \deg(g)$.

Proof: (a) Since $g_n(\mathrm{e}^{2\pi\mathrm{i}t}) = \mathrm{e}^{2\pi int}$, we have that $g_n = \widehat{\varphi_n}$; thus $\deg(f) = \varphi_n(1) = n$.

(b) By 3.2.6, if $f \simeq g$, then $\deg(f) = \deg(g)$.

Conversely, if $\deg(f) = \deg(g) = n$, then we have that $f(\zeta) = f(1) \cdot \widehat{\varphi}(\zeta)$ and $g(\zeta) = g(1) \cdot \widehat{\psi}(\zeta)$, where $\varphi(0) = \psi(0) = 0$ and $\varphi(1) = \psi(1) = n$. Since multiplication by $f(1)$ and by $g(1)$ yields rotations, and so maps that are homotopic to $\mathrm{id}_{\mathbb{S}^1}$, and since by the considerations before 3.2.3 we have $\varphi \simeq \varphi_n \simeq \psi$, it follows that $f \simeq \widehat{\varphi} \simeq \widehat{\varphi_n} \simeq \widehat{\psi} \simeq g$. ☐

3.2.8 Examples.

(a) The map $\mathrm{id}_{\mathbb{S}^1} : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ has degree 1, since $\mathrm{id}_{\mathbb{S}^1} = g_1$.

(b) If $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ is nullhomotopic, than $\deg(f) = 0$, since then $f \simeq g_0$.

(c) The reflection $\rho : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ on the x-axis, that is, the map ρ such that $\rho(\zeta) = \overline{\zeta}$, has degree -1, since $s = g_{-1}$. More generally, since any other reflection $\rho' : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ is conjugate to ρ by a rotation $r : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$, that is, $\rho' = r^{-1} \circ \rho \circ r$, we have that $\deg(\rho') = \deg(\rho) = -1$.

3.2.9 **Proposition**. *Given $f, g : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$, then*

$$\deg(f \cdot g) = \deg(f) + \deg(g)\,,$$

where $f \cdot g : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ denotes the mapping $\zeta \mapsto f(\zeta)g(\zeta)$, using the complex multiplication in $\mathbb{S}^1$.

Proof: If $f \simeq g_m$ and $g \simeq g_n$, then $f \cdot g \simeq g_m \cdot g_n = g_{m+n}$. □

3.2.10 **Proposition**. *Given $f, g : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$, then*

$$\deg(f \circ g) = \deg(f)\deg(g)\,.$$

Proof: If $f \simeq g_m$ and $g \simeq g_n$, then $f \circ g \simeq g_m \circ g_n = g_{mn}$. □

3.2.11 **Corollary**. *If $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ is a homeomorphism, then $\deg(f) = \pm 1$. Consequemtly, $f \simeq \mathrm{id}_{\mathbb{S}^1}$ or $f \simeq \rho$, where ρ is the reflection given by taking complex conjugates.*

Proof: Since $f \circ f^{-1} = \mathrm{id}$, $\deg(f)\deg(f^{-1}) = 1$; this is only possible if $\deg(f) = \deg(f^{-1}) = \pm 1$. In particular, we have that $\deg(f) = \deg(f^{-1})$. □

3.2.12 Definition. We say that a map $f : \mathbb{S}^m \longrightarrow \mathbb{S}^n$ is *odd* if for every $x \in \mathbb{S}^m$, $f(-x) = -f(x)$; we say that the map is *even* if for every $x \in \mathbb{S}^n$, $f(-x) = f(x)$.

3.2.13 **Theorem**.

(a) *If $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ is odd, then $\deg(f)$ is odd.*

(b) *If $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ is even, then $\deg(f)$ is even.*

Proof: (a) By 3.2.2, there is a map $\varphi : I \longrightarrow \mathbb{R}$ such that $\varphi(0) = 0$, $\varphi(1) = \deg(f)$, and

$$f(\mathrm{e}^{2\pi\mathrm{i}t}) = f(1) \cdot \mathrm{e}^{2\pi\mathrm{i}\varphi(t)} .$$

From $-\mathrm{e}^{2\pi\mathrm{i}t} = \mathrm{e}^{2\pi\mathrm{i}(t+\frac{1}{2})}$ and $-f(\mathrm{e}^{2\pi\mathrm{i}t}) = f(-\mathrm{e}^{2\pi\mathrm{i}t}) = f(\mathrm{e}^{2\pi\mathrm{i}(t+\frac{1}{2})})$ it follows that

$$\mathrm{e}^{2\pi\mathrm{i}(\varphi(t)+\frac{1}{2})} = -\mathrm{e}^{2\pi\mathrm{i}\varphi(t)} = \mathrm{e}^{2\pi\mathrm{i}\varphi(t+\frac{1}{2})} ,$$

and therefore

$$\varphi\left(t + \frac{1}{2}\right) = \varphi(t) + \frac{1}{2} + k ,$$

where k is an integer that does not depend on t, since I is connected and φ is continuous. For $t = 0$ one has that $\varphi(\frac{1}{2}) = \varphi(0 + \frac{1}{2}) = \varphi(0) + \frac{1}{2} + k = \frac{1}{2} + k$. For $t = \frac{1}{2}$, one has

$$\deg(f) = \varphi(1) = \varphi\left(\frac{1}{2} + \frac{1}{2}\right) = \varphi\left(\frac{1}{2}\right) + \frac{1}{2} + k = \frac{1}{2} + k + \frac{1}{2} + k = 1 + 2k ,$$

and therefore $\deg(f)$ is odd.

The even case is proved similarly. □

3.2.14 EXERCISE. Prove (b) in the theorem above.

3.2.15 EXERCISE. The set $[\mathbb{S}^1, \mathbb{S}^1]$ has an *additive structure* (namely the structure of an abelian group), given by $[f] + [g] = [f \cdot g]$ (see 3.2.9) and a *multiplicative structure* given by $[f][g] = [f \circ g]$ (see 3.2.10). Prove that $[\mathbb{S}^1, \mathbb{S}^1]$ is a commutative ring with $0 = [g_0]$ ($g_0(\zeta) = 1$ for all $\zeta \in \mathbb{S}^1$) and $1 = [g_1]$ ($g_1(\zeta) = \zeta$ for all $\zeta \in \mathbb{S}^1$) with respect to these structures. Conclude that the function $[\mathbb{S}^1, \mathbb{S}^1] \longrightarrow \mathbb{Z}$ given by $[f] \mapsto \deg(f)$ is a ring isomorphism. According to Exercise 3.1.36, this proves that the first cohomology group of the circle is $\mathbb{Z}$, i.e., $H^1(\mathbb{S}^1) \cong \mathbb{Z}$.

3.2.16 **Proposition**. *The inclusions $i, j : \mathbb{S}^1 \hookrightarrow \mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ given by $i(z) = (z, 1)$, $j(z) = (1, z)$ are not nullhomotopic and are not homotopic to each other; that is, $0 \neq [i] \neq [j] \neq 0$.*

Proof: If i and j were nullhomotopic, then the composites $\mathrm{proj}_1 \circ i = \mathrm{id}_{\mathbb{S}^1}$ and $\mathrm{proj}_2 \circ j = \mathrm{id}_{\mathbb{S}^1}$ would also be nullhomotopic, thus contradicting 3.2.8(a). Similarly, if i and j were homotopic, then the composites $\mathrm{proj}_1 \circ i = \mathrm{id}_{\mathbb{S}^1} = g_1$ and $\mathrm{proj}_1 \circ j = g_0$ would also be homotopic, a result that would contradict 3.2.8(a) and (b). □

Proposition 3.2.16 above, showing that the maps i and j are not homotopic, corresponds to the intuitive idea that each of the two maps “surrounds” a certain “hole”. In fact, i surrounds the “exterior hole” of the tube forming the torus, and

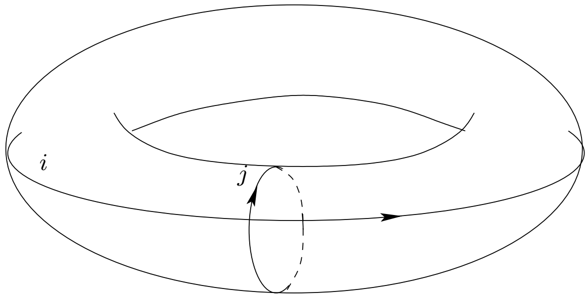


Figure 3.3 The generators i and j on the torus

j the "interior hole," and these two holes are essentially different (see Figure 3.3). Let us say, colloquially, that if we consider the torus as a car tire, then i surrounds the tube and j surounds the rim.

The next example is probably more eloquent. If we bore a hole into the complex plane $\mathbb{C}$, let us say, to obtain the complement of the origin $\mathbb{C}-0$, then the inclusion $i : \mathbb{S}^1 \hookrightarrow \mathbb{C} - 0$ is not nullhomotopic, since if it were, then the map

$$\mathrm{id}_{\mathbb{S}^1} : \mathbb{S}^1 \xrightarrow{i} \mathbb{C} - 0 \xrightarrow{r} \mathbb{S}^1 ,$$

would also be nullhomotopic, where $r(z) = z/|z|$. What this shows is that the map $i : \mathbb{S}^1 \longrightarrow \mathbb{C} - 0$ detects the hole. It is in this sense that we shall systematize in the next section the study of maps $\mathbb{S}^1 \longrightarrow X$ for any topological space X in order to "detect holes" or, in other words, to measure certain kinds of complications in the structure of the space X.

3.2.17 REMARK. Let $f : \mathbb{S}^1 \longrightarrow \mathbb{C}$ be continuous and $z_0 \notin f(\mathbb{S}^1)$. A rasonable question is the following: How many times does the curve described by f turn around z_0? The answer is not always intuitively clear, as shown in Figure 3.4.

The answer is as follows. First, if $r : \mathbb{C} - 0 \longrightarrow \mathbb{S}^1$ is the retraction given by $r(z) = z/|z|$, then the map

$$f_{z_0} : \mathbb{S}^1 \xrightarrow{f} \mathbb{C} - z_0 \xrightarrow{t_{z_0}} \mathbb{C} - 0 \xrightarrow{r} \mathbb{S}^1 ,$$

where $t_{z_0}(z) = z - z_0$, is well defined. Then the answer to the question posed is that the curve described by f surrounds the point z_0 precisely $\deg(f_{z_0})$ times. This number is called the *winding number* of the curve $f(\mathbb{S}^1)$, and we denote it by $W(f, z_0)$. In other words,

$$W(f, z_0) = \deg(f_{z_0}) \quad \text{where} \quad f_{z_0}(\zeta) = \frac{f(\zeta) - z_0}{|f(\zeta) - z_0|} . \tag{3.2.18}$$

For example, see [8] for a systematic and more general study of the degree, the winding number, and other related concepts.

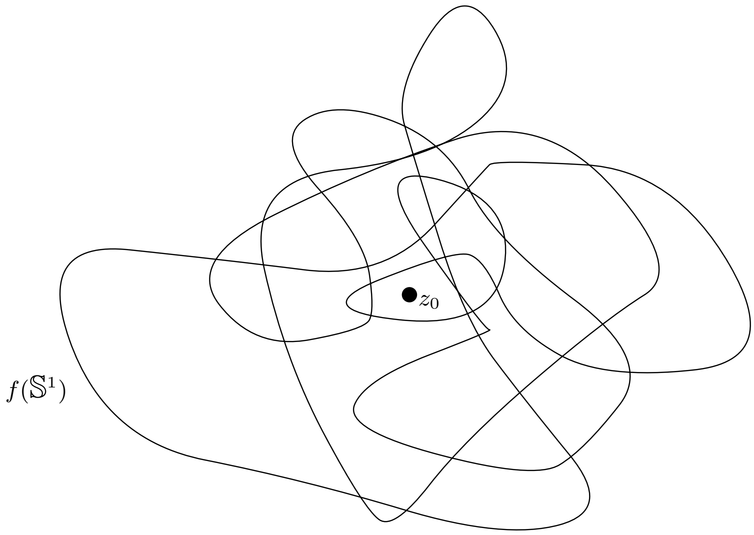


Figure 3.4 How many times does the curve $f(\mathbb{S}^1)$ turn around z_0?

As a matter of fact, when f is differentiable, then the winding number around z_0 corresponds to the number obtained by the Cauchy formula; that is,

$$W(f,z_0) = \deg(f_{z_0}) = \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{S}^1} \frac{f'(\zeta)}{f(\zeta)-z_0} \mathrm{d}\zeta\,.$$

(See [22] or [3].)

Having been able to classify maps $\mathbb{S}^1 \longrightarrow \mathbb{S}^1$ up to homotopy brings many nice consequences. From the fact that $\deg(\mathrm{id}_{\mathbb{S}^1}) = 1$, one has that $\mathrm{id}_{\mathbb{S}^1}$ is not nullhomotopic, and from this we obtain the following.

3.2.19 **Theorem**. *The circle $\mathbb{S}^1$ is not contractible.*

Proof: If it were, then $\mathrm{id}_{\mathbb{S}^1}$ would be nullhomotopic. □

In the example of $i : \mathbb{S}^1 \longrightarrow \mathbb{C} - 0$, we saw that $r : \mathbb{C} - 0 \longrightarrow \mathbb{S}^1$ is a *retraction* of the punctured plane $\mathbb{C} - 0$ onto the subspace $\mathbb{S}^1$; this way of thinking allows us to prove an interesting fact, which is the following.

3.2.20 **Theorem**. *There is no retraction $r : \mathbb{B}^2 \longrightarrow \mathbb{S}^1$, that is, there is no map r such that $r|_{\mathbb{S}^1} = \mathrm{id}_{\mathbb{S}^1}$.*

Proof: Since $\mathbb{B}^2$ is contractible (see 3.1.10(f)), any map defined on $\mathbb{B}^2$ is nullhomotopic, and in particular, r would be so too. But this would be a contradiction, since the composition of r with the inclusion $\mathbb{S}^1 \hookrightarrow \mathbb{B}^2$, which is $\mathrm{id}_{\mathbb{S}^1}$, would also be nullhomotopic. Thus, such an r cannot exist. □

The proposition above allows us to prove a very important result in topology with many applications. It is known as Brouwer's fixed point theorem.

3.2.21 **Theorem**. *Every map* $f : \mathbb{B}^2 \longrightarrow \mathbb{B}^2$ *has a* fixed point*, that is, a point* $x_0 \in \mathbb{B}^2$ *such that* $f(x_0) = x_0$.

Proof: If there were no such x_0, then we would have $f(x) \neq x$ for every $x \in \mathbb{B}^2$. Thus the points x and $f(x)$ would determine a ray that starts at $f(x)$, passes through x, and intersects $\mathbb{S}^1$ in exactly one point $r(x)$ (see Figure 3.5). The map $r \colon \mathbb{B}^2 \to \mathbb{S}^1$ is well defined, continuous, and is also a retraction. However, the existence of such a retraction is denied by Proposition 3.2.20. □

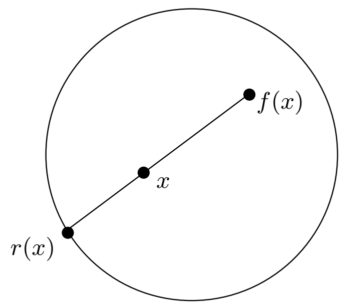


Figure 3.5 The hypothetic retraction $r : \mathbb{B}^2 \longrightarrow \mathbb{S}^1$

3.2.22 Exercise. For a given map f, find an explicit formula for the retraction $r : \mathbb{B}^2 \longrightarrow \mathbb{S}^1$ described in the proof of Brouwer's fixed point theorem 3.2.21.

3.2.23 Exercise. Take $X = \{(x, y, z) \in \mathbb{R}^3 \mid |x| \leq 1, |and| \leq 2, |z| \leq 3\}$ and consider the map $f : X \longrightarrow \mathbb{R}^3$ given by

$$f(x, y, z) = \left(x - \frac{y^2 + z^2 + 1}{14}, y - \frac{x^2 + z^2 + 4}{14}, z - \frac{x^2 + y^2 + 9}{14}\right).$$

Prove that the equation $f(x, y, z) = 0$ has a solution (in X). (*Hint:* Use Brouwer's fixed point theorem 3.2.21.)

The Brouwer fixed point theorem is valid in general. Its proof follows similarly to that of 3.2.21. One needs the following.

3.2.24 **Theorem**. *There is no retraction* $r : \mathbb{B}^n \longrightarrow \mathbb{S}^{n-1}$*, that is, there is no map* r *such that* $r|_{\mathbb{S}^{n-1}} = \mathrm{id}_{\mathbb{S}^{n-1}}$.

The proof requires some arguments, either from analysis or from algebraic topology, which are beyond the scope of this text. However, using this result, using one proves the general Brouwer theorem.

3.2.25 **Theorem**. *Every map $f : \mathbb{B}^n \longrightarrow \mathbb{B}^n$ has a fixed point, namely there is $x_0 \in \mathbb{B}^n$ such that $f(x_0) = x_0$.*

The following result is equivalent to the retraction theorem 3.2.24

3.2.26 **Theorem**. *The identity map $\mathbb{S}^{n-1} \longrightarrow \mathbb{S}^{n-1}$ is nonnullhomotopic.*

Proof: If 3.2.24 holds and $H : \mathbb{S}^{n-1} \times I \longrightarrow \mathbb{S}^{n-1}$ is a nullhomotopy of the identity, then the map $r : \mathbb{B}^n \longrightarrow \mathbb{S}^{n-1}$ given by $r(tx) = H(x,t)$ is a retraction.

Conversely, if $r : \mathbb{B}^n \longrightarrow \mathbb{S}^{n-1}$ is a retraction, then $H(x,t) = r(tx)$ defines a nullhomotopy of the identity. □

The degree concept is so useful that it frequently has applications beyond the limits of topology. A nice example of this is the proof of the fundamental theorem of algebra.

3.2.27 **Theorem**. (Fundamental theorem of algebra) *Every nonconstant polynomial with complex coefficients has a root. That is, if $f(z) = a_0 + a_1 z + \cdots + a_{n-1}z^{n-1} + z^n$, $n > 0$, $a_0, a_1, \ldots, a_{n-1} \in \mathbb{C}$, then there exists $z_0 \in \mathbb{C}$ such that $f(z_0) = 0$.*

Proof: Assuming that f does not have a root, the mapping $z \mapsto f(z)$ would determine a map $f : \mathbb{C} \longrightarrow \mathbb{C} - 0$. If we take $\mu = |a_0| + |a_1| + \cdots + |a_{n-1}| + 1$ and $z \in \mathbb{S}^1$, then

$$\begin{aligned}|f(\mu z) - \mu^n z^n| &\leq |a_0| + \mu|a_1| + \cdots + \mu^{n-1}|a_{n-1}| \\ &\leq \mu^{n-1}(|a_0| + |a_1| + \cdots + |a_{n-1}|) \quad (\text{since } \mu \geq 1) \\ &< \mu^n = |\mu^n z^n| \quad (\text{since } \mu > |a_0| + |a_1| + \cdots + |a_{n-1}|).\end{aligned}$$

Therefore, $f(\mu z)$ lies in the interior of a circle with center at $\mu^n z^n$ and radius $|\mu^n z^n|$, and so the line segment connecting $f(\mu z)$ with $\mu^n z^n$ does not contain the origin. Hence $H(z,t) = (1-t)f(\mu z) + t\mu^n z^n$ determines a homotopy $H : \mathbb{S}^1 \times I \longrightarrow \mathbb{C} - 0$ between the mapping $z \mapsto f(\mu z)$ and the mapping $z \mapsto \mu^n z^n$. Since the first map is nullhomotopic using the nullhomotopy $(z,t) \mapsto f((1-t)\mu z)$, so also is the second map. Therefore, by composing it with the known retraction $r : \mathbb{C} - 0 \longrightarrow \mathbb{S}^1$ given by $r(z) = z/|z|$, we obtain that the map $\mathbb{S}^1 \longrightarrow \mathbb{S}^1$ given by $z \mapsto z^n$ is nullhomotopic. But this last map is g_n, and so we have contradicted 3.2.7. □

Another application of the degree, or more precisely, of the winding number $W(f,z)$ defined above in 3.2.17, is to prove a version of the Jordan curve theorem. This assertion will be based on the following proposition.

3.2.28 **Proposition**. *Let $f : \mathbb{S}^1 \longrightarrow \mathbb{C}$ be continuous, and let z_0 and z_1 be points on the same path component of $\mathbb{C} - f(\mathbb{S}^1)$. Then $W(f,z_0) = W(f,z_1)$.*

Proof: If $\lambda : z_0 \simeq z_1$ is a path, then $f_{\lambda(t)}$ given by

$$f_{l(t)}(\zeta) = \frac{f(\zeta) - l(t)}{|f(\zeta) - l(t)|}$$

(see 3.2.18) is a homotopy from f_{z_0} to f_{z_1}; consequently,

$$W(f,z_0) = \deg(f_{z_0}) = \deg(f_{z_1}) = W(f,z_1)\,.$$

□

The following is a weak version of the famous Jordan curve theorem.

3.2.29 **Theorem**. *Given any map $f : \mathbb{S}^1 \longrightarrow \mathbb{C}$, the complement of its image $\mathbb{C} - f(\mathbb{S}^1)$ contains only one unbounded path component For z inside this component, one has that $W(f,z) = 0$.*

Proof: Since $f(\mathbb{S}^1)$ is compact, being the continuous image of a compact set, the Heine-Borel theorem guarantees that it is bounded. Thus its complement $\mathbb{C} - f(\mathbb{S}^1)$ contains an unbounded component V. If $\mu > 0$ is large enough, then $f(\mathbb{S}^1) \subset D = \{z \in \mathbb{C} \mid |z| \leq \mu\}$, $\mathbb{C} - D \subset \mathbb{C} - f(\mathbb{S}^1)$, and, since D is bounded, $(\mathbb{C} - D) \cap V \neq \emptyset$. Therefore, since $\mathbb{C} - D$ is path connected, $\mathbb{C} - D \subset V$ and V is the only unbounded component of $\mathbb{C} - f(\mathbb{S}^1)$. If $z \in V$ and $z' \in \mathbb{C} - D$, then by 3.2.28, $W(f,z) = W(f,z')$. Moreover, the homotopy

$$H(\zeta,t) = \frac{(1-t)f(\zeta) - z'}{|(1-t)f(\zeta) - z'|}$$

starts at $f_{z'}$ and ends at a constant map, and so one has that $W(f,z') = \deg(f_{z'}) = 0$. □

The classical Jordan curve theorem states that *given an embedding $e : \mathbb{S}^1 \hookrightarrow \mathbb{R}^2$, then the complement $\mathbb{R}^2 - e(\mathbb{S}^1)$ has exactly two components, one bounded and one unbounded.* The latter is the one given by 3.2.29. One can prove that $W(e,z) = \pm 1$ if z lies inside the bounded component.

Another beautiful result of algebraic topology is the Borsuk–Ulam theorem, of which we shall prove only its two-dimensional version. This result implies the nonexistence of an embedding $\mathbb{S}^2 \hookrightarrow \mathbb{R}^2$.

3.2.30 **Theorem**. (Borsuk–Ulam) *Given a continuous map* $f : \mathbb{S}^2 \longrightarrow \mathbb{R}^2$*, there is a point* $x \in \mathbb{S}^2$ *such that* $f(x) = f(-x)$.

Proof: If we assume that $f(x) \neq f(-x)$ for every point $x \in \mathbb{S}^2$, then one cane define two maps, namely

$$\begin{aligned} f_1 : \mathbb{S}^2 \longrightarrow \mathbb{S}^1 \quad &\text{given by} \quad f_1(x) = \frac{f(x) - f(-x)}{|f(x) - f(-x)|}\,, \\ f_2 : \mathbb{B}^2 \longrightarrow \mathbb{S}^1 \quad &\text{given by} \quad f_2(x_1, x_2) = f_1\left(x_1, x_2, \sqrt{1 - x_1^2 - x_2^2}\right). \end{aligned}$$

If we define $g = f_2|_{\mathbb{S}^1} : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$, then we have, on the one hand, that g is nullhomotopic, since the homotopy

$$H : \mathbb{S}^1 \times I \longrightarrow \mathbb{S}^1 \quad \text{given by} \quad H(\zeta, t) = f_2((1-t)\zeta)$$

is a nullhomotopy. On the other hand, g is odd, that is, $g(-\zeta) = -g(\zeta)$, since f_1 is odd. By 3.2.13(a) one has that $\deg(g)$ is odd, thus contradicting that g is nullhomotopic. $\square$

3.2.31 Note. The Borsuk–Ulam theorem is often described in meteorological terms as follows. If we assume that temperature T and athmospheric pressure P are continuous functions of location on the surface of the Earth, then both determine a map $f = (T, P) : \mathbb{S}^2 \longrightarrow \mathbb{R}^2$. The theorem asserts in this case that there exists a pair of antipodal points with the same temperature and pressure.

If $g : \mathbb{S}^2 \longrightarrow \mathbb{S}^1$ is continuous, then it cannot be odd, that is, it cannot happen that $g(-x) = -g(x)$, since the composite

$$\mathbb{S}^2 \xrightarrow{g} \mathbb{S}^1 \hookrightarrow \mathbb{R}^2$$

would be a counterexample to the Borsuk–Ulam theorem 3.2.30. In the proof of this theorem, by assuming the contrary of its assertion, that is, the existence of $f : \mathbb{S}^2 \longrightarrow \mathbb{R}^2$ such that for every $x \in \mathbb{S}^2$, $f(x) \neq f(-x)$, we could construct an odd map $g : \mathbb{S}^2 \longrightarrow \mathbb{S}^1$. We have hence that the Borsuk–Ulam theorem is equivalent to the following.

3.2.32 **Theorem**. *There are no continuous odd maps* $f : \mathbb{S}^2 \longrightarrow \mathbb{S}^1$. $\square$

3.2.33 Remark. There is a general version of the Borsuk–Ulam theorem stating that *given a continuous map* $f : \mathbb{S}^n \longrightarrow \mathbb{R}^n$*, there is a point* $x \in \mathbb{S}^2$ *such that* $f(x) = f(-x)$. As before, this statement is equivalent to saying that *there are no continuous odd maps* $f : \mathbb{S}^n \longrightarrow \mathbb{S}^{n-1}$. In order to prove these facts, more sophisticated machinery is needed (see [2, 11.8.28 and 11.8.29]).

3.2.34 EXERCISE. By assuming that the generalization of the Borsuk–Ulam theorem given in 3.2.33 is true, prove that *if there exists a continuous odd map* $f : \mathbb{S}^m \longrightarrow \mathbb{S}^n$*, then* $m \leq n$. Moreover, prove that this last statement is equivalent to the Borsuk–Ulam theorem. (*Hint:* Observe that if $m \leq n$, then $\mathbb{S}^m \subset \mathbb{S}^n$.)

3.2.35 EXERCISE. Let $f : \mathbb{B}^2 \longrightarrow \mathbb{R}^2$ be an odd map on the boundary, that is, such that if $x \in \mathbb{S}^1$, then $f(-x) = -f(x)$. Prove that there exists $x_0 \in \mathbb{B}^2$ such that $f(x_0) = 0$.

3.2.36 EXERCISE. Consider the following system of equations:

$$\begin{aligned} x \cos y &= x^2 + y^2 - 1 \\ y \cos x &= \tan 2\pi(x^3 + y^3)\,. \end{aligned}$$

Using the last exercise, prove that the system has a solution (x_0, y_0) such that $x_0^2 + y_0^2 \leq 1$.

One last result in this section, whose proof is an application of the Borsuk–Ulam theorem, is the so-called ham sandwich theorem. In order to state it, we need the following preparatory considerations. For each point $a = (a_1, a_2, a_3) \in \mathbb{S}^2$ and each element $d \in \mathbb{R}$, let $E(a, d) \subset \mathbb{R}^3$ be the plane given by the equation

$$\gamma_a(x) = a_1x_1 + a_2x_2 + a_3x_3 - d = 0\,,$$

and let $E^+(a, d)$ and $E^-(a, d)$ be the half-spaces of $\mathbb{R}^3$ such that $\gamma_a(x) \geq 0$ and $\gamma_a(x) \leq 0$, respectively. Obviously, $E^+(-a, -d) = E^-(a, d)$. Let $A_1, A_2, A_3 \subset \mathbb{R}^3$ be subsets such that the functions $f_\nu^\pm : \mathbb{S}^2 \times \mathbb{R} \longrightarrow \mathbb{R}$, where $f_\nu^\pm(a, d)$ is the volume of $A_\nu \cap E^\pm(a, d)$, $\nu = 1, 2, 3$, are well defined and continuous. Moreover, for each $a \in \mathbb{S}^2$ there exists a unique $d_a \in \mathbb{R}$ depending continuously on a and such that $f_1^+(a, d_a) = f_1^-(a, d_a)$. This last condition means that given any family of parallel planes, there exists only one that divides the set A_1 in two portions of equal volume. Clearly, $d_{-a} = -d_a$. Under these conditions, one has the following result.

3.2.37 **Theorem**. (Ham sandwich theorem) *Given subsets* A_1*,* A_2*,* $A_3 \subset \mathbb{R}^3$ *as above, there exists a plane in* $\mathbb{R}^3$ *dividing each of the sets* A_1*,* A_2*,* A_3 *in portions of equal volume.*

Proof: If $f : \mathbb{S}^2 \longrightarrow \mathbb{R}^2$ is the map given by

$$f(a) = (f_2^+(a, d_a), f_3^+(a, d_a))\,,$$

then, by the assumptions, f is well defined and continuous. By the Borsuk–Ulam theorem 3.2.30 there exists $b \in \mathbb{S}^2$ such that $f(b) = f(-b)$. By the properties of d_a and $E^{\pm}(a,d)$, one has for this b that $f_\nu^+(b, d_b) = f_\nu^+(-b, d_{-b}) = f_\nu^+(-b, -d_b) = f_\nu^-(b, d_b)$, as was required. □

3.2.38 Note. As indicated by its name, a gastronomic interpretation of the ham sandwich theorem can be given if we assume that A_1 is the bread, A_2 the butter, and A_3 the ham that were used to prepare a sandwich. The theorem guarantees that it is possible to cut the sandwich with a flat knife, independent of the distribution of the ingredients, in such a way that each of the two pieces contains exactly the same amount of bread, butter, and ham.

3.2.39 Exercise. Prove the Borsuk–Ulam theorem in dimension 1; that is, prove that given a map $f : \mathbb{S}^1 \longrightarrow \mathbb{R}$, there exists $x \in \mathbb{S}^1$ such that $f(x) = f(-x)$. (*Hint:* Use the intermediate value theorem for a convenient function.)

3.2.40 Exercise. State the ham sandwich theorem in $\mathbb{R}^2$ and apply the former exercise to prove it.

3.2.41 Exercise. Indicate which of the following maps f are nullhomotopic and which are not.

(a) $f : \mathbb{S}^n \longrightarrow \mathbb{S}^{n+1}$, $f(x) = (x, 0)$.

(b) $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1 \times \mathbb{S}^1$, $f(\zeta) = (\zeta^2, \zeta^3)$.

(c) $f : \mathbb{S}^1 \times \mathbb{S}^1 \longrightarrow \mathbb{S}^1 \times \mathbb{S}^1$, $f(\xi, \eta) = (\xi\eta, 1)$.

(d) $f : \mathbb{R}^2 - \{0\} \longrightarrow \mathbb{R}^2 - \{0\}$, $f(x, y) = (x^2 - y^2, 2xy)$.

(e) $f : \mathbb{R}^2 - \{0\} \longrightarrow \mathbb{R}^2 - \{0\}$, $f(x, y) = (x^2, y)$.

3.3 Homotopy equivalence

The concept of homotopy equivalence is weaker than that of homeomorphism, but the homotopy methods allow us to distinguish spaces up to homotopy equivalence. If by some homotopy argument, we are able to decide that two spaces are not homotopy equivalent, then we shall immediately conclude that they are not homeomorphic either.

Let us recall that two spaces X and Y are homeomorphic if there exist maps $f : X \longrightarrow Y$ and $g : Y \longrightarrow X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. The concept that we are about to introduce, is obtained by replacing in the two previous equations the equalities by homotopies.

3.3.1 Definition. We say that two spaces X and Y are *homotopy equivalent*, or that they have *the same homotopy type*,* in symbols $X \simeq Y$, if there exist maps $f : X \longrightarrow Y$ and $g : Y \longrightarrow X$ such that

$$g \circ f \simeq \mathrm{id}_X \quad \text{and} \quad f \circ g \simeq \mathrm{id}_Y \,.$$

Each of these two maps f and g is called a *homotopy equivalence* and we may write $f : X \simeq Y$ or $g : Y \simeq X$;. Moreover, each of them is a *homotopy inverse* of the other.

3.3.2 Example. If X is a contractible space, then it is homotopy equivalent to a singular space (a one-point space), that is, $X \simeq *$, since if $\mathrm{id}_X \simeq c_{x_0}$, then the maps $f : X \longrightarrow *$, and $g : * \longrightarrow X$ such that $g(*) = x_0$ are homotopy inverse, because $g \circ f = \mathrm{id}_*$ and $f \circ g = c_{x_0} \simeq \mathrm{id}_X$. In particular, $\mathbb{B}^n \simeq *$.

3.3.3 Exercise. Prove that, conversely to the statement of Example 3.3.2, if a space X is homotopy equivalent to a point, then it is contractible. Namely, one has that X *is contractible if and only if* X *is homotopy equivalent to a point.*

The following statement that shows that the concept of homotopy equivalence is weaker than that of homeomorphism is immediate.

3.3.4 **Proposition**. *If $X \approx Y$, then $X \simeq Y$.* □

It is an easy *exercise* to prove the following statement.

3.3.5 **Proposition**. *The homotopy equivalence relation is, in fact, an equivalence relation in the class of topological spaces.* □

Corresponding to the concept of homotopy of pairs, one has the following concept.

3.3.6 Definition. Two pairs of spaces (X, A) and (Y, B) are *homotopy equivalent*, or they have *the same homotopy type*, in symbols $(X, A) \simeq (Y, B)$, if there exist maps of pairs $f : (X, A) \longrightarrow (Y, B)$ and $g : (Y, B) \longrightarrow (X, A)$ such that

$$g \circ f \simeq \mathrm{id}_{(X,A)} \quad \text{and} \quad f \circ g \simeq \mathrm{id}_{(Y,B)} \,;$$

that is, these composites ar homotopic to the identity maps via homotopies of pairs.

*Mostly in the literature is this second the usual denomination.

We use analogous designations in the relative case, to those of the absolute case. An important special case is the *homotopy equivalence of pointed spaces.*

The following statement is clear.

3.3.7 **Proposition**. *If $f : (X, A) \longrightarrow (Y, B)$ is a homotopy equivalence of pairs, then $f : X \longrightarrow Y$ and $f|_A : A \longrightarrow B$ are homotopy equivalences.* □

The converse of the previous statement is false as is shown in the next exercise.

3.3.8 EXERCISE. Let $P \subset \mathbb{R}^2$ be the comb space and take $A = \{(0, 1)\}$. Prove that the map $f : (X, A) \longrightarrow (X, A)$ given by $f(x) = x_0 = (0, 1)$ satisfies that $f : X \longrightarrow X$ and $f|_A : A \longrightarrow A$ are homotopy equivalences; however, f is not a homotopy equivalence of pairs. (Observe that this assertion shows, in particular, that the converse of the statement in 3.1.21 is not true.)

An important source of homotopy equivalences is given by the following concept.

3.3.9 DEFINITION. A subspace A of a space X is a *deformation retract* of X if there exists a homotopy $H : X \times I \longrightarrow X$ such that

$$\begin{aligned} H(x, 0) &= x && \text{if } x \in X \\ H(x, 1) &\in A && \text{if } x \in X \\ H(a, 1) &= a && \text{if } a \in A. \end{aligned} \tag{3.3.10}$$

Such homotopy H is called a *deformation retraction* of X in A. In particular, the map $r_H : X \longrightarrow A$ given by $r_H(x) = H(x, 1)$ is a *retraction* of X in A; namely, $r_H|_A = \mathrm{id}_A$ and A is a *retract* of X. If additionally to the conditions (3.3.10) we ask

$$H(a, t) = a \quad \text{if } a \in A\,,\ t \in I\,,$$

we say that A is a *strong deformation retract* of X and that H is a *strong deformation retraction.*

3.3.11 REMARK. In the previous definition H is a homotopy between the identity map and the retraction r_H, that is, $H : \mathrm{id}_X \simeq r_H$; if the deformation retraction H is strong, then the homotopy is relative to A, that is, $H : \mathrm{id}_X \simeq r_H$ rel A.

3.3.12 **Theorem**. *If A is a deformation retract of X, then the inclusion $i : A \hookrightarrow X$ is a homotopy equivalence.*

Proof: If $i : A \hookrightarrow X$ is the inclusion and $H : X \times I \longrightarrow X$ is a deformation retraction then, by Remark 3.3.11, $H : \mathrm{id}_X \simeq i \circ r_H$; on the other hand, by the third equation of (3.3.10), $r_H \circ i = r_H|_A = \mathrm{id}_A$. □

In general, the converse is false. However, one has the following.

3.3.13 Exercise. A homotopy $H : X \times I \longrightarrow X$ is called a *weak deformation retraction*, if in (3.3.10) one only requires that the map $r_H : X \longrightarrow A \subset X$ given by $r_H(x) = H(x,1)$ is a *weak retraction*, namely, it is such that $r_H|_A \simeq \mathrm{id}_A : A \longrightarrow A$. Prove that A *is a weak deformation retract of* X *if and only if the inclusion map* $A \hookrightarrow X$ *is homotopy equivalence.*

3.3.14 Exercise. Give an example of a weak deformation retract of X that is not a deformation retract of X. (In particular, the fact that an inclusion $i : A \hookrightarrow X$ is a homotopy equivalence does not imply that A is a deformation retract of X.)

3.3.15 Exercise. Prove that every compact convex subset of $\mathbb{R}^n$ is a (strong) deformation retract of $\mathbb{R}^n$.

3.3.16 Examples.

(a) Let $* \in CX$ be the vertex of the cone over X, $CX = X \times I/X \times \{1\}$; that is, if $q : X \times I \longrightarrow CX$ is the quotient map, then $* = q(x,1)$. The singular space $*$ is a strong deformation retract of CX, via the strong deformation retraction $H : CX \times I \longrightarrow CX$ given by $H(q(x,s),t) = q(x,(1-t)s+t)$.

(b) The unit sphere $\mathbb{S}^n$ is a strong deformation retract of $X = \mathbb{B}^{n+1} - 0$ or of $X = \mathbb{R}^{n+1} - 0$, via $H : X \times I \longrightarrow X$ given by
$$H(x,t) = (1-t)x + t\frac{x}{|x|}\,.$$

(c) The disk with g holes $\mathbb{D}_g$ (see 2.2.9(b)) contains a wedge sum of g circles as a strong deformation retract. Consequently, the handle body $H_g = \mathbb{D}_g \times I$ contains $\mathbb{S}^1_1 \vee \cdots \vee \mathbb{S}^1_g$ as a strong deformation retract. Thus one has that $\mathbb{D}_g$, H_g, and the wedge of g circles are homotopy equivalent spaces. (See Figure 3.6.)

3.3.17 **Theorem**. *Let* Y *be a space obtained from* X *by attaching an* n*-cell via a map*
$$\mathbb{B}^n \supset \mathbb{S}^{n-1} \xrightarrow{\varphi} X\,;$$
that is, $Y = X \cup_\varphi e^n$. *If* y_0 *is the point in* Y *that comes from* $0 \in \mathbb{B}^n \subset X \sqcup \mathbb{B}^n$ *by taking the quotient, then* X *is a strong deformation retract of* $Y - \{y_0\}$, *as can be appreciated in Figure 3.7.*

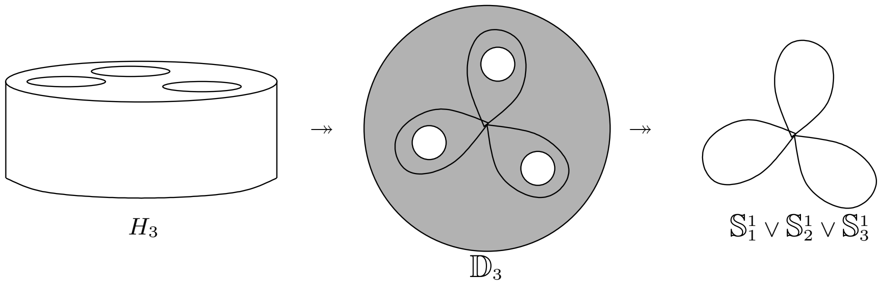


Figure 3.6 H_g, $\mathbb{D}_g$ and $\mathbb{S}^1_1 \vee \cdots \vee \mathbb{S}^1_g$ are homotopy equivalent.

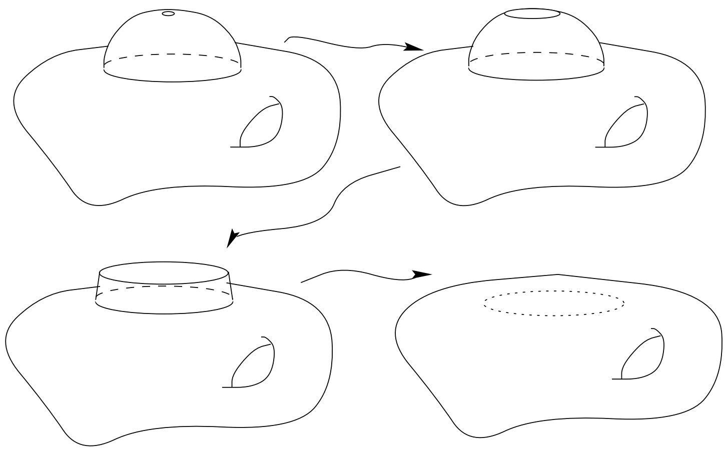


Figure 3.7 Deformation sequence of a space with a punctured cell

Proof: As we already saw in Example 3.3.16(b), the sphere $\mathbb{S}^{n-1}$ is a strong deformation retract of the punctured ball $\mathbb{B}^n - 0$. If $H : (\mathbb{B}^n - 0) \times I \longrightarrow \mathbb{B}^n - 0$ is a strong deformation retraction, then $H' = H \sqcup \mathrm{proj}_X : ((\mathbb{B}^n - 0) \sqcup X) \times I \longrightarrow (\mathbb{B}^n - 0) \sqcup X$ is a strong deformation retraction that determines K in the quotient space. This homotopy K is the desired deformation retraction such that the following diagram commutes.

$$\begin{tikzcd}
((\mathbb{B}^n - 0) \sqcup X) \times I \arrow[r, "H'"] \arrow[d, "q \times \mathrm{id}_I"'] & (\mathbb{B}^n - 0) \sqcup X \arrow[d, "q"] \\
((X \cup_\varphi e^n) - \{y_0\}) \times I \arrow[d, equal] & (X \cup_\varphi e^n) - \{y_0\} \arrow[d, equal] \\
(Y - \{y_0\}) \times I \arrow[r, dashed, "K"'] & Y - \{y_0\}\,.
\end{tikzcd}$$

Since I is locally compact, the vertical arrow on the left hand side is an identification, and this guarantees the continuity of K. Clearly, K starts at the identity map and ends at a retraction $Y - \{y_0\} \longrightarrow X \subset Y - \{y_0\}$, and X remains fixed along the whole deformation. □

Let us now consider the orientable surface S_g, resp. the nonorientable surface N_g, both of genus g. As in 2.2.11 and 2.2.22, it is a quotient space of the regular polygon E_{4g}, resp. E_{2g}. Let

$$p : E_{4g} \longrightarrow S_g \quad \text{resp.} \quad q : E_{2g} \longrightarrow N_g$$

be the quotient map. In particular, the boundary ∂E_{4g}, resp. ∂E_{2g}, is mapped under p, resp. q, to a wedge sum of circles; namely, all the vertices p_i, $i = 1, 2, \ldots, 4g$, resp. $i = 1, 2, \ldots, 2g$, of the polygon are mapped to just one point x_0, and the edges a_i, b_i, resp. a_i, $i = 1, 2, \ldots, g$, are mapped onto the circles α_i, β_i, resp. α_i, $i = 1, 2, \ldots, g$. Taking homeomorphisms

$$\underbrace{\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1}_{2g} \approx p(\partial E_{4g})\,, \text{ resp. } \underbrace{\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1}_{g} \approx q(\partial E_{2g})\,,$$

let

$$\varphi_g : \mathbb{S}^1 \longrightarrow \underbrace{\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1}_{2g}, \text{ resp. } \psi_g : \mathbb{S}^1 \longrightarrow \underbrace{\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1}_{g},$$

be such that the following diagrams commute:

$$\begin{tikzcd} \mathbb{S}^1 \arrow[r, "\varphi_g"] \arrow[d, "\approx"'] & \mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1 \arrow[d, "\approx"] \\ \partial E_{4g} \arrow[r, "p"'] & p(\partial E_{4g})\,, \end{tikzcd} \qquad \begin{tikzcd} \mathbb{S}^1 \arrow[r, "\psi_g"] \arrow[d, "\approx"'] & \mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1 \arrow[d, "\approx"] \\ \partial E_{2g} \arrow[r, "q"'] & q(\partial E_{2g})\,. \end{tikzcd}$$

Then we have the following consequence of this discussion.

3.3.18 **Proposition**. *The surface S_g is obtained by attaching a 2-cell to $p(\partial E_{4g})$ with the attaching map $\mathbb{S}^1 \approx \partial E_{4g} \longrightarrow p(\partial E_{4g})$, namely*

$$S_g \approx \underbrace{\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1}_{2g} \cup_{\varphi_g} e^2\,.$$

Analogously, the surface N_g is obtained by attaching a 2-cell to $q(\partial E_{2g})$ with the attaching map $\mathbb{S}^1 \approx \partial E_{2g} \longrightarrow q(\partial E_{2g})$, namely

$$N_g \approx \underbrace{\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1}_{g} \cup_{\psi_g} e^2\,.$$

□

From this proposition and 3.3.17, we obtain the following consequence.

3.3.19 **Corollary**. *If $x \in S_g$ and $y \in N_g$ are any points, then there are homotopy equivalences*

$$S_g - \{x\} \simeq \underbrace{\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1}_{2g} \qquad N_g - \{y\} \simeq \underbrace{\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1}_{g}\,.$$

□

3.3.20 NOTE. By 2.2.28 one has that each of the surfaces S_g and N_g are homogeneous, and thus it is irrelevant in the previous corollary, what points are $x \in S_g$ and $y \in N_g$.

Coming back to the general case, let us assume that Y is obtained from X by attaching a 2-cell; that is to say, $Y = X \cup_\varphi e^2$, with an attaching map $\varphi : \mathbb{S}^1 \longrightarrow X$. There is a path in Y given by the composite

$$\lambda_\varphi : I \xrightarrow{q} \mathbb{S}^1 \xrightarrow{\varphi} X \longrightarrow X \cup_\varphi e^2 = Y\,,$$

where $q(t) = \mathrm{e}^{2\pi i t}$. Clearly, this path is a *loop*, that is, $\lambda_\varphi(0) = \lambda_\varphi(1)$. There is a commutative diagram of pairs

$$\begin{tikzcd}
(I, \partial I) \arrow[r, "\lambda_\varphi"] \arrow[d] & (Y, y_0) \\
(\mathbb{S}^1, 1) \arrow[r, hook] & (\mathbb{B}^2, 1)\,, \arrow[u]
\end{tikzcd}$$

and since $\mathbb{B}^2$ is contractible, the loop λ_φ is nullhomotopic (relative to $\partial I = \{0, 1\}$; see Figure 3.8).

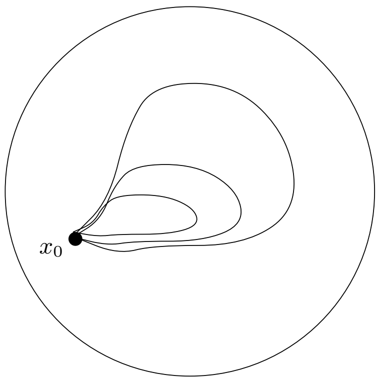


Figure 3.8 Every loop can be contracted inside a cell.

3.3.21 DEFINITION. Let us call λ_φ the *canonical loop* associated to the attaching space $Y = X \cup_\varphi e^2$.

3.3.22 **Corollary**. *For every attaching space $Y = X \cup_\varphi e^2$, the canonical loop is nullhomotopic; that is, one has that*

$$\lambda_\varphi \simeq c_{y_0} \text{ rel } \partial I\,,$$

where $y_0 = \varphi(1) \in Y$. □

3.3.23 Exercise. Similarly to Corollary 3.3.22, given a space Y obtained by attaching an n-cell to a space X with an attaching map $\varphi : \mathbb{S}^{n-1} \longrightarrow X$, namely $Y = X \cup_\varphi e^n$, prove that the homotopy class of the map $\gamma_\varphi : \mathbb{S}^{n-1} \xrightarrow{\varphi} X \hookrightarrow Y$ is trivial. In other words, prove that the map γ_φ is nullhomotopic.

No confusion should arise if we call

$$\alpha_i, \beta_i : (I, \partial I) \longrightarrow (S_g, x_0)\,;$$

$$\alpha_i : (I, \partial I) \longrightarrow (N_g, x_0)$$

the loops defined by the maps

$$\alpha_i : t \mapsto p((1-t)p_{4i-3} + tp_{4i-2})\,, \quad \beta_i : t \mapsto p((1-t)p_{4i-2} + tp_{4i-1})\,;$$

$$\alpha_i : t \mapsto q((1-t)p_{2i-1} + tp_{2i})\,,$$

respectively, where $p : E_{4g} \longrightarrow S_g$ and $q : E_{2g} \longrightarrow N_g$ are the quotient maps and the points p_i are as in 2.2.11 and 2.2.22. The canonical loop λ_{φ_g} in S_g travels during the lapse $\left[0, \frac{1}{4g}\right]$ the same as does the loop α_1 (reparametrized by $[0,1] \longrightarrow \left[0, \frac{1}{4g}\right]$); during the lapse $\left[\frac{1}{4g}, \frac{2}{4g}\right]$, the same as β_1; during the lapse $\left[\frac{2}{4g}, \frac{3}{4g}\right]$, the same as α_1, but in the opposite sense; during the lapse $\left[\frac{3}{4g}, \frac{4}{4g}\right]$, the same as β_1, but in the opposite sense, and, in general, during the lapse $\left[\frac{4i-4}{4g}, \frac{4i-3}{4g}\right]$, the same as α_i; during the lapse $\left[\frac{4i-3}{4g}, \frac{4i-2}{4g}\right]$, the same as β_i; during the lapse $\left[\frac{4i-2}{4g}, \frac{4i-1}{4g}\right]$, the same as α_i, but in the opposite sense, and during the lapse $\left[\frac{4i-1}{4g}, \frac{4i}{4g}\right]$, the same as β_i, but in the opposite sense, $i = 1, \dots, g$. We can express this by writing simply $\lambda_{\varphi_g} = \alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1} \cdots \alpha_g\beta_g\alpha_g^{-1}\beta_g^{-1}$, (see 4.4.6(c) in the following chapter). Similarly, in the nonorientable case for the surfaces N_g, we express the canonical loop as $\lambda_{\psi_g} = \alpha_1^2 \cdots \alpha_g^2$.

From 3.3.22, we obtain the following statement.

3.3.24 **Theorem**. *The loops $\lambda_{\varphi_g} = \alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1} \cdots \alpha_g\beta_g\alpha_g^{-1}\beta_g^{-1}$ in S_g; and $\alpha_1^2 \cdots \alpha_g^2$ in N_g are nullhomotopic.* □

This theorem will appear again in the next chapter when we compute the fundamental groups of the surfaces.

To finish this section, we recall the mapping cylinder M_f of a map $f : X \longrightarrow Y$, defined in 1.2.5, and that is obtained by identifying each point $(x, 0) \in X \times I$ with $f(x) \in Y$. If we denote the quotient map by $q : (X \times I) \sqcup Y \longrightarrow M_f$, we have a homotopy $H : M_f \times I \longrightarrow M_f$ given by

$$H(z,t) = \begin{cases} q(x, (1-t)s) & \text{if } z = q(x,s)\,, (x,s) \in X \times I, \\ q(y) & \text{if } z = q(y)\,, y \in Y, \end{cases}$$

that is clearly a strong deformation retraction of M_f onto $Y \subset M_f$. We have proved the following.

3.3.25 **Proposition**. *Given a map $f : X \longrightarrow Y$, the subspace $Y \subset M_f$ is a strong deformation retract of the cylinder M_f, with retraction $r : M_f \longrightarrow Y$ given by*

$$r(z) = \begin{cases} f(x) & \text{if } z = q(x,s)\,,\,(x,s) \in X \times I\,, \\ y & \text{if } z = q(y)\,,\,y \in Y\,. \end{cases}$$

□

Therefore, the inclusion $i : Y \hookrightarrow M_f$ is a homotopy equivalence with inverse r. Moreover, if we take the inclusion $j : X \hookrightarrow M_f$ given by $j(x) = q(x,1)$, then $f = r \circ j$; that is, up to homotopy, every map f can be decomposed as an inclusion† followed by a homotopy equivalence. We have, in particular, the following result.

3.3.26 **Theorem**. *A map $f : X \longrightarrow Y$ is a homotopy equivalence if and only if X (or more precisely, $q(X \times \{1\})$), is a (strong) deformation retract of the mapping cylinder M_f of f.*

Proof: Let $j : X \hookrightarrow M_f$ be the inclusion given by $j(x) = q(x,1)$. Since, as we already said, the inclusion $i : Y \hookrightarrow M_f$ is a homotopy equivalence with inverse $r : M_f \longrightarrow Y$, and since $f = r \circ j$, the map f is a homotopy equivalence if and only if j is one. It is thus enough to prove that j is a homotopy equivalence if and only if X is a strong deformation retract of M_f.

Assume in the first place that X is a (strong) deformation retract of M_f. In this case, by 3.3.12, j is a homotopy equivalence.

Conversely, if $j : X \hookrightarrow M_f$ is a homotopy equivalence with inverse $g : M_f \longrightarrow X$, let $F : X \times I \longrightarrow X$ be a homotopy $g \circ j \simeq \mathrm{id}_X$ and let $G : M_f \times I \longrightarrow M_f$ be a homotopy $j \circ g \simeq \mathrm{id}_{M_f}$. Define a retraction $r : M_f \longrightarrow j(X)$ by

$$r(z) = \begin{cases} q(g(z),1) & \text{if } z = q(y),\ y \in Y, \\ q(gq(x,2s)) & \text{if } z = q(x,s),\ 0 \le s \le \frac{1}{2}, \\ q(F(x,2s-1),1) & \text{if } z = q(x,s),\ \frac{1}{2} \le s \le 1. \end{cases}$$

With this retraction we define a deformation $H : M_f \times I \longrightarrow M_f$ by

$$H(z,t) = \begin{cases} G(z,1-2t) & \text{if } t \le \frac{1}{2}, \\ rG(z,2t-1) & \text{if } t \ge \frac{1}{2}. \end{cases}$$

This is a deformation retraction (it is not yet a strong deformation and we have to transform it into one. This can be done using the Lemma 3.3.27 below, where we take $Z = M_f$ and identify X with $j(X)$.) □

†This inclusion is particularly “decent,” since it is a cofibration (cf. 3.4.8 below or for details [2, 4.2.8(b)]).

If we identify X with $j(X) \subset M_f = Z$, we have that *there exists a retraction* $\sigma : Z \times I \longrightarrow (Z \times \{0\}) \cup (X \times I)$, defined by

$$\sigma(q(x,s),t) = \begin{cases} (q(x, \frac{2s}{2-t}), 0) & \text{if } 0 \le s \le \frac{2-t}{2}, \\ (q(x,1), \frac{2s+t-2}{2s-1}) & \text{if } \frac{2-t}{2} \le s \le 1, \end{cases}$$
$$\sigma(q(y),t) = (q(y),0),$$

where $q : (X \times I) \sqcup Y \longrightarrow M_f$ is the canonical quotient map, and $x \in X$, $s \in I$, and $y \in Y$. Therefore, by 3.1.19, there exists a retraction $\rho : Z \times I \times I \longrightarrow (Z \times I \times \{1\}) \cup (W \times I)$.

3.3.27 **Lemma**. *Let $X \subset Z$ be such that there exists a retraction $r : Z \longrightarrow X$. If there exists a deformation retraction $H : Z \times I \longrightarrow Z$ of Z in X, then there exists a strong deformation retraction $L : Z \times I \longrightarrow Z$ of Z in X.*

Proof: Let $r : Z \longrightarrow X$ be the retraction associated to H, namely $r(z) = H(z,1) \in X$, and take $W = (Z \times \{0\}) \cup (X \times I) \cup (Z \times \{1\}) \subset Z \times I$. Define $G : (Z \times I \times \{1\}) \cup (W \times I) \longrightarrow Z$ by

$$\begin{aligned} G(z,0,t) &= z \\ G(x,s,t) &= H(x,st) \\ G(z,1,t) &= H(r(z),t) \\ G(z,s,1) &= H(z,s)\,, \end{aligned}$$

where $z \in Z$, $x \in X$, $s,t \in I$. The homotopy $L : Z \times I \longrightarrow Z$ given by

$$L(z,t) = G\rho(z,t,0),$$

where $\rho : Z \times I \times I \longrightarrow (Z \times I \times \{1\}) \cup (W \times I)$, as above, is a strong deformation retraction of Z in X, as one may verify directly. □

Let X and Y be two topological spaces. If there is a homotopy equivalence $f : X \longrightarrow Y$, then by 3.3.26, X is a strong deformation retract of the space $Z = M_f$. Moreover, Y is always a strong deformation retract of Z. Conversely, if X and Y are each a strong deformation retract of some space Z, then, since both are homotopy equivalent to Z, they are homotopy equivalent to each other. Thus we have proved the following.

3.3.28 **Corollary**. *Two topological spaces X and Y are homotopy equivalent if and only if there exists a space Z that contains both as strong deformation retracts.*

□

3.3.29 EXERCISE. Let X and Y be homotopy equivalent topological spaces. Prove or disprove by a counterexample each of the following assertions:

(a) X connected $\Longrightarrow Y$ connected.

(b) X path connected $\Longrightarrow Y$ path connected.

(c) X compact $\Longrightarrow Y$ compact.

(d) X Hausdorff $\Longrightarrow Y$ Hausdorff.

State corresponding assertions for other topological properties and prove or disprove them.

3.3.30 EXERCISE. Prove that $X_1 \simeq Y_1$, $X_2 \simeq Y_2 \Longrightarrow X_1 \times X_2 \simeq Y_1 \times Y_2$. Is this fact also true for infinite products?

3.3.31 EXERCISE. We know by 3.3.19, that if we remove a point from the torus, then we obtain a space of the same homotopy type of the wedge sum of two circles. What can be said if we remove two points instead? And what about more than two?

3.3.32 EXERCISE. Prove that if we remove from $\mathbb{R}^3$ the unit circle in $\mathbb{R}^2 \subset \mathbb{R}^3$, then the remaining space is homotopy equivalent to the wedge sum $\mathbb{S}^1 \vee \mathbb{S}^2$.

3.3.33 EXERCISE. Prove that if we remove from $\mathbb{S}^n$ the unit circle $\mathbb{S}^1 \subset \mathbb{S}^n$, canonically embedded, then the remaining space is homotopy equivalent to $\mathbb{S}^{n-2}$.

3.3.34 EXERCISE. Prove that $X \simeq X'$, $Y \simeq Y' \Longrightarrow [X, Y] \approx [X', Y']$.

3.3.35 EXERCISE. Consider the set

$$\mathcal{H}(X) = \{[f] \mid f : X \longrightarrow X \text{ is a homotopy equivalence}\}.$$

Prove that the operation given by the composition, $[f] \cdot [g] = [f \circ g]$ turns this set into a group. Verify that the group $\mathcal{H}(\mathbb{S}^1)$ es cyclic of order 2, namely that it is isomorphic to $\mathbb{Z}/2$.

3.3.36 EXERCISE. What particular space is the mapping cylinder of the reflection $g_{-1} : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$?

3.3.37 EXERCISE. The space of $n \times n$ matrices with real coefficients $\mathrm{M}_n(\mathbb{R})$, or with complex coefficients $\mathrm{M}_n(\mathbb{C})$, is homeomorphic to the euclidian space $\mathbb{R}^{n^2}$, or $\mathbb{C}^{n^2} = \mathbb{R}^{4n^2}$. The determinant $\det : \mathrm{M}_n(\mathbb{R}) \longrightarrow \mathbb{R}$, or $\det : \mathrm{M}_n(\mathbb{C}) \longrightarrow \mathbb{C}$, is a continuous function, hence, the subspaces

$$\mathrm{GL}_n(\mathbb{R}) = \det^{-1}(\mathbb{R} - 0), \qquad \mathrm{GL}_n(\mathbb{C}) = \det^{-1}(\mathbb{C} - 0)$$

are open sets. These spaces consist of the invertible matrices, and together with matrix multiplication, they constitute groups called the *general linear group of real $n \times n$ matrices* and *general linear group of complex $n \times n$ matrices* (see 2.4.1). Let $\mathrm{O}_n \subset \mathrm{GL}_n(\mathbb{R})$ and $\mathrm{U}_n \subset \mathrm{GL}_n(\mathbb{C})$ be the *orthogonal* and the *unitary* groups; that is, $M \in \mathrm{O}_n$, resp. $M \in \mathrm{U}_n$, if and only if $MM^* = 1$, where M^* represents the transposed matrix, resp. the conjugate transposed matrix, of M, and 1 is the identity matrix. Prove that O_n and U_n are strong deformation retracts of $\mathrm{GL}_n(\mathbb{R})$ and $\mathrm{GL}_n(\mathbb{C})$, respectively. (*Hint:* Use the the Gram-Schmidt orthonormalization process.)

3.4 HOMOTOPY EXTENSION

In this section, we shall apply the Tietze-Urysohn extension theorem to prove some results about extension of homotopies. The Tietze-Urysohn theorem [21, 9.1.32] is as follows. The contents are inspired by tom Dieck's book [7].

3.4.1 **Theorem**. *Let X be a normal space and take a closed set $A \subset X$. Consider a family of intervals $\{I_\lambda\}_{\lambda\in\Lambda}$ in $\mathbb{R}$ and take the product $Y = \prod_\lambda I_\lambda$. Then every continuous map $F : A \longrightarrow Y$ admits a continuous extension $f : X \longrightarrow Y$. (In particular, Y can be taken as $\mathbb{R}^n$.)* □

As a consequence of this, we obtain the next.

3.4.2 **Theorem**. *Let X be a normal space and take a closed set $A \subset X$. If $f : A \longrightarrow \mathbb{S}^n$ is continuous, then there is an open neighborhood V of A and a continuous extension $g : V \longrightarrow \mathbb{S}^n$ of f.*

Proof: Consider the composite $f' : A \xrightarrow{f} \mathbb{S}^n \hookrightarrow \mathbb{R}^{n+1}$. By the Tietze–Urysohn theorem, there is a continuous extension $F : X \longrightarrow \mathbb{R}^{n+1}$ of f'. Put $V = F^{-1}(\mathbb{R}^{n+1} - \{0\}$ and define $g : V \longrightarrow \mathbb{S}^n$ by

$$g(x) = \frac{F(x)}{|F(x)|}\,.$$

This is the desired extension. □

Clearly not every map f can be continuously extended to all of X. For instance, if $X = \mathbb{R}^{n+1}$, $A = \mathbb{S}^n$, and $f = \mathrm{id}_{\mathbb{S}^n}$, then there is no extension, since otherwise f would be nullhomotopic.

We have the following.

3.4.3 **Theorem**. *Let X and $X \times I$ be normal spaces, let $A \subset X$ be closed. Then every map $h : X \times \{0\} \cup A \times I \longrightarrow \mathbb{S}^n$ has a continuous extension $H : X \times I \longrightarrow \mathbb{S}^n$.*

Proof: By the previous result, there is a neighborhood V of $X \times \{0\} \cup A \times I$ in $X \times I$ and an extension $g : V \longrightarrow \mathbb{S}^n$. Since $X \times I$ is normal, by the Urysohn lemma [21, 9.1.25] there is a function $f : X \times I \longrightarrow I$ such that $f|_{A \times I} = 1$ and $f|_{X \times I - V} = 0$. Put $s(x) = \min\{f(x,t) \mid t \in I\}$. Hence for all $x \in X$, the point $(x, s(x)t)$ lies inside V. If we define $H(x,t) = g(x, s(x)t)$, then H is the desired extension. □

The following result, which reformulates the previous one, states that if $A \subset X$ is a closed subset of a normal space, which has the property that $X \times I$ is also normal, then the pair (X, A) has the *homotopy extension property* –HEP for short– with respect to the spheres.

3.4.4 **Theorem**. *Let X and $X \times I$ be normal spaces, let $A \subset X$ be closed and let $f : X \longrightarrow \mathbb{S}^n$ be continuous. Given a homotopy $H : A \times I \longrightarrow \mathbb{S}^n$ such that $H(a,0) = f(a)$ for all $a \in A$, then H can be extended to another homotopy $K : X \times I \longrightarrow \mathbb{S}^n$ such that $K(x,0) = f(x)$ for all $x \in X$. In a diagram*

$$\begin{tikzcd}
 & X \arrow[dr, "j_0"] \arrow[drr, bend left, "f"] & & \\
A \arrow[ur, hook, "i"] \arrow[dr, "j_0"'] & & X \times I \arrow[r, dashed, "K"] & \mathbb{S}^n \\
 & A \times I \arrow[ur, hook, "i \times \mathrm{id}"'] \arrow[urr, bend right, "H"'] & &
\end{tikzcd}$$

In particular, if $g : A \longrightarrow \mathbb{S}^n$ is nullhomotopic, then g can be continuously extended to a map $h : X \longrightarrow \mathbb{S}^n$. □

3.4.5 DEFINITION. A subset $E \subseteq \mathbb{R}^n$ is called a *Euclidean neighborhood retract*, or *ENR* for short, if there is an open neighborhood V of E and a retraction $r : U \longrightarrow E$.

Clearly Theorem 3.4.4 remains true if instead of $\mathbb{S}^n$ we put an ENR E.

3.4.6 **Theorem**. *Let $A \subset \mathbb{S}^{n+1}$ be closed and let $f : A \longrightarrow \mathbb{S}^n$ be continuous. Then there is an extension $g : \mathbb{S}^{n+1} - F \longrightarrow \mathbb{S}^n$ of f, where F is a finite subset of $\mathbb{S}^{n+1}$.*

Proof: This proof makes use of some results on smooth maps, which can be seen in [8]. Take $y \in \mathbb{R}^{n+1}$ such that $|y| < 1$. There is a retraction $r : \mathbb{R}^{n+1} - \{y\} \longrightarrow \mathbb{S}^n$. We may consider f as a map $A \longrightarrow \mathbb{R}^{n+1} - \{y\}$, so that we only need to find an extension $F : \mathbb{S}^{n+1} - F \longrightarrow \mathbb{R}^{n+1} - \{y\}$. By Theorem 3.4.4, we only need to extend some map g which is homotopic to f. By Theorem 3.4.1, there is a continuous extension $F : \mathbb{S}^{n+1} \longrightarrow \mathbb{R}^{n+1}$ of f. Now we may take a C^∞-map $G : \mathbb{S}^{n+1} \longrightarrow \mathbb{R}^{n+1}$ such that $|G(x) - f(x)| \leq \frac{1}{2}$ for all $x \in \mathbb{S}^{n+1}$. This is possible by applying the Weierstrass approximation theorem (see??????). Let $y \in \mathbb{R}^{n+1}$ be a regular value of G such that $|y| \leq \frac{1}{2}$, which exists by Sard's theorem (see?????). Therefore the set $F = G^{-1}(y)$ is finite and $G : \mathbb{S}^{n+1} - F \longrightarrow \mathbb{R}^{n+1} - \{y\}$ is an extension of $g = G|_A$. But by the conditions imposed, the linear homotopy $H : \mathbb{S}^{n+1} - F \longrightarrow \mathbb{R}^{n+1} - \{y\}$ given by $H(x,t) = (1-t)f(x) + tg(x)$ is well defined.

One may complete the previous theorem giving a more precise description of the exceptional set F.

3.4.7 **Theorem**. *One may choose the set F in 3.4.6 so that it has at most one point in each component of the complement $\mathbb{S}^{n+1} - A$.*

Proof: Let us first delete a point z from $\mathbb{S}^{n+1}$, and we consider $\mathbb{S}^{n+1} - \{z\}$ as if it were $\mathbb{R}^{n+1}$. Under this convention, take

$$B_\varepsilon(u) = \{x \in \mathbb{R}^{n+1} \mid |u - x| \leq \varepsilon\} \subset \mathbb{R}^{n+1} - A\,,$$

so that F does not meet the boundary $S_\varepsilon(u)$ of $B_\varepsilon(u)$. Restrict the map F to $\mathbb{S}^{n+1} - (B^\circ_\varepsilon(u) - F)$. For any point $x \in B^\circ_\varepsilon(u)$, there is a retraction $r : B_\varepsilon(u) - \{x\} \longrightarrow S_\varepsilon(u)$. Hence we may extend $F|_{\mathbb{S}^{n+1} - B^\circ_\varepsilon(u)}$ to $\mathbb{S}^{n+1} - (B^\circ_\varepsilon(u) \cup F) \cup (B_\varepsilon(u) - \{x\}$ by applying the map $r \circ F$ on $B_\varepsilon(u) - \{x\}$. Hence if F is an extension of f, so is also the newly constructed map.

The described extension process can be used in two ways. If there are several points of F in $B^\circ_\varepsilon(u)$, then the only missing point is $x \in B_\varepsilon(u)$, and thus we have reduced the set F. If $x, y \in F$ lie in the same component of $\mathbb{S}^{n+1} - A$, then we can find a finite system of points $x = x_0, \ldots, x_k = y$ in this component, such that for each i, the points x_i and x_{i+1} lie in some adequate ball $B_\varepsilon(u)$. First we may enlarge F by adding the points $x_1, \ldots, x_{k-1}$, and then successively suppressing $x_0, \ldots, x_{k-1}$. Repeating this process we get the desired result.

3.4.8 Definition. An inclusion $i : A \hookrightarrow X$ is said to have the *homotopy extension property* –HEP for short– for a class $\mathcal{C}$ of topological spaces if given a homotopy $H : A \times I \longrightarrow Z$, where $Z \in \mathcal{C}$, such that $H(a, 0) = f(a)$ for all $a \in A$, then H can be extended to another homotopy $K : X \times I \longrightarrow Z$ such that $K(x,0) = f(x)$ for all $x \in X$. In a diagram

$$\begin{tikzcd}
 & X \arrow[dr, "j_0"] \arrow[rr, bend left, "f"] & & \\
A \arrow[ur, hook, "i"] \arrow[dr, "j_0"'] & & X \times I \arrow[r, dashed, "K"] & Z \\
 & A \times I \arrow[ur, hook, "i\times\mathrm{id}"'] \arrow[urr, bend right, "H"'] & &
\end{tikzcd}$$

i is said to be a *cofibration* if it has the HEP for the class of all topological spaces. The cofibration is c *closed* if $A \subset X$ is a closed subset.

3.4.9 EXERCISE. Show that if $i : A \hookrightarrow X$ is a closed cofibration if and only if there is a retraction $r : X \times I \longrightarrow (X \times \{0\}) \cup (A \times I)$. (*Hint:* Consider the diagram

$$\begin{tikzcd}
 & X \arrow[dr, "j_0"] \arrow[rr, bend left, "i'"] & & \\
A \arrow[ur, hook, "i"] \arrow[dr, "j_0"'] & & X \times I \arrow[r, dashed, "r"] & (X \times \{0\}) \cup (A \times I) \\
 & A \times I \arrow[ur, hook, "i\times\mathrm{id}"'] \arrow[urr, bend right, "j'"'] & &
\end{tikzcd}$$

where $i' : X \longrightarrow X \times \{0\} \cup (A \times I)$ is the embedding given by $x \mapsto (x,0)$ and $j' : A \times I \longrightarrow (X \times \{0\}) \cup (A \times I)$ is the inclusion. Conversely, given f and H as in the definition, define $K = (f,H) \circ r$, where $(f,H)(x,0) = f(x)$ if $(x,0) \in X \times \{0\}$ and $(f,H)(a,t) = H(a,t)$ if $(a,t) \in A \times I$.)

The following result characterizes a cofibration (see 3.1.19) using a local variant of the concept of a deformation retraction given in 3.3.9.

3.4.10 **Theorem**. *An inclusion $i : A \hookrightarrow X$ is a cofibration if and only if there is a function $u : X \longrightarrow \mathbb{R}^+ = \{t \in \mathbb{R} \mid t \geq 0\}$ and a homotopy $\varphi : X \times I \longrightarrow X$ such that*

(1) $A \subset u^{-1}(0)$,

(2) $\varphi(x,0) = x$ *for all* $x \in X$,

(3) $\varphi(a,t) = a$ *for all* $a \in A$ *and all* $t \in I$,

(4) $\varphi(x,t) \in A$ *for all* $x \in X$ *and all* $t > u(x)$.

Proof: If i is a cofibration, then by 3.4.9 there is a retraction $r : X \times I \longrightarrow (X \times \{0\}) \cup (A \times I)$. Define u and φ by

$$u(x) = \max\{t - \mathrm{proj}_I r(x,t) \mid t \in I\} \quad \text{and} \quad \varphi(x,t) = \mathrm{proj}_X r(x,t)\,.$$

Conversely, given u and φ define the retraction $r : X\times I \longrightarrow (X\times\{0\})\cup(A\times I)$ by

$$r(x,t) = \begin{cases} (\varphi(x,t),0)\,, & \text{if } t \le u(x) \\ (\varphi(x,t),t-u(x))\,, & \text{if } t \ge u(x) \end{cases}$$

We finish this section with the following result about the product of two cofibrations.

3.4.11 **Theorem**. *Let $i : A \hookrightarrow X$ and $j : B \hookrightarrow Y$ be cofibrations so that $A \subset X$ is closed. Then $k : (A\times Y)\cup(X\times B) \hookrightarrow X\times Y$ is a cofibration.*

Proof: By Theorem 3.4.10, for i we have a function $u : X \longrightarrow \mathbb{R}^+$ and a map $\varphi : X\times I \longrightarrow X$ and for j we have a function $v : Y \longrightarrow \mathbb{R}^+$ and a map $\psi : Y\times I \longrightarrow Y$, which fulfill (1)–(4). Define a function

$$w : X\times Y \longrightarrow \mathbb{R}^+ \quad \text{by} \quad (x,y) \longmapsto \min\{u(x),v(y)\}\,,$$

and define a map

$$\eta : X\times Y\times I \longrightarrow X\times Y \text{ by } (x,y,t) \longmapsto (\varphi(x,\min\{t,v(y)\}),\psi(y,\min\{t,u(x)\}))\,.$$

One may now verify that w and η satisfy (1)–(4) for $i\times j$:

(1) Take $(a,y) \in A\times Y$. Then

$$w(a,y) = \min\{u(a),v(y)\} = \min\{0,v(y)\} = 0\,.$$

Now take $(x,b) \in X\times B$. Then

$$w(x,b) = \min\{u(x),v(b)\} = \min\{u(x),0\} = 0\,.$$

Hence $(A\times Y)\cup(X\times B) \subset w^{-1}(0)$.

(2) For $(x,y) \in X\times Y$ one has

$$\begin{aligned}\eta(x,y,0) &= (\varphi(x,\min\{0,v(y)\}),\psi(y,\min\{0,u(x)\})) \\ &= (\varphi(x,0),\psi(y,0)) = (x,y)\,.\end{aligned}$$

(3) Take $(a,y,t) \in A\times Y\times I$. Then

$$\begin{aligned}\eta(a,y,t) &= (\varphi(a,\min\{t,v(y)\}),\psi(y,\min\{t,u(a)\})) \\ &= (a,\psi(y,0)) = (a,y) \quad \text{since } u(a)=0\,.\end{aligned}$$

Now take $(x,b,t) \in X\times B\times I$. Then

$$\begin{aligned}\eta(x,b,t) &= (\varphi(x,\min\{t,v(b)\}),\psi(b,\min\{t,u(x)\})) \\ &= (\varphi(x,0),b) = (x,b) \quad \text{since } v(b)=0\,.\end{aligned}$$

(4) Suppose

$$t > w(x,y) = \min\{u(x), v(y)\} = \begin{cases} u(x) & \text{if } u(x) \leq v(y) \\ v(y) & \text{if } v(y) \leq u(x) \end{cases}$$

In the first case,

$$\min\{t, v(y)\} \geq u(x) \text{“} \Rightarrow \text{''} \ \varphi(x, \max\{t, v(y)\}) \in A\,,$$

and in the second case,

$$\min\{t, u(x)\} \geq v(y) \text{“} \Rightarrow \text{''} \ \psi(y, \max\{t, u(x)\}) \in B\,.$$

In any case we have

$$\eta(x,y,t) = (\varphi(x, \max\{t, v(y)\}), \psi(y, \max\{t, u(x)\})) \in (A \times Y) \cup (X \times B)\,.$$

3.4.12 EXERCISE. Show that given two closed cofibrations $A \hookrightarrow X$ and $B \hookrightarrow X$, then $A \cap B \hookrightarrow X$ is a closed cofibration.

3.4.13 EXERCISE. Assume that the following

$$\begin{tikzcd} A \arrow[r, "i"] \arrow[d, "f"'] & X \arrow[d, "g"] \\ B \arrow[r, "j"'] & Y \end{tikzcd}$$

is a pushout diagram. Show that if i is a cofibration, then j is also a cofibration. Furthermore, show that if f is a homotopy equivalence, then so is g. Conclude that if A is contractible, then the quotient map $q : X \longrightarrow X/A$ is a homotopy equivalence

3.4.14 EXERCISE. Let $i : A \hookrightarrow X$ be a cofibration and let $H : X \times I \longrightarrow Y$ be a homotopy. Assume furthermore that $K : (A \times I) \times I \longrightarrow Y$ is another homotopy starting with $H \circ (i \times \mathrm{id}_X)$ and ending with a constant homotopy $K' : A \times I \longrightarrow Y$ rel $A \times \{0,1\}$, namely $K(a,t,1) = K'(a,0) = K(a,0,1)$. Show that there is a homotopy $H' : H_0 \simeq H_1 : X \longrightarrow Y$, where $H_\nu(x) = H(x,\nu)$, $\nu = 0,1$, such that H' remains constant on A.

3.4.15 EXERCISE. Assume that $i : A \hookrightarrow X$ is a cofibration and that $r : X \longrightarrow A$ is a strong deformation retraction. Take a strong deformation $H : X \times I \longrightarrow X$, that is $H(x,0) = x$, $H(x,1) = r(x)$, and $H(a,t) = a$, for $x \in X$, $a \in A$, and $t \in I$, compose with the reverse homotopy of $H|_{A \times I}$, and apply Exercise 3.4.14. Explain what happens!

3.5 DOMAIN INVARIANCE

In this section we shall prove the domain invariance theorem 1.1.8 stated in Chapter 1. This theorem an its consequences about the dimension and the boundary invariance theorems are due to Brouwer. The proof given here uses essentially the separation property of disjoint compact sets that $\mathbb{R}^n$ has (and which is a consequence of the Hausdorff separation axiom replacing two points by two disjoint compact sets), as well as the retraction theorem 3.2.24 for general n, of which we only proved the case $n = 2$. Other proofs make use of some algebraic topology results.

We start with some elementary concepts. Consider a subspace $X \subset \mathbb{R}^n$ and a point $x_0 \in \mathbb{R}^n - X$. We define the *winding map* by

$$w_{X,x_0} = w_{x_0} : X \longrightarrow \mathbb{S}^{n-1}\,, \quad x \longmapsto \frac{x - x_0}{|x - x_0|}\,.$$

If X is compact, then $\mathbb{R}^n - X$ decomposes in open path-components. Only one of these components is unbounded. In this case, $X \subset B_r(0) = \{x \in \mathbb{R}^n \mid |x| \leq r,$ and since the complement $\mathbb{R}^n - B_r(0)$ is connected, then it is contained in one component V of $\mathbb{R}^n - X$, which is the unbounded one, all other components lie inside the ball $B_r(0)$.

3.5.1 **Proposition**. *Let $X \subset \mathbb{R}^n$ be compact. Then the following hold:*

(a) *A given point $x_0 \in \mathbb{R}^n - X$ lies in the unbounded component if and only if the winding map w_{X,x_0} is nullhomotopic.*

(b) *Two points $x, y \in \mathbb{R}^n - X$ lie in the same component if and only if the winding maps $w_{X,x}$ and $w_{X,y}$ are homotopic.*

Proof: (a) Consider an affine homeomorphism $a : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ of the form $a(x) = rx + b$, where $r > 0$ and $b \in \mathbb{R}^n$. We use a to transform X to $Y = a(X)$ and x_0 to $y_0 = a(x_0)$. Then clearly $w_{X,x_0} = w_{Y,y_0} \circ a$. Since a is a homeomorphism, w_{X,x_0} is nullhomotopic if and only if w_{Y,y_0} is nullhomotopic. Also x_0 lies in the unbounded component of $\mathbb{R}^n - X$ if and only if y_0 lies in the unbounded component of $\mathbb{R}^n - Y$. Choosing r large enough and b conveniently, we may assume that $x_0 = 0$ and $X \subset \mathbb{B}^n$ ($\mathbb{B}^n$ the unit ball).

If we now suppose that x_0 lies in the unbounded component K of $\mathbb{R}^n - X$, then there is a path $\sigma : x_0 = 0 \simeq \sigma(1) = z_0 \in \mathbb{R}^n - \mathbb{B}^n$. Consider the homotopy

$$H : X \times I \longrightarrow \mathbb{S}^{n-1}\,, \quad (x,t) \longmapsto \frac{x - \lambda(t)}{|x - \lambda(t)|}\,.$$

Then $H_1(x) = H(x,1) \neq q = |z_0|^{-1}z_0$ and thus its image lies in the complement $\mathbb{S}^{n-1} - \{z_0\}$, which is contractible. Hence H_1 is nullhomotopic.

Conversely, if the component K where 0 lies, is bounded, then $\overline{K} \subset \mathbb{B}^n$ and $K \cup X$ is closed, since its complement is a union of components of $\mathbb{R}^n - X$, which by 1.3.6 are open. If $w = w_{X,0}$ is nullhomotopic, then by 3.4.4 we can extend w to a map $W : K \cup X \longrightarrow \mathbb{S}^{n-1}$. Define $r : \mathbb{B}^n \longrightarrow \mathbb{S}^{n-1}$ by

$$r(x) = \begin{cases} W(x) & \text{if } x \in K \cup X \\ |x|^{-1}x & \text{if } x \in \mathbb{B}^n - K\,. \end{cases}$$

Since both parts of the definition coincide in X, then r is well defined and continuous. Furthermore $r|_{\mathbb{S}^{n-1}} = \mathrm{id}_{\mathbb{S}^{n-1}}$, hence it is a retraction, and thus contradicts the retraction theorem 3.2.24.

(b) Assume that x and y lie in different components. We may decompose $\mathbb{R}^n - X$ as a disjoint union $A \cup B$, where (A, B) is a separation such that $x \in A$ and $y \in B$. One of the two sets A or B is bounded, let us say that it is A. The map $w_y|_X$ is defined on $\mathbb{R}^n - \{y\}$, thus in particular it is defined on on $A \cup X \subset \mathbb{R}^n - \{y\}$. But at the end of the proof of part (a) we saw that one cannote extend $w_x|X$ to $A \cup X$. But if w_x and w_y were homotopic, then by 3.4.4, both maps could be simultaneously extended. □

We shall now prove a duality theorem for subsets of a sphere.

3.5.2 **Theorem**. *Let $X \subset \mathbb{S}^n$ be a closed subset which is different from $\mathbb{S}^n$. Then the complement $\mathbb{S}^n - X$ is connected if and only if any map $f : X \longrightarrow \mathbb{S}^{n-1}$ is nullhomotopic.*

Proof: If $\mathbb{S}^n - X$ is disconnected, then take points $x, y \in \mathbb{S}^n - X$ which lie in different components. $\mathbb{S}^n - \{y\}$ is homeomorphic to $\mathbb{R}^n$, and, up to the homeomorphism, x lies in the bounded component. By 3.5.1, there is a nullhomotopic map $f : X \longrightarrow \mathbb{S}^{n-1}$.

Conversely, if $\mathbb{S}^n - X$ is connected and $f : X \longrightarrow \mathbb{S}^{n-1}$ is continuous, then by 3.4.7, f can be extended to a continuous map $g : \mathbb{S}^n - \{x\} \longrightarrow \mathbb{S}^{n-1}$. Since $\mathbb{S}^n - \{x\}$ is contractible, g and thus f too, is nullhomotopic. □

Under the same assumptions, we have the following consequences.

3.5.3 **Corollary**. (a) *If X is contractible, then $\mathbb{S}^n - X$ is connected.*

(b) *If X is is homeomorphic to $\mathbb{S}^{n-1}$, then there are nonnullhomotopic maps $X \longrightarrow \mathbb{S}^{n-1}$. Hence $\mathbb{S}^n - X$ has at least two components.*

Proof: (a) is immediate and (b) follows from 3.2.26.

3.5.4 **Theorem**. *Consider $A \subset X \subset \mathbb{S}^n$ such that the pair (X, A) is homeomorphic to the pair $(\mathbb{B}^n, \mathbb{S}^{n-1})$. Then the complement $\mathbb{S}^n - A$ consists of two components, namely $\mathbb{S}^n - X$ and $X - A$.*

Proof: By Corollary 3.5.3 (a), $\mathbb{S}^n - X$ is connected. On the other hand, the complement $X - A$ is homeomorphic to $\mathbb{B}^n - \mathbb{S}^n$ and hence it is connected. But $\mathbb{S}^n - A = (\mathbb{S}^n - X) \cup (X - A)$, and so the two sets $(\mathbb{S}^n - X)$ and $(X - A)$ must be the components of $\mathbb{S}^n - A$.

3.5.5 **Theorem**. (Domain invariance) *Take subsets $X, Y \subset \mathbb{S}^n$ such that X is homeomorphic to Y. Then, if X is open in $\mathbb{S}^n$, Y must be open too.* □

Proof: Let $\varphi : X \longrightarrow Y$ be a homeomorphism, and take $y \in Y$. If $x \in X$ is such that $y = \varphi(x) \in Y$, then take a neighborhood U of x such that $\overline{U}$ is compact and contained in X, as well as $(\overline{U}, \partial U) \approx (\mathbb{B}^= n, \mathbb{S}^{n-1})$. Put $(W, B) = (\varphi(\overline{U}), \varphi(\partial U))$. By 3.5.4, $W - B$ is open. Since $y \in W - B \subset Y$, y is an interior point of Y and thus Y is open.

As a consequence of the previous result, we obtain the domain invariance theorem for manifolds.

3.5.6 **Theorem**. (Domain invariance for manifolds) *Let M and N be topological manifolds of dimension n, and take subsets $X \subset M$, $Y \subset N$ such that X is homeomorphic to Y. Then, if X is open in M, Y must be open in N.*

Proof: Let $\varphi : X \longrightarrow Y$ be a homeomorphism, and take $y \in Y$. If $x \in X$ is such that $y = \varphi(x) \in Y$, then take neighborhoods U of x in M and V of y in N such that U is contained in X, V is contained in Y, $\varphi(U) \subset V$, and both U and V are homeomorphic to $\mathbb{R}^n$. Let $\xi : U \longrightarrow \mathbb{S}^n$ and $\eta : V \longrightarrow \mathbb{S}^n$ be embeddings with open images $\xi(U) = U'$ and $\eta(V) = V'$. Hence we have the the map $\eta \circ \varphi \circ \xi^{-1} : U' \longrightarrow V'$ is a homeomorphism of U' onto a set $V'' \subset V'$. By Theorem 3.5.5, V'' is open in $\mathbb{S}^n$, and so it is open in V' too. Hence $\varphi\xi^{-1}(U') = \varphi(U)$ is open in V, and thus in N too. Since $y \in \varphi(U) \subset N$, one has that Y is open.

Chapter 4 The fundamental group

The fundamental group is probably the most important concept of algebraic topology. This will be the first properly algebraic invariant of a topological space to be studied in this book. We shall associate to a topological space this group, which in general is not abelian and whose structure provides us with valuable information about the space.

4.1 Definition and general properties

In this section we start giving the definition of the fundamental group, which in the beginning depends on the basic concept of a path inside a topological space.

4.1.1 Definition. Let X be a topological space and take points $x_0, x_1 \in X$. A *path* from x_0 to x_1 is a map $\omega : I \longrightarrow X$ such that $\omega(0) = x_0$ and $\omega(1) = x_1$ (see Figure 4.1). As before, we denote it by $\omega : x_0 \simeq x_1$. The point x_0 will be called the *origin* (or *beginning*) of ω, and x_1 the *destination* (or *end point*) of ω, and both will be called *extreme points* of the path. If both extreme points coincide, that is, if $x_0 = x_1$, we say that the path is *closed* or simply that it is a *loop* based at x_0.

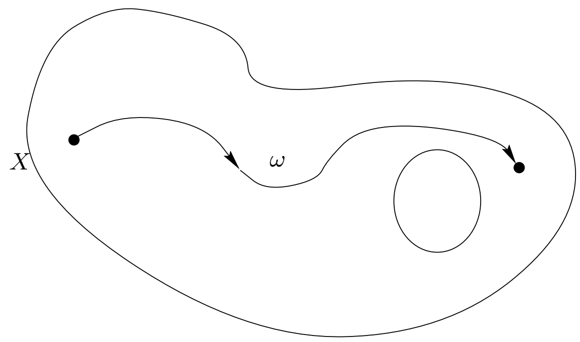


Figure 4.1 Path in a topological space

4.1.2 EXAMPLES.

(a) If $x \in X$, then $c_x : I \longrightarrow X$ given by $c_x(t) = x$ for every $t \in I$ is the *constant path* or *constant loop.*

(b) Let $ZX = X \times I$ be the cylinder over X. Then for each $x \in X$ the path $\omega_x : I \longrightarrow ZX$ given by $\omega_x(t) = (x, t)$ is the *generatrix* over x; similarly, if $CX = ZX/X \times \{1\}$ is the cone over X, the same formula for ω_x determines the generatrix over x in the cone.

(c) More generally, given $f : X \longrightarrow Y$, if M_f and C_f represent the mapping cylinder and the mapping cone of f, respectively, then the maps $\omega_x : I \longrightarrow M_f$ and $\omega_x : I \longrightarrow C_f$ given by $\omega_x(t) = \overline{(x,t)}$, where the bar represents the corresponding images in the quotient spaces, determine the *generatrices* of the cylinder and the cone.

(d) If $f : \mathbb{S}^1 \longrightarrow X$ is continuous, then $\lambda_f : I \longrightarrow X$ given by $\omega_f(t) = f(\mathrm{e}^{2\pi \mathrm{i}t})$ is the *associated loop* of f.

(e) Take $n \in \mathbb{Z}$. The path $\omega_n : I \longrightarrow \mathbb{S}^1$ given by $\omega_n(t) = \mathrm{e}^{2\pi \mathrm{i}nt}$ is the loop of *degree* n in the circle. It has the effect of wrapping around $\mathbb{S}^1$ n times (counterclockwise if $n > 0$, clockwise if $n < 0$, and if $n = 0$, it does not wrap around) as t runs along I; ω_n is the associated loop of the map $g_n : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ defined in 3.2.7(a).

(f) In the torus $T^2 = \mathbb{S}^1 \times \mathbb{S}^1$, the paths $\omega_1^1, \omega_1^2 : I \longrightarrow T^2$ given by $\omega_1^1(t) = (\mathrm{e}^{2\pi \mathrm{i}t}, 1) = (\omega_1(t), 1)$ and $\omega_1^2(t) = (1, \mathrm{e}^{2\pi \mathrm{i}t}) = (1, \omega_1(t))$ are loops, which will be called the *unitary equatorial loop* and the *unitary meridional loop.* (See 3.2.16.) More generally, we have in T^2 the loops $\omega_m^1, \omega_n^2 : I \longrightarrow T^2$ given by $\omega_m^1(t) = (\omega_m(t), 1)$ and $\omega_n^2(t) = (1, \omega_n(t))$.

Figure 4.2 shows the generators ω_1^1 and ω_1^2 in the torus.

In general, as one can see in the preceding examples as well as in Figure 4.1, as the parameter t varies from 0 to 1, the point $\omega(t)$ describes a curve or path in X connecting the points x_0 and x_1. Two paths $\omega, \sigma : I \longrightarrow X$ are *equal* if as maps they are equal, that is, if for every $t \in I$, $\omega(t) = \sigma(t)$. It is not enough that they have the same images. For instance, the loops ω_n in $\mathbb{S}^1$ defined in 4.1.2(b) are all different from each other. Given any numbers $a < b \in \mathbb{R}$ and any map $\gamma : [a, b] \longrightarrow X$, the canonical homeomorphism $I \longrightarrow [a, b]$ given by $t \mapsto (1-t)a + tb$ transforms γ into a new path $\widehat{\gamma} : I \longrightarrow X$ such that $\widehat{\gamma}(t) = \gamma((1-t)a + tb)$, so that in principle, any such map γ is canonically a path. For technical reasons, it is convenient always to assume $a = 0$ and $b = 1$.

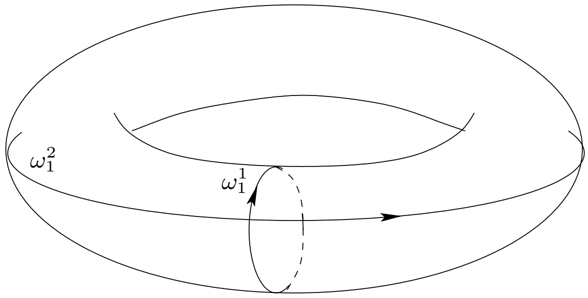


Figure 4.2 The generators ω_1^1 and ω_1^2 in the torus

4.1.3 Exercise. Prove that giving a path $\sigma : x_o \simeq x_1$ in X is equivalent to giving a homotopy $H : c_{x_0} \simeq c_{x_1} : * \longrightarrow X$, where c_x represents the map from the one-point space $*$ into X with value x.

As in the case of loops, as we saw in the last chapter, it is sometimes possible to multiply paths by each other as well as to define inverses, as we shall now see.

4.1.4 Definition. Given a path $\omega : I \longrightarrow X$, we define the *inverse path* as $\overline{\omega} : I \longrightarrow X$, where $\overline{\omega}(t) = \omega(1-t)$. If $\omega : x_0 \simeq x_1$, then $\overline{\omega} : x_1 \simeq x_0$. Two paths $\omega, \sigma : I \longrightarrow X$ are *connectable* if $\omega(1) = \sigma(0)$; in this case one can define the *product* of ω and σ as the path $\omega\sigma : I \longrightarrow X$ given by

$$(\omega\sigma)(t) = \begin{cases} \omega(2t) & \text{if } 0 \leq t \leq \frac{1}{2}, \\ \sigma(2t-1) & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

If ω is closed, namely a loop, we may define $\omega\omega$, and we denote this path by ω^2. More generally, we can in this case define ω^n as $\omega^{n-1}\omega$ for $n \geq 2$.

Nonetheless, in general, $\omega\overline{\omega} \neq c_{x_0}$, $c_{x_0}\omega \neq \omega$, etc. This bad behavior is corrected with the following definition.

4.1.5 Definition. Two paths $\omega_0, \omega_1 : I \longrightarrow X$ are said to be *homotopic* if they have the same extreme points x_0 and x_1 and there exists a *homotopy* $H : I \times I \longrightarrow X$ such that $H(s,0) = \omega_0(s)$, $H(s,1) = \omega_1(s)$, $H(0,t) = x_0$, $H(1,t) = x_1$, for every $s, t \in I$; that is, H is a homotopy relative to $\{0,1\}$. This we denote, as usual, by $H : \omega_0 \simeq \omega_1$ rel ∂I; if it is not necessary to emphasize the homotopy, then the fact that ω_0 and ω_1 are homotopic is simply denoted by $\omega_0 \simeq \omega_1$. Figure 4.3 illustrates this concept. If a loop ω is homotopic to the constant loop c_{x_0}, that is, $\omega \simeq c_{x_0}$, one says that it is *nullhomotopic* or *contractible*.

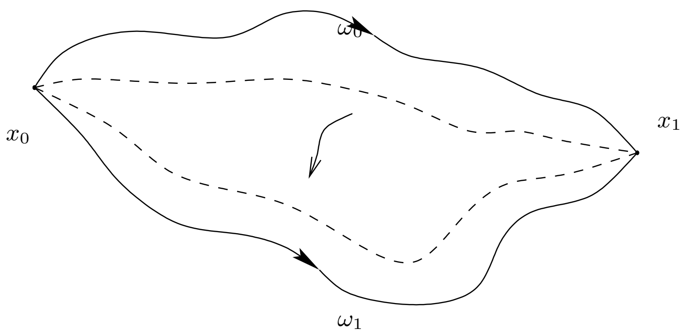


Figure 4.3 A homotopy of paths

4.1.6 Note. The notation $H : \omega \simeq \sigma$, for a homotopy of paths, that is analogous to the notation $\omega : x \simeq y$ for a path from x to y is justified, since H can be seen as a path in the function space $\mathcal{T}op(I, X)$ furnished with the compact-open topology (see [21]).

4.1.7 Exercise. Prove that, indeed, giving a homotopy $H : \omega \simeq \sigma$ (not necessarily relative to the extreme points, namely such that only $H(s, 0) = \omega(s)$ and $H(s, 1) = \sigma(s)$ hold), is equivalent to giving a path in $\mathcal{T}op(I, X)$ with origin ω and destination σ.

In relation to the comments following Definition 4.1.4, we have the following lemma.

4.1.8 **Lemma**. *Let $\omega : x_0 \simeq x_1$, $\sigma : x_1 \simeq x_2$ and $\gamma : x_2 \simeq x_3$ be paths in X. Then one has the following facts.*

(a) $\omega(\sigma\gamma) \simeq (\omega\sigma)\gamma$.

(b) $c_{x_0}\omega \simeq \omega\,, \quad \omega c_{x_1} \simeq \omega$.

(c) $\omega\overline{\omega} \simeq c_{x_0}\,, \quad \overline{\omega}\omega \simeq c_{x_1}$.

Proof:
(a) The homotopy $H : I \times I \longrightarrow X$ given by

$$H(s,t) = \begin{cases} \omega(\frac{4s}{2-t}) & \text{if } 0 \leq s \leq \frac{2-t}{4}, \\ \sigma(4s + t - 2) & \text{if } \frac{2-t}{4} \leq s \leq \frac{3-t}{4}, \\ \gamma(\frac{4s+t-3}{t+1}) & \text{if } \frac{3-t}{4} \leq s \leq 1, \end{cases}$$

is well defined and is such that $H : \omega(\sigma\gamma) \simeq (\omega\sigma)\gamma$.

(b) The homotopies $H, K : I \times I \longrightarrow X$ given by

$$H(s,t) = \begin{cases} x_0 & \text{if } 0 \le s \le \frac{1-t}{2}, \\ \omega(\frac{2s+t-1}{t+1}) & \text{if } \frac{1-t}{2} \le s \le 1, \end{cases}$$

$$K(s,t) = \begin{cases} \omega(\frac{2s}{t+1}) & \text{if } 0 \le s \le \frac{1+t}{2}, \\ x_1 & \text{if } \frac{1+t}{2} \le s \le 1, \end{cases}$$

are well defined and are such that $H : c_{x_0}\omega \simeq \omega$ and $K : \omega c_{x_1} \simeq \omega$.

(c) The homotopies $H, K : I \times I \longrightarrow X$ given by

$$H(s,t) = \begin{cases} \omega(2s(1-t)) & \text{if } 0 \le s \le \frac{1}{2}, \\ \omega(2(1-s)(1-t)) & \text{if } \frac{1}{2} \le s \le 1, \end{cases}$$

$$K(s,t) = \begin{cases} \omega(2(1-s)(1-t)) & \text{if } 0 \le s \le \frac{1}{2}, \\ \omega(2s(1-t)) & \text{if } \frac{1}{2} \le s \le 1, \end{cases}$$

are well defined and are such that $H : \omega\overline{\omega} \simeq c_{x_0}$ and $K : \overline{\omega}\omega \simeq c_{x_1}$. □

In what follows, we shall frequently write the expresion

$$\omega_1\omega_2 \cdots \omega_k ,$$

without parentheses, which, if it is not stated otherwise, means the path

$$\omega_1\omega_2 \cdots \omega_k(t) = \begin{cases} \omega_1(kt) & \text{if } 0 \le t \le \frac{1}{k}, \\ \omega_2(kt-1) & \text{if } \frac{1}{k} \le t \le \frac{2}{k}, \\ \quad\vdots & \qquad\vdots \\ \omega_k(kt-k+1) & \text{if } \frac{k-1}{k} \le t \le 1, \end{cases}$$

that is, all paths in the product are *uniformly* traveled.

Similarly to 3.1.3, we have the following.

4.1.9 **Lemma**. *The relation $\omega \simeq \sigma$ is an equivalence relation.*

Proof: The homotopy $H(s,t) = \omega(s)$ proves that $\omega \simeq \omega$.

If $H : \omega \simeq \sigma$, then $\overline{H} : I \times I \longrightarrow X$, given by $\overline{H}(s,t) = H(s, 1-t)$, is such that $\overline{H} : \sigma \simeq \omega$.

Finally, if $H : \omega \simeq \sigma$ and $K : \sigma \simeq \gamma$, then the homotopy $L : I \times I \longrightarrow X$ defined by

$$L(s,t) = \begin{cases} H(s,2t) & \text{if } 0 \le t \le \frac{1}{2}, \\ K(s,2t-1) & \text{if } \frac{1}{2} \le t \le 1, \end{cases}$$

is a homotopy relative to $\{0, 1\}$, is well defined, and satisfies $L : \omega \simeq \gamma$. □

In what follows we shall denote the equivalence class of ω by $[\omega]$ and we shall call it the *homotopy class* of ω. We are especially interested in homotopy classes of loops based at a specific point x and in particular, in the class $[c_x]$, which will be denoted by 1_x or, if there is no danger of confusion, by 1.

If $H : \omega_0 \simeq \omega_1$ and $K : \sigma_0 \simeq \sigma_1$, then the homotopy $HK : I \longrightarrow X$ given by

$$HK(s,t) = \begin{cases} H(2s,t) & \text{if } 0 \leq s \leq \frac{1}{2}, \\ K(2s-1,t) & \text{if } \frac{1}{2} \leq s \leq 1, \end{cases}$$

is well defined and is such that $HK : \omega_0\sigma_0 \simeq \omega_1\sigma_1$. Hence we may define the *product* of the homotopy classes of two connectable paths ω and σ by the formula

$$[\omega][\sigma] = [\omega\sigma]\,.$$

Using this and 4.1.8 we have the following result.

4.1.10 **Proposition**. *Let $\omega : w \simeq x$, $\sigma : x \simeq y$, and $\gamma : y \simeq z$ be paths in X. Then the following identities hold:*

(a) $[\omega]([\sigma][\gamma]) = ([\omega][\sigma])[\gamma]$.

(b) $1_w[\omega] = [\omega] = [\omega]1_x$.

(c) $[\omega][\overline{\omega}] = 1_w$, $[\overline{\omega}][\omega] = 1_x$.
(For this reason, $[\overline{\omega}]$ is denoted by $[\omega]^{-1}$.) □

Thanks to (a), we have that the product of homotopy classes of paths is associative. Hence there shall not be any confusion if one writes simply $[\omega][\sigma][\gamma]$.

4.1.11 Exercise. Prove that if $\omega_n : I \longrightarrow \mathbb{S}^1$, $n \in \mathbb{Z}$, is as in 4.1.2(b), then $[\omega_n] = [\omega_1]^n$. (*Hint:* $\omega_1^2 = \omega_2$; proceed by induction over n.)

The concept of fundamental group depends on a base point $x_0 \in X$.

If we restrict 4.1.10 to loops (closed paths), we have the following result.

4.1.12 **Theorem** and Definition. *Let (X, x_0) be a pointed space. Then the homotopy set*

$$\pi_1(X, x_0) = \{[\lambda] \mid \lambda \text{ is a loop based at } x_0\}$$

is a group with respect to the multiplication $[\lambda][\mu] = [\lambda\mu]$ with neutral element $1 = 1_{x_0} = [c_{x_0}]$ and with $[\lambda]^{-1}$ as the inverse of each $[\lambda]$. This group is called the *fundamental group* of X based at the point x_0. □

Let $f : (X, x_0) \longrightarrow (Y, y_0)$ be a pointed map. If $\lambda : I \longrightarrow X$ is a loop based at x_0, then the composite $f \circ \lambda : I \longrightarrow Y$ is a loop based at y_0. Besides, if c_{x_0} is the constant loop in X, then $f \circ c_{x_0} = c_{y_0}$ is the constant loop in Y, and given loops λ and μ in X, one has

$$f \circ (\lambda\mu) = (f \circ \lambda)(f \circ \mu)\,.$$

4.1.13 EXERCISE. Prove the last assertion in its general form, that is, if $f : X \longrightarrow Y$ is continuous and λ and μ are connectable paths in X, then $f \circ \lambda$ and $f \circ \mu$ are connectable in Y and $f \circ (\lambda\mu) = (f \circ \lambda)(f \circ \mu)$.

4.1.14 **Theorem**. *A pointed map $f : (X, x_0) \longrightarrow (Y, y_0)$ induces a group homomorphism*

$$f_* : \pi_1(X, x_0) \longrightarrow \pi_1(Y, y_0)\,,$$

given by $f_*([\lambda]) = [f \circ \lambda]$.

Proof: If $H : \lambda_0 \simeq \lambda_1 \text{ rel } \partial I$ is a homotopy of loops in X based at x_0, that is, $H(s, 0) = \lambda_0(s)$, $H(s, 1) = \lambda_1(s)$, $H(0, t) = x_0 = H(1, t)$, then clearly $f \circ H : f \circ \lambda_0 \simeq f \circ \lambda_1 \text{ rel } \partial I$, so that the function $f_*([\lambda]) = [f \circ \lambda]$ is well defined.

The remarks before the statement of the theorem prove that $f_*([\lambda\mu]) = [f \circ (\lambda\mu)] = [(f \circ \lambda)(f \circ \mu)] = f_*([\lambda])f_*([\mu])$, which shows that f_* is a group homomorphism. $\square$

In fact, the construction of the fundamental group is functorial; that is, it behaves well with respect to maps, as the following immediate result shows.

4.1.15 **Theorem**. *Let (X, x_0), (Y, y_0), and (Z, z_0) be pointed spaces and let $f : (X, x_0) \longrightarrow (Y, y_0)$ and $g : (Y, y_0) \longrightarrow (Z, z_0)$ be pointed maps. Then one has the following properties:*

(a) $\mathrm{id}_{X*} = 1_{\pi_1(X,x_0)} : \pi_1(X, x_0) \longrightarrow \pi_1(X, x_0)$.

(b) $(g \circ f)_* = g_* \circ f_* : \pi_1(X, x_0) \longrightarrow \pi_1(Z, z_0)$. $\square$

The conditions (a) and (b) above show that the correspondence

$$\begin{tikzcd} X \arrow[d, "f"'] \arrow[r, phantom, "\longmapsto"] & \pi_1(X) \arrow[d, "f_*"] \\ Y & \pi_1(Y) \end{tikzcd}$$

is a functor from the category $\mathcal{T}op_*$ of pointed spaces and pointed maps to the category $\mathcal{G}rp$ of groups and homomorphisms (see 3.1.27).

4.1.16 EXERCISE. Recall from 4.1.2(d) that given a pointed map $f : \mathbb{S}^1 \longrightarrow X$, there is an induced loop $\lambda_f : I \longrightarrow X$. Conversely, given a loop $\lambda : I \longrightarrow X$ (based at x_0), it induces a pointed map $f_\lambda : \mathbb{S}^1 \longrightarrow X$, since the exponential $I \longrightarrow \mathbb{S}^1$ is a quotient map and λ is compatible with it. Prove that this correspondence establishes a bijection

$$\pi_1(X) \cong [\mathbb{S}^1, X]_*\,,$$

where X is based at x_0 and $\mathbb{S}^1$ is based at 1. Moreover, prove that this bijection is natural, namely the diagram

$$\begin{tikzcd}
\pi_1(X) \arrow[r, "\cong"] \arrow[d, "f_*"'] & {[\mathbb{S}^1, X]_*} \arrow[d, "f_*"] \\
\pi_1(Y) \arrow[r, "\cong"'] & {[\mathbb{S}^1, Y]_*}\,,
\end{tikzcd}$$

is commutative (cf. 3.1.32(b)).

4.1.17 EXAMPLES.

(a) If $\lambda : I \longrightarrow \mathbb{R}^n$ is a loop based at 0, then the homotopy $H(s,t) = (1-t)\lambda(s)$ is a nullhomotopy. Hence $[\lambda] = 1 \in \pi_1(\mathbb{R}^n, 0)$. Therefore, $\pi_1(\mathbb{R}^n, 0) = 1$; that is, the fundamental group of $\mathbb{R}^n$ is the trivial group.

(b) As in the previous example, one can prove that $\pi_1(\mathbb{B}^n, 0) = 1$.

(c) Recall that a subset $X \subset \mathbb{R}^n$ is convex if given two points $x, y \in X$, then for every $t \in I$, $(1-t)x + ty \in X$; that is, the straight line segment joining x and y lies inside X. Given any point $x_0 \in X$ and any loop $\lambda : I \longrightarrow X$ based at x_0, the homotopy $H(s,t) = (1-t)\lambda(s) + tx_0$ is a nullhomotopy relative to ∂I. Therefore, $[\lambda] = 1 \in \pi_1(X, x_0)$. Hence the fundamental group of any convex set is trivial.

(d) If X is (strongly) contractible to $x_0 \in X$, then every loop $\lambda : I \longrightarrow X$ based at x_0 is nullhomotopic, as the nullhomotopy $H(s,t) = D(\lambda(s), t)$ shows, where $D : X \times I \longrightarrow X$ is a contraction, that is, $D(x,0) = x$, $D(x,1) = x_0 = D(x_0, t)$, $t \in I$. Therefore, $\pi_1(X, x_0) = 1$; that is, *the fundamental group of every contractible space is trivial.*

4.1.18 **Proposition**. *Let (X, x_0) and (Y, y_0) be pointed spaces. Then the function*

$$\varphi = (\mathrm{proj}_{X*}, \mathrm{proj}_{Y*}) : \pi_1(X \times Y, (x_0, y_0)) \longrightarrow \pi_1(X, x_0) \times \pi_1(Y, y_0)$$

is a group isomorphism (see 4.3.7(c)*).*

Proof: The function is clearly a homomorphism. If $\lambda : I \longrightarrow X \times Y$ is a loop satisfying $\varphi([\lambda]) = (1,1)$, then the loops $\lambda_1 = \mathrm{proj}_X \circ \lambda : I \longrightarrow X$ and $\lambda_2 = \mathrm{proj}_Y \circ \lambda : I \longrightarrow Y$ are nullhomotopic, say through the nullhomotopies $H_1 : I \times I \longrightarrow X$ and $H_2 : I \times I \longrightarrow Y$. Therefore, $H = (H_1, H_2) : I \longrightarrow X \times Y$ is a nullhomotopy of the loop $(\lambda_1, \lambda_2) = \lambda : I \longrightarrow X \times Y$. Consequently, $[\lambda] = 1$, and φ is a monomorphism.

On the other hand, if $([\lambda_1], [\lambda_2]) \in \pi_1(X, x_0) \times \pi_1(Y, y_0)$ is an arbitrary element, then the loop $\lambda = (\lambda_1, \lambda_2) : I \longrightarrow X \times Y$ is such that $\varphi([\lambda]) = ([\lambda_1], [\lambda_2])$. So φ is an epimorphism. $\square$

4.1.19 Exercise. Prove that the isomorphism given in the preceding proposition is natural (in both X and Y; cf. 3.1.34).

Up to now, we have only had explicit examples of trivial fundamental groups. In the next section we shall see examples of nontrivial fundamental groups.

In what follows we shall analyze the relationship between the fundamental groups of a space X with respect to two different base points x_0 and x_1.

If $x_0 \in X$ lies in the path component X_0 of X and λ is a loop in X based at x_0, then, since I is path connected, the image of λ lies in X_0. Moreover, if $H : \lambda \simeq \mu$ is a homotopy in X, then the image of the homotopy also lies inside X_0. These remarks establish the truth of the following statement.

4.1.20 **Proposition**. *Let X be a pointed space with base point x_0. If X_0 is the path component of X containing $x_0 \in X$, then the inclusion map $i : X_0 \hookrightarrow X$ induces an isomorphism $i_* : \pi_1(X_0, x_0) \xrightarrow{\cong} \pi_1(X, x_0)$.* $\square$

Proposition 4.1.20 allows us to restrict the analysis of the fundamental group to path-connected spaces. Indeed for such spaces the fundamental group is well defined, up to isomorphism, independent of the base point. More precisely, we have the following result.

4.1.21 **Theorem**. *Let $\omega : x_0 \simeq x_1$ be a path in X. There is an isomorphism*

$$\varphi_\omega : \pi_1(X, x_1) \longrightarrow \pi_1(X, x_0)$$

given by $\varphi_\omega([\lambda]) = [\omega][\lambda][\omega]^{-1}$.

Proof: Since λ is a loop based at x_1, ω and λ are connectable, and so also are $\omega\lambda$ and $\overline{\omega}$; therefore, the function φ_ω is well defined, and indeed it depends only on the class $[\omega]$.

To see that it is a homomorphism, we have by 4.1.10 that

$$\varphi_\omega([\lambda][\mu]) = [\omega][\lambda][\mu][\overline{\omega}] = [\omega][\lambda][\overline{\omega}][\omega][\mu][\overline{\omega}] = \varphi_\omega([\lambda])\varphi_\omega([\mu])\,.$$

Hence φ_ω is a homomorphism.

Moreover, the homomorphism $\varphi_{\overline{\omega}} : \pi_1(X, x_0) \longrightarrow \pi_1(X, x_1)$ is clearly the inverse of φ_ω. □

4.1.22 EXERCISE. Check that in fact, $\varphi_\omega \circ \varphi_{\overline{\omega}} = 1_{\pi_1(X,x_0)}$ and $\varphi_{\overline{\omega}} \circ \varphi_\omega = 1_{\pi_1(X,x_1)}$.

If in Theorem 4.1.21 we take in particular ω to be a loop based at x_0, that is, such that $[\omega] \in \pi_1(X, x_0)$, then φ_ω is precisely the *inner automorphism* of $\pi_1(X, x_0)$ given by conjugation with the element $[\omega]$.

4.1.23 REMARK. Theorem 4.1.21 allows us to write $\pi_1(X)$ for a path-connected space X without reference to the base point. Notice, however, that in general there is no canonical isomorphism between the fundamental group at two different base points. Therefore, $\pi_1(X)$ is really a family of isomorphic groups.

The concept introduced in what follows will be an important concept in this textbook, as it also is in general.

4.1.24 DEFINITION. A topological space X is said to be *simply connected* if it is path connected (0-connected) and for some base point $x_0 \in X$ the fundamental group $\pi_1(X, x_0)$ is trivial. Frequently, a simply connected space is also called *1-connected.*

The spaces given in 4.1.17 are all simply connected spaces. We have the following characterization of this concept.

4.1.25 **Proposition**. *Let X be a path-connected space. The following are equivalent.*

(a) *X is simply connected.*

(b) *$\pi_1(X, x) = 1$ for every point $x \in X$.*

(c) *Every loop $\lambda : I \longrightarrow X$ is nullhomotopic.*

(d) *$\omega \simeq \sigma$ rel ∂I for any two paths with the same extreme points x and y.*

Proof: (a) $\Leftrightarrow$ (b) follows from Theorem 4.1.21, since, because X is path connected, there is always a path $\omega : x_0 \simeq x$ in X.

(b) $\Rightarrow$ (c), for if $\lambda : I \longrightarrow X$ is a loop based at x, then $[\lambda] \in \pi_1(X, x) = 1$. Hence $[\lambda] = 1$; that is, λ is nullhomotopic.

(c) $\Rightarrow$ (d), since $\omega\overline{\sigma}$ is a loop based at x and so is nullhomotopic; that is, $\omega\overline{\sigma} \simeq c_x$. Therefore, by Lemma 4.1.8,

$$(\omega\overline{\sigma})\sigma \simeq c_x\sigma\,.$$

But by the same lemma the left-hand side is homotopic to $\omega(\overline{\sigma}\sigma) \simeq \omega$, while the right-hand side is homotopic to σ. Hence, since $\simeq$ is an equivalence relation, $\omega \simeq \sigma$.

(d) $\Rightarrow$ (a), for if $[\lambda] \in \pi_1(X, x_0)$, then since λ and c_{x_0} have the same extreme points, $\lambda \simeq c_{x_0}$; that is, $[\lambda] = 1$. Hence $\pi_1(X, x_0) = 1$, and so X is simply connected. □

Let $f, g : (X, x_0) \longrightarrow (Y, y_0)$ be homotopic maps between pointed spaces and let $H : X \times I \longrightarrow Y$ be a homotopy relative to $\{x_0\}$. If $\lambda : I \longrightarrow X$ is a loop in X based at x_0, then as we saw above, $f \circ \lambda$ and $g \circ \lambda$ are loops in Y based at y_0; moreover, the homotopy $(s, t) \mapsto H(\lambda(s), t)$ is a homotopy between the loops $f \circ \lambda$ and $g \circ \lambda$ relative to $\{0, 1\}$, i.e., $[f \circ \lambda]$ and $[g \circ \lambda]$ are the same element in $\pi_1(Y, y_0)$. Thus, we have shown the following.

4.1.26 **Proposition**. *Let $f, g : (X, x_0) \longrightarrow (Y, y_0)$ be homotopic maps of pointed spaces. Then $f_* = g_* : \pi_1(X, x_0) \longrightarrow \pi_1(Y, y_0)$.* □

Indeed, the result above has a stronger version; one has the following theorem.

4.1.27 **Theorem**. *Let $f, g : X \longrightarrow Y$ be homotopic maps and, if $H : f \simeq g$ is a homotopy, let $\gamma : I \longrightarrow Y$ be the path given by $\gamma(t) = H(x_0, t)$, for some point $x_0 \in X$. Then $f_* = \varphi_\gamma \circ g_* : \pi_1(X, x_0) \longrightarrow \pi_1(Y, f(x_0))$, where φ_γ is as in* 4.1.21.

Proof: Take $[\lambda] \in \pi_1(X, x_0)$ and let $F : I \times I \longrightarrow Y$ be given by

$$F(s,t) = \begin{cases} H(\lambda(2(1-t)s), 2st) & \text{if } 0 \le s \le \frac{1}{2}, \\ H(\lambda(1 + 2t(s-1)), t + (1-t)(2s-1)) & \text{if } \frac{1}{2} \le s \le 1. \end{cases}$$

It is straightforward to check that F is a homotopy relative to $\{0, 1\}$ of the path product $(f \circ \lambda)\gamma$ to $\gamma(g \circ \lambda)$. Therefore, $[f \circ \lambda][\gamma] = [\gamma][g \circ \lambda]$, that is, $f_*([\lambda]) = \varphi_\gamma g_*([\lambda])$. □

By the theorem above, we have that the fundamental group is a *homotopy invariant*; i.e., it depends only on the homotopy type of the space. The following holds.

4.1.28 **Theorem**. *If $f : X \longrightarrow Y$ is a homotopy equivalence, then the induced homomorphism $f_* : \pi_1(X, x_0) \longrightarrow \pi_1(Y, f(x_0))$ is an isomorphism for every point $x_0 \in X$.*

Proof: Let $g : Y \longrightarrow X$ be a homotopy inverse of f; hence $g \circ f \simeq \mathrm{id}_X$ and $f \circ g \simeq \mathrm{id}_Y$. By 4.1.27, we have

$$\begin{aligned}(g \circ f)_* &= \varphi_\gamma : \pi_1(X, x_0) \longrightarrow \pi_1(X, gf(x_0))\,,\\ (f \circ g)_* &= \varphi_\mu : \pi_1(Y, f(x_0)) \longrightarrow \pi_1(Y, fgf(x_0))\,,\end{aligned}$$

for certain paths γ in X and μ in Y. That is, $g_* \circ f_*$ and $f_* \circ g_*$ are group isomorphisms with the inverse of the first being α, say. So, $g_* \circ (f_* \circ \alpha) = 1$ and $((f_* \circ \alpha) \circ g_*) \circ f_* = f_*$, but since f_* is an epimorphism, $(f_* \circ \alpha) \circ g_* = 1$; that is, g_* is an isomorphism. Therefore, since $(\alpha \circ g_*) \circ f_* = 1$ and $\alpha \circ g_*$ is an isomorphism, so is f_*. □

4.1.29 NOTE. Let $A \subset X$ and take $x_0 \in A$. Then, the inclusion $i : A \hookrightarrow X$ induces a homomorphism $i_* : \pi_1(A, x_0) \longrightarrow \pi_1(X, x_0)$, which, as shown by the case $A = \mathbb{S}^1 \subset \mathbb{B}^2 = X$, it is not in general a monomorphism. However, if λ is a loop in A representing an element in $\pi_1(A, x_0)$, then $i_*([\lambda])$ is represented by the loop $i \circ \lambda$, which is essentially the same loop λ, but now thought of as a loop in X. As is shown by the special case mentioned above, the fact that λ is a loop in A that is contractible in X does not mean that it is contractible in A; that is, if $i_*([\lambda]) = 0$, then it does not necessarily follow that $[\lambda] = 0$.

From 3.3.11, we obtain the following statement.

4.1.30 **Proposition**. *If $A \subset X$ is a defomation retract, then the inclusion $i : A \hookrightarrow X$ induces an isomorphism $i_* : \pi_1(A, x_0) \xrightarrow{\cong} \pi_1(X, x_0)$.* □

If $\lambda : I \longrightarrow X$ is a loop based at x_0, then λ determines a pointed map $\widetilde{\lambda} : (\mathbb{S}^1, 1) \longrightarrow (X, x_0)$ given by $\widetilde{\lambda}(\mathrm{e}^{2\pi \mathrm{i} t}) = \lambda(t)$. Conversely, a pointed map $f : (\mathbb{S}^1, 1) \longrightarrow (X, x_0)$ determines a loop λ_f based at x_0 given by $\lambda_f(t) = f(\mathrm{e}^{2\pi \mathrm{i} t})$. In other words, we have the next statement.

4.1.31 **Proposition**. *The function $\pi_1(X, x_0) \longrightarrow [\mathbb{S}^1, 1; X, x_0]$ given by $[\lambda] \mapsto [\widetilde{\lambda}]$ is bijective.* □

More generally, we have the following.

4.1.32 **Theorem**. *Let X be path connected, and let*

$$\Phi : \pi_1(X, x_0) \longrightarrow [\mathbb{S}^1, X]$$

be given by $\Phi([\lambda]) = [\widetilde{\lambda}]$ by ignoring the base points. Then Φ is surjective. Moreover, if $\alpha, \beta \in \pi_1(X, x_0)$, then $\Phi(\alpha) = \Phi(\beta)$ if and only if there exists $\gamma \in \pi_1(X, x_0)$ such that $\alpha = \gamma\beta\gamma^{-1}$; that is, α and β are conjugates.

Proof: Every map $f : \mathbb{S}^1 \longrightarrow X$ is homotopic to a map $g : \mathbb{S}^1 \longrightarrow X$ such that $g(1) = x_0$, since if $\sigma : f(1) \simeq x_0$ is some path, then the homotopy

$$H(s,t) = \begin{cases} \sigma(t - 3s) & \text{if } 0 \leq s \leq \frac{t}{3}, \\ f(\mathrm{e}^{2\pi \mathrm{i}(\frac{3s-t}{3-2t})}) & \text{if } \frac{t}{3} \leq s \leq \frac{3-t}{3}, \\ \sigma(3s + t - 3) & \text{if } \frac{3-t}{3} \leq s \leq 1, \end{cases}$$

is such that $H(s, 0) = f(\mathrm{e}^{2\pi \mathrm{i}s})$ and $H(s, 1)$ is the product loop $\overline{\sigma}\lambda_f\sigma$; in other words, the homotopy $K : \mathbb{S}^1 \times I \longrightarrow X$ given by $K(\mathrm{e}^{2\pi\mathrm{i}s}, t) = H(s, t)$ starts at f and ends at a map g such that $g(1) = \sigma(1) = x_0$. This shows that Φ is surjective.

Let us now assume that $\Phi([\lambda]) = \Phi([\mu])$; then we have a homotopy $L : \mathbb{S}^1 \times I \longrightarrow X$ such that $L(\mathrm{e}^{2\pi\mathrm{i}s}, 0) = \lambda(s)$ and $L(\mathrm{e}^{2\pi\mathrm{i}s}, 1) = \mu(s)$. Thus, the path $\sigma : I \longrightarrow X$ given by $\sigma(t) = L(1, t)$ is a loop representing an element $\gamma = [\sigma] \in \pi_1(X, x_0)$. Thanks to the homotopy

$$F(s,t) = \begin{cases} H(2(1-t)s, 2st) & \text{if } 0 \leq s \leq \frac{1}{2}, \\ H(1 + 2t(s-1), t + (1-t)(2s-1)) & \text{if } \frac{1}{2} \leq s \leq 1, \end{cases}$$

which is analogous to the one in the proof of 4.1.27, where $H(s,t) = L(\mathrm{e}^{2\pi\mathrm{i}s}, t)$, one has $\lambda\sigma \simeq \sigma\mu$.

Conversely, if $\lambda\sigma \simeq \sigma\mu$, then there exists a homotopy $H : \lambda \simeq \sigma\mu\overline{\sigma}$ rel ∂I. So $K(\mathrm{e}^{2\pi\mathrm{i}s}, t) = H(s, t)$ is a well-defined homotopy from $\widetilde{\lambda}$ to $\widetilde{\sigma\mu\overline{\sigma}}$. On the other hand, the homotopy

$$G(s,t) = \begin{cases} \sigma(3s + t) & \text{if } 0 \leq s \leq \frac{1-t}{3}, \\ \mu(\frac{3s+t-1}{1+2t}) & \text{if } \frac{1-t}{3} \leq s \leq \frac{2+t}{3}, \\ \sigma(3 - 3s + t) & \text{if } \frac{2+t}{3} \leq s \leq 1, \end{cases}$$

is such that $G : \sigma\mu\overline{\sigma} \simeq \mu$ and $G(0, t) = \sigma(t) = G(1, t)$; therefore, it defines a homotopy $M : \mathbb{S}^1 \times I \longrightarrow X$ such that $M(\mathrm{e}^{2\pi\mathrm{i}s}, t) = G(s, t)$, starting at $\widetilde{\sigma\mu\overline{\sigma}}$ and ending at $\widetilde{\mu}$. Thus the homotopies K and M may be composed to yield one from $\widetilde{\lambda}$ to $\widetilde{\mu}$; that is, $\Phi([\lambda]) = \Phi([\mu])$. $\square$

4.2 The fundamental group of the circle

The circle $\mathbb{S}^1$ is path connected, and thus its fundamental group is independent of the choice of base point. The natural base point is $1 \in \mathbb{S}^1$. In Section 3.2 in the preceding chapter we did all the necessary computations to understand this group. We shall use the results of that section, and as there, we keep close to the approach of [23]. The following lemma will be very useful.

4.2.1 **Lemma**. *The loop product of two loops in $\mathbb{S}^1$ is homotopic to the product of the loops realized as maps with complex values.*

Proof: Let λ, $\mu : I \longrightarrow \mathbb{S}^1$ be two loops. Take the homotopy

$$H(s,t) = \begin{cases} \lambda(2s) & \text{if } 0 \leq s \leq \frac{1-t}{2}, \\ \lambda(\frac{2s-t+1}{2}) \cdot \mu(\frac{2s+t-1}{2}) & \text{if } \frac{1-t}{2} \leq s \leq \frac{1+t}{2}, \\ \mu(2s-1) & \text{if } \frac{1+t}{2} \leq s \leq 1, \end{cases}$$

where $\zeta \cdot \eta$ represents the product in $\mathbb{S}^1$ of the unit complex numbers ζ and η. This homotopy starts with the loop product $\lambda\mu$ and ends with the complex product of complex maps $\lambda \cdot \mu$. □

By the previous lemma, we have that if $[\lambda], [\mu] \in \pi_1(\mathbb{S}^1, 1)$, then $[\lambda][\mu] = [\lambda \cdot \mu]$, and therefore, since the complex product is commutative, we have that $[\lambda][\mu] = [\mu][\lambda]$; that is, we have the following consequence of the previous lemma.

4.2.2 **Lemma**. *The fundamental group of the circle $\pi_1(\mathbb{S}^1, 1)$ is abelian.* □

4.2.3 NOTE. One can give a direct proof of the fact that the fundamental group of the circle is abelian. To start, let $\lambda, \mu : I \longrightarrow \mathbb{S}^1$ be loops. The homotopy $H : I \times I \longrightarrow \mathbb{S}^1$ given by

$$H(s,t) = \begin{cases} \mu(2st) \cdot \lambda(2(1-t)s) & \text{if } 0 \leq s \leq \frac{1}{2}, \\ \mu(t + (1-t)(2s-1)) \cdot \lambda(1 + 2t(s-1)) & \text{if } \frac{1}{2} \leq s \leq 1, \end{cases}$$

where $\zeta \cdot \eta$ is the product of the complex numbers ζ and η in $\mathbb{S}^1$, is such that $H : \lambda\mu \simeq \mu\lambda$; that is, $[\lambda][\mu] = [\mu][\lambda]$.

The homotopy above is indeed the composite of two maps, namely of the map $f : I \times I \longrightarrow I \times I$ given by

$$f(s,t) = \begin{cases} (2(1-t)s, 2st) & \text{if } 0 \leq s \leq \frac{1}{2}, \\ (1 + 2t(s-1), t + (1-t)(2s-1)) & \text{if } \frac{1}{2} \leq s \leq 1, \end{cases}$$

and the map $g : I \times I \longrightarrow \mathbb{S}^1$ given by $g(s,t) = \mu(t) \cdot \lambda(s)$. The map f takes the sides $\{0\} \times I$ and $\{1\} \times I$ of the square onto the vertices $(0,0)$ and $(1,1)$, respectively, and the sides $I \times \{0\}$ and $I \times \{1\}$ to $I \times \{0\} \cup \{1\} \times I$ and $\{0\} \times I \cup I \times \{1\}$, respectively. On the other hand, the map g "translates" the loop λ in $\mathbb{S}^1$ along the loop μ. What this looks like is shown in Figure 4.4.

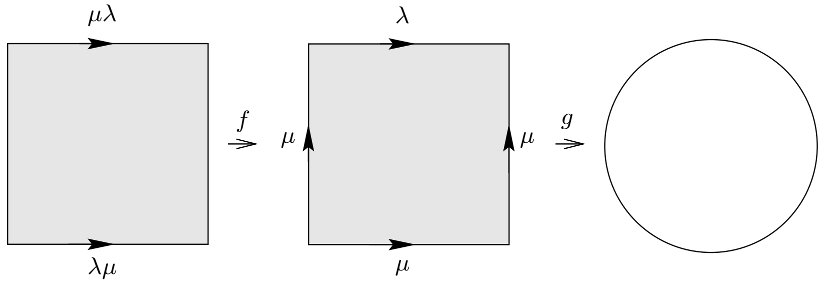


Figure 4.4 The fundamental group of the circle is abelian

4.2.4 Exercise. A group G is called a *topological group* if it is also a topological space and both the multiplication $G \times G \longrightarrow G$ and the function $G \longrightarrow G$ that sends each element to its inverse, are continuous maps. Prove that the fundamental group of every (path-connected) topological group G based at 1, that is, $\pi_1(G, 1)$, is abelian. (*Hint:* One may use the same proof as given for 4.2.1.)

4.2.5 Exercise. Let G be a topological group (or an H-space; see [2, 2.7.2]). Prove that if λ, $\mu : I \longrightarrow G$ are loops, then $[\lambda][\mu] = [\lambda \cdot \mu]$, where $\cdot$ represents the group multiplication. Use this to show that $\pi_1(G, 1)$ is abelian. (*Hint:* Use [2, 2.10.10].)

Let us recall the function $\deg : [\mathbb{S}^1, \mathbb{S}^1] \longrightarrow \mathbb{Z}$ defined in 3.2.5, and the function $\Phi : \pi_1(\mathbb{S}^1, 1) \longrightarrow [\mathbb{S}^1, \mathbb{S}^1]$ of the previous section. Let $\Psi = \deg \circ \Phi : \pi_1(\mathbb{S}^1, 1) \longrightarrow \mathbb{Z}$. We summarize what we did in Section 3.2 in the following result.

4.2.6 **Theorem.** *$\Psi : \pi_1(\mathbb{S}^1, 1) \longrightarrow \mathbb{Z}$ is a group isomorphism.*

Proof: By 3.2.7 and by 4.1.27, since in this case φ_γ is the identity, Ψ is bijective. Thus it is enough to check that it is a group homomorphism. Take $\alpha = [\lambda], \beta = [\mu] \in \pi_1(\mathbb{S}^1, 1)$; by 4.2.1, $\alpha\beta = [\lambda \cdot \mu]$. If $\widetilde{\lambda}, \widetilde{\mu} : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ are representatives of $\Phi(\alpha), \Phi(\mu)$, respectively, then $\Psi(\alpha\beta) = \Psi([\lambda \cdot \mu]) = \deg(\widetilde{\lambda \cdot \mu}) = \deg(\widetilde{\lambda} \cdot \widetilde{\mu}) = \deg(\widetilde{\lambda}) + \deg(\widetilde{\mu}) = \Psi(\alpha) + \Psi(\beta)$, where the next to the last equality comes from 3.2.9. $\square$

Let $\gamma_n : I \longrightarrow \mathbb{S}^1$ be given by $\gamma_n(t) = \mathrm{e}^{2\pi i n t} = g_n(\mathrm{e}^{2\pi \mathrm{i} t})$. Then $\Phi([\gamma_n]) = [g_n]$, and thus $\Psi([\gamma_n]) = \deg(g_n) = n$. Hence in particular, $\Psi([\gamma_1]) = 1$ is a generator of $\mathbb{Z}$ as an infinite cyclic group. We have thus the following result.

4.2.7 **Theorem.** *$\pi_1(\mathbb{S}^1, 1)$ is an infinite cyclic group generated by $[\gamma_1]$, that is, by the homotopy class of the loop $t \mapsto \mathrm{e}^{2\pi \mathrm{i} t}$.* $\square$

4.2.8 Definition. The class $[\gamma_1]$ is called the *canonical generator* of the infinite cyclic group $\pi_1(\mathbb{S}^1, 1)$.

If one works with a path-connected space, then as we already proved in 4.1.21, its fundamental group is essentially independent of the base point. In what follows, whenever the base point either is clear or irrelevant, we shall denote the fundamental group of a path-connected space X simply by $\pi_1(X)$.

4.2.9 Examples. If a space X has the same homotopy type of $\mathbb{S}^1$, then $\pi_1(X) \cong \mathbb{Z}$; we have the following:

(a) $\pi_1(\mathbb{C} - 0) \cong \mathbb{Z}$. The isomorphism is defined by $[\lambda] \mapsto W(f_\lambda, 0)$, the winding number around 0 of the map $f_\lambda : \mathbb{S}^1 \longrightarrow \mathbb{C}$ given by $f_\lambda(\mathrm{e}^{2\pi \mathrm{i} t}) = \lambda(t)$.

(b) If Y is contractible and $X = Y \times \mathbb{S}^1$, then, by 4.1.18 and 4.1.17(d), $\pi_1(X) \cong \pi_1(Y) \times \pi_1(\mathbb{S}^1) \cong \pi_1(\mathbb{S}^1) \cong \mathbb{Z}$. In particular, if $X = \mathbb{B}^2 \times \mathbb{S}^1$ is a *solid torus*, $\pi_1(X) \cong \mathbb{Z}$.

(c) If M is the Moebius band, then $\pi_1(M) \cong \mathbb{Z}$. In fact, the *equatorial loop* $\lambda_e : I \longrightarrow M$ such that $\lambda_e(t) = q(t, \frac{1}{2})$, where $q : I \times I \longrightarrow M$ is the canonical identification, represents a generator of $\pi_1(M)$.

The following example, in particular, is very important. It is an immediate consequence of 4.1.18 and 4.2.7.

4.2.10 Example. If $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ is the torus and $x_0 = (1, 1) \in \mathbb{T}^2$, then

$$\pi_1(\mathbb{T}^2, x_0) \cong \mathbb{Z} \oplus \mathbb{Z}\,. \tag{4.2.11}$$

Moreover, if $\gamma_1^1, \gamma_1^2 : I \longrightarrow \mathbb{T}^2$ are the *canonical loops* $\gamma_1^1(t) = (\gamma_1(t), 1)$, $\gamma_1^2(t) = (1, \gamma_1(t))$, then we may reformulate (4.2.11) by saying that $\pi_1(\mathbb{T}^2, x_0)$ is the free abelian group generated by the classes $\alpha_1 = [\gamma_1^1]$ and $\alpha_2 = [\gamma_1^2]$.

As a generalization of the previous example, we may prove immediately by induction the following.

4.2.12 **Proposition**. *Let*

$$\mathbb{T}^n = \underbrace{\mathbb{S}^1 \times \cdots \times \mathbb{S}^1}_{n}\,.$$

Then $\pi_1(\mathbb{T}^n)$ is the free abelian group generated by the classes $[\gamma_1^1], \ldots, [\gamma_1^n]$ defined by

$$\gamma_1^i(t) = (1, \ldots, \underbrace{\gamma_1(t)}_{i}, \ldots, 1) \in \mathbb{T}^n\,.$$

□

Let $g_n : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ be the map of degree n given by $g_n(\zeta) = \zeta^n$. For the canonical loop $\gamma_1 : I \longrightarrow \mathbb{S}^1$, such that $[\gamma_1]$ is the canonical generator of $\pi_1(\mathbb{S}^1)$, one has that $g_n \circ \gamma_1 = \gamma_n$, so that $(g_n)_*([\gamma_1]) = [\gamma_n] = [\gamma_1]^n$ (since by the considerations prior to 4.2.7, $\Psi(\gamma_n) = n$). Hence $g_{n*} : \pi_1(\mathbb{S}^1) \longrightarrow \pi_1(\mathbb{S}^1)$ is $g_{n*}(\alpha) = \alpha^n$. Since $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ has degree n implies $f \simeq g_n$, we therefore have the following theorem.

4.2.13 **Theorem**. *Let $f : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ satisfy $\deg(f) = n$. Then the homomorphism $f_* : \pi_1(\mathbb{S}^1) \longrightarrow \pi_1(\mathbb{S}^1)$ is given by $f_*(\alpha) = \alpha^n$.* □

4.2.14 Note. Strictly speaking, in the previous theorem one has the homomorphism $f_* : \pi_1(\mathbb{S}^1, 1) \longrightarrow \pi_1(\mathbb{S}^1, f(1))$; thus the statement of the theorem can be more precisely applied to the composite

$$\pi_1(\mathbb{S}^1, 1) \xrightarrow{f_*} \pi_1(\mathbb{S}^1, f(1)) \xrightarrow{\left(r_{f(1)^{-1}}\right)_*} \pi_1(\mathbb{S}^1, 1)\,,$$

where $r_{f(1)^{-1}} : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ is the rotation in $\mathbb{S}^1$ given by multiplying by $f(1)^{-1}$, which is homotopic to the identity.

Another interesting and useful example is the following.

4.2.15 Example. Let $f_{bd}^{ac} : \mathbb{S}^1 \times \mathbb{S}^1 \longrightarrow \mathbb{S}^1 \times \mathbb{S}^1$ be given by $f_{bd}^{ac}(\zeta, \eta) = (\zeta^a \cdot \eta^b, \zeta^c \cdot \eta^d)$, $a, b, c, d \in \mathbb{Z}$. Then, by 4.2.13 and 4.2.10, $(f_{bd}^{ac})_* : \pi_1(\mathbb{T}^2) \longrightarrow \pi_1(\mathbb{T}^2)$ is such that $(f_{bd}^{ac})_*(\alpha_1) = \alpha_1^a \alpha_2^c$ and $(f_{bd}^{ac})_*(\alpha_2) = \alpha_1^b \alpha_2^d$, if $\alpha_1, \alpha_2 \in \pi_1(\mathbb{T}^2)$ are as in 4.2.10.

4.2.16 Exercise. Check all details of the assertions in the example above and characterize the values of a, b, c, d for which $(f_{bd}^{ac})_*$ is an isomorphism. What can be said about the map f_{bd}^{ac} for these values?

4.2.17 Exercise. Let $\varphi : \pi_1(\mathbb{T}^2) \longrightarrow \pi_1(\mathbb{T}^2)$ be any homomorphism. Prove that there exists $f : \mathbb{T}^2 \longrightarrow \mathbb{T}^2$ such that $f_* = \varphi$. Moreover, show that if φ is an isomorphism, then f can be chosen to be a homeomorphism. (*Hint:* Use Example 4.2.15.)

4.2.18 Exercise. Prove that $\mathbb{T}^m \approx \mathbb{T}^n$ if and only if $m = n$.

4.2.19 Exercise. Prove that a loop $\lambda : I \longrightarrow \mathbb{S}^1$ is such that $[\lambda] \in \pi_1(\mathbb{S}^1)$ is a generator if and only if $W(f_\lambda, 0) = \pm 1$, where $f_\lambda : \mathbb{S}^1 \longrightarrow \mathbb{C}$ is given by $f_\lambda(\mathrm{e}^{2\pi i t}) = \lambda(t)$ and W is the winding number function.

4.2.20 Exercise. If M is the Moebius band and $f : \mathbb{S}^1 \longrightarrow \partial M$ is a homeomorphism, prove that the loop $\lambda_f : I \longrightarrow M$ given by $\lambda_f(t) = f(\mathrm{e}^{2\pi i t}) \in M$ satisfies $[\lambda_f] = \alpha^2$ for α one of the generators of $\pi_1(M) \cong \mathbb{Z}$ (see 4.2.9(c)). Conclude that the boundary ∂M is not a retract of M.

4.3 The Seifert–van Kampen theorem

A very useful tool is a formula that in some cases allows us to compute the fundamental group of certain spaces in terms of the fundamental groups of parts of them. Before going to the general formula, as an example of it, let us first analyze a special case.

4.3.1 **Proposition**. *Let $X = X_1 \cup X_2$ with X_1, X_2 open subsets. If X_1 and X_2 are simply connected and $X_1 \cap X_2$ is path connected, then X is simply connected.*

Proof: Let $\lambda : I \longrightarrow X$ be a loop based at $x_0 \in X_1 \cap X_2$. We have that the set $\{\lambda^{-1}(X_1), \lambda^{-1}(X_2)\}$ is an open cover of I. There exists a number $\delta > 0$, called the Lebesgue number of this cover (see [21]), such that if $0 \leq t - s < \delta$, then $[s,t] \subset \lambda^{-1}(X_1)$ or $[s,t] \subset \lambda^{-1}(X_2)$. Hence, one has a partition $0 = t_0 < t_1 < \cdots < t_k = 1$ of the interval I such that

$$\lambda([t_0, t_1]) \subset X_1, \quad \lambda([t_1, t_2]) \subset X_2, \ \ldots, \lambda([t_{k-1}, t_k]) \subset X_2 \,.$$

Since $\lambda(t_i) \in X_1 \cap X_2$, there exist paths $\omega_i : x_0 \simeq \lambda(t_i)$ in $X_1 \cap X_2$, $i = 1, 2, \ldots, k-1$; let moreover ω_0 as well as ω_k denote the constant path at $x_0 = \lambda(t_0) = \lambda(0) = \lambda(1) = \lambda(t_k)$. The loops

$$\mu_i(t) = \begin{cases} \omega_{i-1}(3t) & \text{if } 0 \leq t \leq \frac{1}{3}, \\ \lambda_i(3t-1) & \text{if } \frac{1}{3} \leq t \leq \frac{2}{3}, \\ \omega_i(3-3t) & \text{if } \frac{2}{3} \leq t \leq 1, \end{cases}$$

where $\lambda_i(t) = \lambda((1-t)t_{i-1} + tt_i)$ is the portion of λ in the interval $[t_{i-1}, t_i]$, $i = 1, 2, \ldots, k$, lie in X_1 or in X_2, and therefore they are contractible in X_1 or in X_2 and hence in X; that is, $\mu_i \simeq 0$ in X. Since $\lambda \simeq \mu_1 \mu_2 \cdots \mu_k$, we have that λ is contractible, that is, $\lambda \simeq 0$. Figure 4.5 shows the proof graphically. □

An important application is given in the next example.

4.3.2 Example. If $n > 1$, then the sphere $\mathbb{S}^n$ is simply connected, that is, $\pi_1(\mathbb{S}^n) = 1$. For if $N = (0, \ldots, 0, 1)$ and $S = (0, \ldots, 0, -1)$ are the poles of the sphere and $X_1 = \mathbb{S}^n - S$, $X_2 = \mathbb{S}^n - N$, then the hypotheses of 4.3.1 hold, since X_1 and X_2, being homeomorphic to $\mathbb{R}^n$, are contractible, and $X_1 \cap X_2$ is path connected, since $X_1 \cap X_2 \approx \mathbb{S}^{n-1} \times (-1, 1) \simeq \mathbb{S}^{n-1}$.

4.3.3 Note. For $n = 1$, the preceding example obviously does not hold, because $(\mathbb{S}^1 - N) \cap (\mathbb{S}^1 - S)$ is not path connected. Indeed, in this case, $\pi_1(\mathbb{S}^1) \cong \mathbb{Z}$, as shown in 4.2.7.

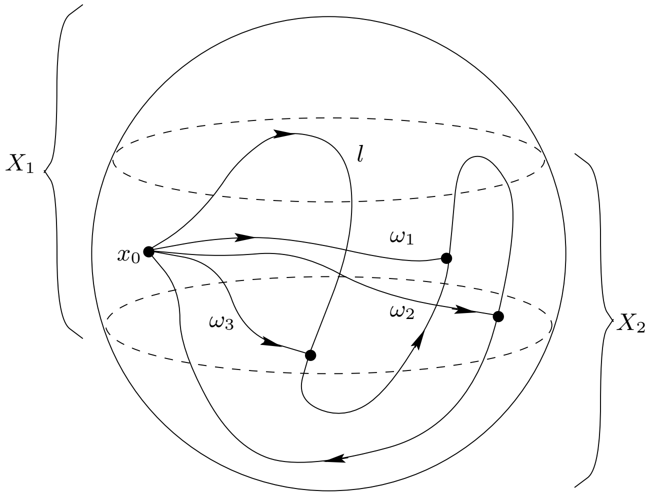


Figure 4.5 The Seifert–van Kampen theorem in the 2-sphere

4.3.4 EXERCISE. Recall the suspension of a space X defined as $\Sigma X = X \times I/\!\sim$, where $(x,s) \sim (y,t)$ if and only if $x = y$ and $s = t$ or $s = t = 0$ or 1. Prove that *if X is path connected, then its (reduced) suspension ΣX is simply connected.*

The Seifert–van Kampen theorem is a generalization of 4.3.1, because it allows one to compute the fundamental group of a union of open subspaces if one knows the fundamental groups of each of them and the way that the fundamental group of the intersection relates to these.

Since we are going to use more delicate concepts of group theory, at this point we shall make an algebraic parenthesis to make ideas more precise.

4.3.5 DEFINITION. Let G be a group and let $A \subset G$ be any subset. The subgroup

$$G_A = \cap\{H \subset G \mid H \text{ is a subgroup such that } A \subset H\}$$

is called the *subgroup of G generated by A* and the normal subgroup

$$N_A = \cap\{H \subset G \mid H \text{ is a normal subgroup such that } A \subset H\}$$

is known as the *normal subgroup of G generated by A*. G_A consists of 1 and the elements of the form

$$g = a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k}\,,$$

where $a_1, a_2, \ldots, a_k \in A$ and $n_1, n_2, \ldots, n_k \in \mathbb{Z}$; similarly, N_A consists of products of all conjugates of the elements of G_A, namely conjugates of elements of the form described above. If $G = G_A$ we say that A *generates* G or that the elements of A are *generators* of G. Analogously, if $N_A = G$ we say that A *generates G normally* or that the elements of A are *normal generators* of G.

4.3.6 Note. For abelian groups with the additive notation, Definition 4.3.5 can be expressed by saying that $A \subset G$ generates G if for $0 \neq g \in G$ there exist $a_1, \ldots, a_k \in A$, $n_1, \ldots, n_k \in \mathbb{Z}$ such that $g = n_1 a_1 + \cdots + n_k a_k$ (in this case all elements $a_1, \ldots, a_k$ can be assumed to be different).

4.3.7 Examples.

(a) The empty set $\emptyset$ generates the trivial group $G = 1$.

(b) If G is generated by only one element a, then G is *cyclic*; if $a^n \neq 1$ for every $n \in \mathbb{Z}$, $n \neq 0$, then G is *infinite cyclic* and consists of all elements $a^0 = 1, a^{\pm 1}, a^{\pm 2}, \ldots$; in this case the function $n \mapsto a^n$ determines an isomorphism $\mathbb{Z} \longrightarrow G$. If $a^k = 1$ for some $k > 0$ and this k is minimal, then G is *cyclic of order* k and consists of all elements $a^0 = 1, a, a^2, \ldots, a^{k-1}$. The function $n \mapsto a^n$ determines an isomorphism $\mathbb{Z}_k \longrightarrow G$.

(c) Given groups G_1 and G_2, one has the *product group* $G_1 \times G_2$, that as a set is the cartesian product and is provided with the multiplication coordinate by coordinate There are group inclusions $j_1 : G_1 \hookrightarrow G_1 \times G_2$ given by $g_1 \mapsto (g_1, 1)$, and $j_2 : G_2 \hookrightarrow G_1 \times G_2$ given by $g_2 \mapsto (1, g_2)$, so that we may view both groups as subgroups of the product. Since $(g_1, g_2) = (g_1, 1) \cdot (1, g_2) = g_1 \cdot g_2$, the group $G_1 \times G_2$ is generated by the elements of $G_1 \cup G_2 \subset G_1 \times G_2$. Moreover, one has that $g_1 \cdot g_2 = g_2 \cdot g_1$.

Analogous to (c) we have the following. Let G_1 and G_2 be groups. We define a group $G_1 * G_2$ that contains G_1 and G_2 as subgroups and that is generated by the union $G_1 \cup G_2$ but does not satisfy the relation $g_1 \cdot g_2 = g_2 \cdot g_1$, as it is the case in the product.

4.3.8 Definition. Let G_ν $(\nu = 1, 2)$ be groups and let F be the set of finite sequences $(x_1, \ldots, x_n)$, $n \geq 0$, that satisfy

(a) $x_j \in G_\nu$, $j = 1, \ldots n$;

(b) $x_j \neq 1$, $j = 1, \ldots n$;

(c) $x_j \in G_\nu \implies x_{j+1} \notin G_\nu$; that is, two consecutive elements lie in different groups.

In particular, for $n = 0$ we write $(\ )$ for the empty sequence. Take $g \in G_\nu$ and let $\overline{g} : F \longrightarrow F$ be the function given by

$$\overline{g}(x_1, x_2, \ldots, x_n) = \begin{cases} (x_1, x_2, \ldots, x_n) & \text{if } g = 1, \\ (g, x_1, x_2, \ldots, x_n) & \text{if } g \neq 1 \text{ and } x_1 \notin G_\nu, \\ (gx_1, x_2, \ldots, x_n) & \text{if } g \neq 1,\ x_1 \in G_\nu \text{ and } gx_1 \neq 1, \\ (x_2, \ldots, x_n) & \text{if } g \neq 1,\ x_1 \in G_\nu \text{ and } gx_1 = 1. \end{cases}$$

In particular, $\overline{g}(\ ) = (g)$ if $g \neq 1$, and $\overline{g}(x_1) = (\ )$ if $gx_1 = 1$. If $g, h \in G_\nu$, $g, h \neq 1$ are such that $\overline{g} = \overline{h}$, then, in particular, $(g) = \overline{g}(\ ) = \overline{h}(\ ) = (h)$, so that $g = h$; if $1 \in G_\nu$, then $\overline{1} = 1_F : F \longrightarrow F$ is the identity and if $g, h \in G_\nu$, then $\overline{gh} = \overline{g} \circ \overline{h}$, as one may easily prove. Hence we have that $g \mapsto \overline{g}$ determines an inclusion of groups $G_\nu \hookrightarrow \mathcal{P}(F) = \{f : F \longrightarrow F \mid f \text{ is bijective}\}$; that is, $\mathcal{P}(F)$ is the *permutation group* of F (this is thus a *representation* of the groups G_1 and G_2). The *free product* $G_1 * G_2$ is the subgroup of $\mathcal{P}(F)$ generated by $G_1 \cup G_2$. There are canonical inclusions (group monomorphisms) $i_1 : G_1 \hookrightarrow G_1 * G_2$ and $i_2 : G_2 \hookrightarrow G_1 * G_2$.

4.3.9 **Proposition**. *For every element $g \in G_1 * G_2$, $g \neq 1$, there is a unique representation $g = x_1 \cdots x_n$, so that the sequence $(x_1, \ldots, x_n)$ lies in F, that is, such that it satisfies conditions* (a), (b), *and* (c) *of* 4.3.8.

Proof: g is a permutation of the elements of F different from the identity, that, lying in the subgroup of $\mathcal{P}(F)$ generated by $G_1 \cup G_2$, is a product $x_1 \cdots x_n$, where each x_i is a permutation of F coming from G_1 or G_2. Reducing (that is, eliminating elements 1 or putting together consecutive elements that lie in the same group), if it is necessary, we may assume that the sequence $(x_1, \ldots, x_n)$ satisfies conditions (a), (b), and (c) of 4.3.8.

It is therefore enough to see that this representation is unique. If $g = x_1 \cdots x_n$ is a representation of $g \neq 1$, then one may inductively verify that $g(\ ) = (x_1, \ldots, x_n)$. If, moreover, $g = y_1 \cdots y_m$, so that $(y_1, \ldots, y_m) \in F$, then also $g(\ ) = (y_1, \ldots, y_m)$; consequently, $(x_1, \ldots, x_n) = (y_1, \ldots, y_m)$ and the representation is unique. □

4.3.10 NOTE. Given elements with unique representation $g = x_1 \cdots x_m$, $h = y_1 \cdots y_n \in G_1 * G_2$, we have that $g^{-1} = x_m^{-1} \cdots x_1^{-1}$ is its unique representation, and $gh = x_1 \cdots x_m y_1 \cdots y_n$ is a representation that can still be *reduced*, that is, if $x_m, y_1 \in G_\nu$, then the product $x_m y_1$ can be taken as one element; if this happens to be 1, then it is removed and one takes the product $x_{m-1} y_2$ as only one element, and so forth, if necessary.

4.3.11 REMARK. The following are properties of the free product:

(a) If $G_1 \neq 1 \neq G_2$, then $G_1 * G_2$ is not an abelian group, since by the uniqueness of the representation of an element g proved in 4.3.9, if $g = x_1 x_2$ is such that $1 \neq x_1 \in G_1$ and $1 \neq x_2 \in G_2$, then g is different from the element $h = x_2 x_1$, so that x_1 and x_2 do not commute.

(b) If $G_1 = 1$, then $G_1 * G_2 = G_2$, since the only possible representation of an element $g \in G_1 * G_2$ is $g = x_1$, $x_1 \in G_2$, so that the mapping $g \mapsto x_1$ determines the desired equality.

(c) G_1 and G_2, as subgroups of $G_1 * G_2$, are such that $G_1 \cap G_2 = 1$; that is, their intersection is the trivial subgroup.

(d) There is a natural epimorphism

$$\gamma : G_1 * G_2 \longrightarrow G_1 \times G_2$$

such that the kernel $\ker(\gamma)$ contains the *commutators* $x_1x_2x_1^{-1}x_2^{-1}$ such that $x_1 \in G_1$, $x_2 \in G_2$. Namely, let $x_1 \cdots x_n$ be the unique representation of an element $g \in G_1 * G_2$. If $x_1 \in G_1$, then

$$\gamma(g) = (x_1x_3 \cdots, x_2x_4 \cdots);$$

similarly, if $x_1 \in G_2$, then

$$\gamma(g) = (x_2x_4 \cdots, x_1x_3 \cdots).$$

γ is thus a well-defined epimorphism. If $g \in G_1 * G_2$ is a commutator, that is, if $g = x_1x_2x_1^{-1}x_2^{-1}$, with $x_1 \in G_{\nu_1}$ and $x_2 \in G_{\nu_2}$, $\nu_1 \neq \nu_2$, then $\gamma(g) = (x_1x_1^{-1}, x_2x_2^{-1}) = (1,1) = 1$ or $\gamma(g) = (x_2x_2^{-1}, x_1x_1^{-1}) = (1,1) = 1$ according to whether $\nu_1 = 1$ or $\nu_1 = 2$.

The free group has a *universal property.*

4.3.12 **Theorem**. *Let $f_1 : G_1 \longrightarrow H$ and $f_2 : G_2 \longrightarrow H$ be group homomorphisms. Then there exists a unique homomorphism $f : G_1 * G_2 \longrightarrow H$ such that $f \circ i_1 = f_1$ and $f \circ i_2 = f_2$; that is, such that the diagram*

(4.3.13)
$$\begin{tikzcd} G_1 \arrow[r, hook, "i_1"] \arrow[dr, "f_1"'] & G_1 * G_2 \arrow[d, dashed, "f"] & G_2 \arrow[l, hook', "i_2"'] \arrow[dl, "f_2"] \\ & H . & \end{tikzcd}$$

commutes. Moreover, if $g = x_1x_2x_3 \cdots x_n$ is the unique representation, then f is such that

(4.3.14)
$$f(g) = f_{\nu_1}(x_1)f_{\nu_2}(x_2)f_{\nu_3}(x_3) \cdots f_{\nu_n}(x_n),$$

where $f_{\nu_i} = f_\nu$, si $x_i \in G_\nu$, $i = 1, \ldots, n$, $\nu = 1, 2$.

Proof: For $g = 1$, $f(g) = 1$; if $g = x_1 \cdots x_n \in G_1 * G_2$, then define f by (4.3.14); f is well defined and makes Diagram (4.3.13) commutative. □

4.3.15 NOTE. The homomorphism $\gamma : G_1 * G_2 \longrightarrow G_1 \times G_2$ of 4.3.11(d), is the one corresponding to $f_1 = j_1 : G_1 \hookrightarrow G_1 \times G_2$ and $f_2 = j_2 : G_2 \hookrightarrow G_1 \times G_2$, according to the preceding theorem.

Let $f_1 : G_1 \longrightarrow H_1$ and $f_2 : G_2 \longrightarrow H_2$ be group homomorphisms. If $j_1 : H_1 \hookrightarrow H_1 * H_2$ and $j_2 : H_2 \hookrightarrow H_1 * H_2 = H$ are the canonical inclusions, then there are homomorphisms $j_1 \circ f_1 : G_1 \longrightarrow H$ and $j_2 \circ f_2 : G_2 \longrightarrow H$, that by the universal property of the free product, define a unique homomorphism $f : G_1 * G_2 \longrightarrow H$ such that $f \circ i_1 = j_1 \circ f_1$ and $f \circ i_2 = j_2 \circ f_2$. This homomorphism f is such that if $x_1x_2x_3 \cdots x_n \in G_1 * G_2$, then $f(x_1x_2x_3 \cdots x_n) = f_{\nu_1}(x_1)f_{\nu_2}(x_2)f_{\nu_3}(x_3) \cdots f_{\nu_n}(x_n) \in H_1 * H_2$, where $f_{\nu_i} = f_\nu : G_\nu \longrightarrow H_\nu$ if $x_i \in G_\nu$, $i = 1, \ldots, n$, $\nu = 1, 2$. Such an f is denoted by

$$f_1 * f_2 : G_1 * G_2 \longrightarrow H_1 * H_2 \,.$$

In fact, we have that the construction of the free product is functorial, as we shall see now.

4.3.16 **Theorem**. *If $f_1 : G_1 \longrightarrow H_1$ and $f_2 : G_2 \longrightarrow H_2$, as well as $f_1' : H_1 \longrightarrow K_1$ and $f_2' : H_2 \longrightarrow K_2$ are group homomorphisms, then*

(a) $1_{G_1} * 1_{G_2} = 1_{G_1 * G_2} : G_1 * G_2 \longrightarrow G_1 * G_2$,

(b) $(f_1' \circ f_1) * (f_2' \circ f_2) = (f_1' * f_2') \circ (f_1 * f_2) : G_1 * G_2 \longrightarrow K_1 * K_2$. □

4.3.17 Remark. Definition 4.3.8 of the free product is, indeed, independent of the set of indices where ν runs. Equally well one may define the free product $G_1 * \cdots * G_n$ of any finite family of groups, or even more generally, the free product $\ast_\nu G_\nu$ of any family of groups $\{G_\nu\}$. In any case, the set of generators is $\cup_\nu G_\nu \subset \mathcal{P}(F)$.

Let us go back again to topology and take a topological space $X = X_1 \cup X_2$, $X_1 \cap X_2 \neq \emptyset$. Take $x_0 \in X_1 \cap X_2$; by the functoriality of the fundamental group, the commutative diagram of inclusions of topological spaces

$$\begin{tikzcd} X_1 \cap X_2 \arrow[r, hook, "i_1"] \arrow[d, hook, "i_2"'] & X_1 \arrow[d, hook, "j_1"] \\ X_2 \arrow[r, hook, "j_2"'] & X \end{tikzcd}$$

induces a commutative diagram of group homomorphisms

$$\begin{tikzcd} \pi_1(X_1 \cap X_2, x_0) \arrow[r, "i_{1*}"] \arrow[d, "i_{2*}"'] & \pi_1(X_1, x_0) \arrow[d, "j_{1*}"] \\ \pi_1(X_2, x_0) \arrow[r, "j_{2*}"'] & \pi_1(X, x_0)\,. \end{tikzcd}$$

We have the next assertion.

4.3.18 **Lemma**. *If X_1 and X_2 are open sets in X and are such that they as well as $X_1 \cap X_2$ are 0-connected, then $\pi_1(X, x_0)$ is generated by the images of $\pi_1(X_1, x_0)$ and $\pi_1(X_2, x_0)$ under j_{1*} and j_{2*}, respectively. Therefore, the homomorphism*

$$\varphi : \pi_1(X_1, x_0) * \pi_1(X_2, x_0) \longrightarrow \pi_1(X, x_0)$$

induced by j_{1*} and j_{2*} according to 4.3.12, is an epimorphism.

Proof: The proof of this result is essentially the same as the one given for 4.3.1. Namely, let $[\lambda] \in \pi_1(X, x_0)$ be an arbitrary element and take a partition $0 = t_0 < t_1 < \cdots < t_k = 1$ such that

$$\lambda([t_0, t_1]) \subset X_1, \quad \lambda([t_1, t_2]) \subset X_2, \ \ldots \ , \lambda([t_{k-1}, t_k]) \subset X_\nu \,, \quad \nu = 1 \text{ or } 2\,.$$

Thus $\lambda(t_i) \in X_1 \cap X_2$, $i = 0, \ldots, k$. For each i, take $\lambda_i(t) = \lambda((1-t)t_{i-1} + tt_i)$ and for $i = 1, \ldots, k-1$ let $\omega_i : I \longrightarrow X$ be a path between x_0 and $\lambda(t_i)$ in $X_1 \cap X_2$; moreover, take $\omega_0 = c_{x_0} = \omega_k$. This way we have loops

$$\mu_i(t) = \begin{cases} \omega_{i-1}(3t) & \text{if } 0 \le t \le \frac{1}{3}, \\ \lambda_i((3t-1) & \text{if } \frac{1}{3} \le t \le \frac{2}{3}, \\ \omega_i(3-3t) & \text{if } \frac{2}{3} \le t \le 1, \end{cases}$$

that lie either in X_1 or in X_2, and that therefore represent elements of $\pi_1(X, x_0)$ either in the image of $\pi_1(X_1, x_0)$ or of $\pi_1(X_2, x_0)$. Hence, since

$$[\mu_1][\mu_2] \cdots [\mu_{k-1}][\mu_k] = [\lambda_1][\lambda_2] \cdots [\lambda_{k-1}][\lambda_k] = [\lambda]\,,$$

this arbitrary element lies in the group generated by the images of j_{1*} and j_{2*}. $\square$

If we call the epimorphism φ above $j_{1*} \cdot j_{2*}$, from this last result we can deduce that

$$\pi_1(X, x_0) \cong \pi_1(X_1, x_0) * \pi_1(X_2, x_0) / \ker(j_{1*} \cdot j_{2*})\,.$$

In what follows we shall compute $\ker(j_{1*} \cdot j_{2*})$.

Take $\alpha \in \pi_1(X_1 \cap X_2, x_0)$. Then $j_{1*} \cdot j_{2*}(i_{1*}(\alpha)) = j_{1*}i_{1*}(\alpha) = j_{2*}i_{2*}(\alpha) = j_{1*} \cdot j_{2*}(i_{2*}(\alpha))$, and so $i_{1*}(\alpha)i_{2*}(\alpha)^{-1} \in \ker(j_{1*} \cdot j_{2*})$. Consequently, $\ker(j_{1*} \cdot j_{2*})$ contains the normal subgroup of $\pi_1(X_1, x_0) * \pi_1(X_2, x_0)$ generated by the elements of the form $i_{1*}(\alpha)i_{2*}(\alpha)^{-1}$ for $\alpha \in \pi_1(X_1 \cap X_2, x_0)$. We shall see in what follows that both groups coincide. To that end, we shall require some previous results. To simplify the lenguage, we shall denote by G the group $\pi_1(X_1, x_0) * \pi_1(X_2, x_0)$, by $N \subset G$, the normal subgroup generated by the elements of the form $i_{1*}(\alpha)i_{2*}(\alpha)^{-1}$ for $\alpha \in \pi_1(X_1 \cap X_2, x_0)$. Moreover, if $\lambda : I \longrightarrow X$ is a loop based at x_0 that lies either in X_ν or in $X_1 \cap X_2$, we shall denote its homotopy class either $[\lambda]_\nu \in \pi_1(X_\nu, x_0)$ or $[\lambda]_{12} \in \pi_1(X_1 \cap X_2, x_0)$, $\nu = 1, 2$, respectively. Of course, if λ lies in $X_1 \cap X_2$, then $i_{\nu*}([\lambda]_{12}) = [\lambda]_\nu$; thus, the cosets $[\lambda]_1 N$ and $[\lambda]_2 N$ are equal. We have proved the following.

4.3.19 **Lemma**. *Let λ be a loop on $X_1 \cap X_2$ based at x_0; then $[\lambda]_1 N = [\lambda]_2 N \in G/N$.* □

Let $\lambda : I \longrightarrow X$ be a contractible loop based at x_0 and let $H : I \times I \longrightarrow X$ be a nulhomotopy of λ; that is, $H(s, 0) = \lambda(s)$ and $H(0, t) = H(1, t) = H(s, 1) = x_0$, $s, t \in I$. We have the following construction:

(a) We decompose the square $I^2 = I \times I$ into a lattice of subsquares, as shown in Figure 4.6, in such a way that for each subsquare Q either $H(Q) \subset X_1$ or $H(Q) \subset X_2$. This is possible since, being X_1 and X_2 open sets, one may take a Lebesgue number for the open cover $\{H^{-1}(X_1), H^{-1}(X_2)\}$ of I^2 and the length of the side of each subsquare shorter than one half of this number.

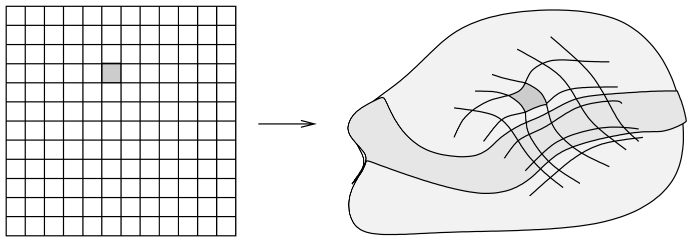


Figure 4.6 Lebesgue subdivision of the square

(b) For each vertex v of the lattice, let $\mu_v : I \longrightarrow X$ be an *auxiliary* path between x_0 and $H(v)$, and let $\overline{\mu}_v$ be the inverse path, so that if $H(v)$ lies either in X_ν, $\nu = 1, 2$, or in $X_1 \cap X_2$, then λ_v also lies thereon. This is possible since all three subspaces are path connected.

(c) For each edge a of the lattice, if we consider it as a path $a : I \longrightarrow I^2$ (in the increasing direction), then $H \circ a : I \longrightarrow X$ is a path between $H(a(0))$ and $H(a(1))$, that by (a) lies either in X_1 or in X_2. Therefore, $\lambda_a = \mu_{a(0)}(H \circ a)\overline{\mu}_{a(1)}$ is a loop, as shown in Figure 4.7(b), that lies either in X_1 or in X_2, as well. Consequently, the elements $[\lambda_a]_1$ or $[\lambda_a]_2$ in G are well defined. Denote by $\widehat{a} \in G/N$ either the coset $[\lambda_a]_1 N$ or $[\lambda_a]_2 N$. By Lemma 4.3.19, if both cosets are defined, then they coincide. Finally, let $a_1^0, \dots, a_n^0$ be the edges at the bottom of the lattice and let $a_1^n, \dots, a_n^n$ be the edges at the top, as indicated in Figure 4.7(a).

4.3.20 **Lemma**. *In G/N the equality $\widehat{a}_1^0 \widehat{a}_2^0 \cdots \widehat{a}_n^0 = 1$ holds.*

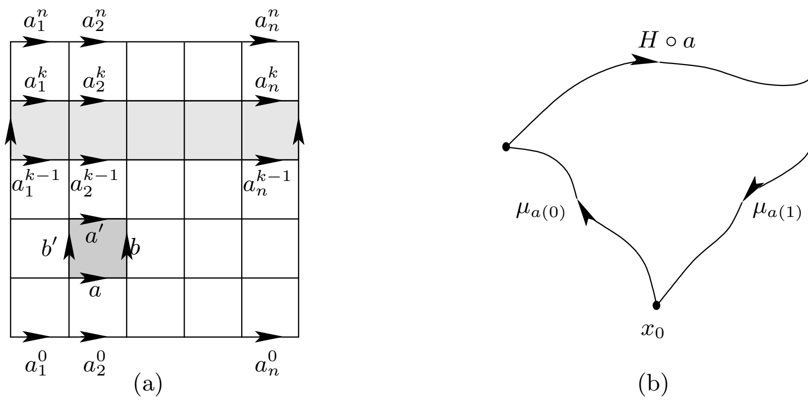


Figure 4.7 Each edge in the Lebesgue subdivision of the square determines a loop in X.

Proof: Let Q be a fixed subsquare with edges a, b, a', b', as indicated in Figure 4.7(a). In Q one has that $ab \simeq b'a'$ rel ∂I, since Q is simply connected (because it is contractible). If we apply H and connect the corresponding auxiliary paths, then we obtain that $\lambda_a \lambda_b \simeq \lambda_{b'} \lambda_{a'}$ rel ∂I either in X_1 or in X_2, according to whether $H(Q) \subset X_1$ or $H(Q) \subset X_2$. In any case, $\widehat{a}\widehat{b} = \widehat{b'}\widehat{a'}$ in G/N, that is, for any subsquare Q one has the equality $\widehat{a} = \widehat{b'}\widehat{a'}\widehat{b}^{-1}$.

If we now take a whole row of subsquares, as the one shown in shade in Figure 4.7(a), and multiply the corresponding equalities, then we obtain

$$\widehat{a}_1^{i-1}\widehat{a}_2^{i-1} \cdots \widehat{a}_n^{i-1} = \widehat{a}_1^{i}\widehat{a}_2^{i} \cdots \widehat{a}_n^{i} \,,$$

since the elements $\widehat{b}_i^j$ corresponding to the middle edges cancel out and $\widehat{b}_i^0 = 1 = \widehat{b}_i^n$, since the homotopy H is constant on the vertical sides of I^2. Inductively, we obtain the equality

$$\widehat{a}_1^{0}\widehat{a}_2^{0} \cdots \widehat{a}_n^{0} = \widehat{a}_1^{n}\widehat{a}_2^{n} \cdots \widehat{a}_n^{n} \,.$$

But, since H is constant also on the top side of I^2, $\widehat{a}_i^n = 1$, $i = 1, 2, \ldots, n$, and this proves the desired equality. $\square$

We are now ready to prove the Seifert–van Kampen theorem, namely to identify $\ker(j_{1*} \cdot j_{2*})$.

4.3.21 **Theorem**. (Seifert–van Kampen) *Take $X = X_1 \cup X_2$ with X_1, X_2 open. If X_1, X_2 and $X_1 \cap X_2$ are nonempty and path connected, then, for $x_0 \in X_1 \cap X_2$,*

$$\pi_1(X, x_0) \cong \pi_1(X_1, x_0) * \pi_1(X_2, x_0)/N \,,$$

where N is the normal subgroup generated by the set

$$\{i_{1*}(\alpha) i_{2*}(\alpha)^{-1} \mid \alpha \in \pi_1(X_1 \cap X_2, x_0)\} \,.$$

Proof: In terms of the notation introduced above, we have to prove that $\ker(j_{1*}\cdot j_{2*})$ is $N = N_{\{i_{1*}(\alpha)i_{2*}(\alpha)^{-1}|\alpha\in\pi_1(X_1\cap X_2,x_0)\}}$.

Let $\beta \in G$ be an element such that $(j_{1*}\cdot j_{2*})(\beta) = 1$; we shall see that $\beta \in N$. We may write $\beta = \alpha_1\alpha_2\cdots\alpha_k \in G$, where either $\alpha_i \in \pi_1(X_1, x_0)$ or $\alpha_i \in \pi_1(X_2, x_0)$, even though we do not require that this decomposition is necessarily reduced. Let λ_i be a loop in X_ν that represents α_i, $\nu = 1$ or 2. Since $(j_{1*}\cdot j_{2*})(\beta) = 1$, $[\lambda_1][\lambda_2]\cdots[\lambda_k] = 1$ in $\pi_1(X, x_0)$, where the homotopy classes are taken in X. We subdivide I in k subintervals of the same length and we take the loop $\lambda : I \longrightarrow X$ that in the ith interval coincides with λ_i conveniently reparametrized, $i = 1, 2, \ldots, k$. The last equality means that λ is a contractible loop. Let $H : I^2 \longrightarrow X$ be a nullhomotopy for λ. We decompose I^2 in subsquares as in the preceding construction, in such a way that each of the k subintervals of I is the union of some of the edges $a_1^0, a_2^0, \ldots, a_n^0$ in Figure 4.7(a). This is possible if we take n to be a large enough multiple of k.

If the first interval, where λ_1 is defined, is the union of $a_1^0, a_2^0, \ldots, a_{i_1}^0$, then $\lambda_1 \simeq \lambda_{a_1^0}\lambda_{a_2^0}\cdots\lambda_{a_{i_1}^0}$ rel ∂I either in X_1 or in X_2, according to whether λ_1 lies in X_1 or in X_2. In any case, one has that $\alpha_1 N = \widehat{a}_1^0\widehat{a}_2^0\cdots\widehat{a}_{i_1}^0$ in G/N. There are corresponding equalities for the remaining $k-1$ subintervals, which when multiplied with each other, yield the equality $\beta N = (\alpha_1 N)(\alpha_2 N)\cdots(\alpha_k N) = \widehat{a}_1^0\widehat{a}_2^0\cdots\widehat{a}_n^0$ in G/N. From Lemma 4.3.20 we then have that $\beta N = 1 \in G/N$, that is, $\beta \in N$, as we wished to prove. □

4.3.22 **Corollary**. *Under the assumptions of* 4.3.21 *one has the following.*

(a) *If X_2 is simply connected, then*

$$j_{1*} : \pi_1(X_1, x_0) \longrightarrow \pi_1(X, x_0)$$

is an epimorphism and $\ker j_{1*}$ *is the normalizer of the subgroup*

$$i_{1*}(\pi_1(X_1 \cap X_2, x_0))\,.$$

(b) *If $X_1 \cap X_2$ is simply connected, then*

$$j_{1*}\cdot j_{2*} : \pi_1(X_1, x_0) * \pi_1(X_2, x_0) \longrightarrow \pi_1(X, x_0)$$

is an isomorphism.

(c) *If X_2 and $X_1 \cap X_2$ are simply connected, then*

$$j_{1*} : \pi_1(X_1, x_0) \longrightarrow \pi_1(X, x_0)$$

is an isomorphism. □

Take $X = X_1 \cup X_2$ so that $X_1, X_2 \subset X$ are closed and there are open neighborhoods V_1 of X_1 and V_2 of X_2 in X, and strong deformation retractions $r_1 : V_1 \longrightarrow X_1$, $r_1 : V_2 \longrightarrow X_2$, that restrict to a strong deformation retraction $r = r_1|_{V_1 \cap V_2} = r_2|_{V_1 \cap V_2} : V_1 \cap V_2 \longrightarrow X_1 \cap X_2$. In this case, Theorem 4.3.21 converts into the following.

4.3.23 **Theorem**. *Take $X = X_1 \cup X_2$ with X_1, X_2 closed sets that satisfy the conditions above. If X_1, X_2, and $X_1 \cap X_2$ are nonempty and path connected, then*

$$\pi_1(X, x_0) \cong \pi_1(X_1, x_0) * \pi_1(X_2, x_0)/N_{\{i_{1*}(\alpha)i_{2*}(\alpha)^{-1} | \alpha \in \pi_1(X_1 \cap X_2, x_0)\}} ,$$

for $x_0 \in X_1 \cap X_2$. □

4.3.24 EXERCISE. Let M be a connected n-manifold, $n \geq 3$, and take $M^* = M - B^\circ$, where B is an n-ball embedded in M. Prove that $\pi_1(M^*) \cong \pi_1(M)$. (*Hint:* $M = M_1 \cup M_2$, where $M_1 = M - \{b\}$, and $M_2 = B^\circ$, where $b \in B^\circ$. Then the inclusion $M^* \hookrightarrow M_1$ is a homotopy equivalence, M_2 is contractible, and $M_1 \cap M_2 \approx \mathbb{S}^{n-1}$ is simply connected. Apply 4.3.22(c).)

4.3.25 EXERCISE. Let M and N be connected n-manifolds, $n \geq 3$. Prove that

$$\pi_1(M \# N) \cong \pi_1(M) * \pi_1(N) ,$$

if $M \# N$ is their connected sum (2.1.32). (*Hint:* Apply the previous exercise.)

4.3.26 EXERCISE. Under the assumptions of the Seifert–van Kampen theorem prove that the fundamental group $\pi_1(X, x_0)$ of X has the following universal property that characterizes it: *Given homomorphisms*

$$f_1 : \pi_1(X_1, x_0) \longrightarrow H \quad \textit{and} \quad f_2 : \pi_1(X_2, x_0) \longrightarrow H$$

such that $f_1 \circ i_{1*} = f_2 \circ i_{2*} : \pi_1(X_1 \cap X_2, x_0) \longrightarrow H$, there exists a unique homomorphism

$$f : \pi_1(X, x_0) \longrightarrow H$$

such that $f \circ j_{1*} = f_1$ and $f \circ j_{2*} = f_2$; that is, such that if the outer square in the diagram

$$\begin{tikzcd}
 & \pi_1(X_1 \cap X_2, x_0) \arrow[dl, "i_{1*}"'] \arrow[dr, "i_{2*}"] & \\
\pi_1(X_1, x_0) \arrow[r, "j_{1*}"] \arrow[dr, "f_1"'] & \pi_1(X, x_0) \arrow[d, dashed, "f"] & \pi_1(X_2, x_0) \arrow[l, "j_{2*}"'] \arrow[dl, "f_2"] \\
 & H &
\end{tikzcd}$$

commutes, then the two triangles obtained also commute. (This universal property means that $\pi_1(X, x_0)$ is the colimit of the diagram

$$\begin{tikzcd} & \pi_1(X_1 \cap X_2, x_0) \arrow[dl, "i_{1*}"'] \arrow[dr, "i_{2*}"] & \\ \pi_1(X_1, x_0) & & \pi_1(X_2, x_0)\,. \end{tikzcd}$$

(Cf. 1.2.2 and observe that the statement means that a pushout diagram in the category $\mathcal{T}op$ of topological spaces maps to a pushout diagram in the category $\mathcal{G}rp$ of groups under the fundamental group functor.)

4.4 Applications of the Seifert–van Kampen theorem

Several constructions in topological spaces can be analyzed from the point of view of the Seifert–van Kampen theorem in order to study their fundamental group. Consider, in the first place, the following assertion.

4.4.1 **Proposition**. *The fundamental group of a wedge of k copies of the circle, $\mathbb{S}^1_1 \vee \cdots \vee \mathbb{S}^1_k$, is freely generated by the elements*

$$\alpha_1, \ldots, \alpha_k \in \pi_1(\mathbb{S}^1_1 \vee \cdots \vee \mathbb{S}^1_k, x_0)\,,$$

where x_0 is the base point of the wedge obtained from all of the elements $1 \in \mathbb{S}^1_i$, and the class α_i is represented by the canonical loop $\lambda_i : I \longrightarrow \mathbb{S}^1_i \hookrightarrow \mathbb{S}^1_1 \vee \cdots \vee \mathbb{S}^1_k$ given by $\lambda_i(t) = \mathrm{e}^{2\pi \mathrm{i}t} \in \mathbb{S}^1_i$. Therefore,

$$\pi_1(\underbrace{\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1}_{k}) \cong \underbrace{\mathbb{Z} * \cdots * \mathbb{Z}}_{k}.$$

Proof: By induction on k. For a wedge of two circles, $X = \mathbb{S}^1_1 \vee \mathbb{S}^1_2$, take $X_1 = \mathbb{S}^1_1 \vee (\mathbb{S}^1_2 - \{-1\})$ and $X_2 = (\mathbb{S}^1_1 - \{-1\}) \vee \mathbb{S}^1_2$. Then X, X_1, and X_2 satisfy the hypotheses of the Seifert–van Kampen theorem, and since $X_1 \cap X_2$ is homeomorphic to the open cross $\mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R}$, it is contractible. Thus using 4.3.22(b) and the fact that the inclusions $\mathbb{S}^1_1 \hookrightarrow X_1$ and $\mathbb{S}^1_2 \hookrightarrow X_2$ induce isomorphisms in the fundamental groups, one has that $\pi_1(\mathbb{S}^1_1, 1) * \pi_1(\mathbb{S}^1_2, 1) \longrightarrow \pi_1(\mathbb{S}^1_1 \vee \mathbb{S}^1_2, x_0)$ is an isomorphism. Moreover, since the classes α_1 and α_2 come from the canonical generators of $\pi_1(\mathbb{S}^1_1, 1)$ and $\pi_1(\mathbb{S}^1_2, 1)$, they are the generators of $\pi_1(\mathbb{S}^1_1 \vee \mathbb{S}^1_2, x_0)$ as a free group. Therefore, the group $\pi_1(\mathbb{S}^1 \vee \mathbb{S}^1, x_0)$ is isomorphic to $\mathbb{Z} * \mathbb{Z}$.

If for a wedge of $k-1$ copies of $\mathbb{S}^1$ the result is true, then take

$$X_1 = \mathbb{S}^1_1 \vee \cdots \vee \mathbb{S}^1_{k-1} \vee (\mathbb{S}^1_k - \{-1\})\,,$$

which has the same homotopy type via the inclusion of $\mathbb{S}^1_1 \vee \cdots \vee \mathbb{S}^1_{k-1}$, and take

$$X_2 = (\mathbb{S}^1_1 - \{-1\}) \vee \cdots \vee (\mathbb{S}^1_{k-1} - \{-1\}) \vee \mathbb{S}^1_k\,,$$

which also via the inclusion has the same homotopy type of $\mathbb{S}^1_k$. Since $X_1 \cap X_2$ is homeomorphic to a "star" with $2k$ rays, it is contractible, and, again by 4.3.22(b),

$$\pi_1(\mathbb{S}^1_1 \vee \cdots \vee \mathbb{S}^1_{k-1}, x_0) * \pi_1(\mathbb{S}^1_k, 1) \longrightarrow \pi_1(\mathbb{S}^1_1 \vee \cdots \vee \mathbb{S}^1_k, x_0)$$

is an isomorphism. And as was the case for $k = 2$, we have that $\alpha_1, \ldots, \alpha_k$ are its generators as a free group, as we wanted to prove. □

4.4.2 EXERCISE. Prove the preceding proposition using version 4.3.23 of the Seifert–van Kampen theorem insetad of 4.3.22(b).

The Seifert–van Kampen theorem can be used to study the fundamental group of a space with a cell attached.

4.4.3 **Proposition**. *For a path-connected space Y, let $f : \mathbb{S}^{n-1} \longrightarrow Y$ be continuous, $n \geq 3$. If $y_0 \in Y$, then the canonical inclusion $i : Y \hookrightarrow Y \cup_f e^n$ induces an isomorphism*

$$i_* : \pi_1(Y, y_0) \xrightarrow{\cong} \pi_1(Y \cup_f e^n, y_0)\,.$$

Proof: Let $X = Y \cup_f e^n$ and let $q : \mathbb{B}^n \sqcup Y \longrightarrow X$ be the identification. The subspaces $X_1 = q((\mathbb{B}^n - \{0\}) \sqcup Y)$ and $X_2 = q(\overset{\circ}{\mathbb{B}}{}^n)$ are open. Notice that the canonical inclusion $Y \hookrightarrow X_1$ is a homotopy equivalence and that X_2 is contractible. Moreover, the intersection $X_1 \cap X_2 \approx \overset{\circ}{\mathbb{B}}{}^n - \{0\}$, which has the same homotopy type of the sphere $\mathbb{S}^{n-1}$, is simply connected, since $n \geq 3$. Therefore, by 4.3.22(c), if $x_0 \in X_1 \cap X_2$, then the inclusion $X_1 \hookrightarrow X$ induces an isomorphism $\pi_1(X_1, x_0) \longrightarrow \pi_1(X, x_0)$.

Take now a path $\omega : x_0 \simeq y_0$ in X_1. Then the homomorphism induced by the inclusion $i_* : \pi_1(Y, y_0) \longrightarrow \pi_1(X, y_0)$ factors as indicated in the commutative diagram

$$\begin{tikzcd}
\pi_1(Y, y_0) \arrow[r, "i_*"] \arrow[d, "\cong"] & \pi_1(X, y_0) \arrow[dd, "\cong"', "\varphi_\omega"] \\
\pi_1(X_1, y_0) \arrow[d, "\varphi_\omega"', "\cong"] & \\
\pi_1(X_1, x_0) \arrow[r, "\cong"] & \pi_1(X, x_0)\,,
\end{tikzcd}$$

where the unnamed isomorphisms are induced by inclusions and the φ_ω are the isomorphisms of 4.1.21 in X_1 and in X, respectively. Therefore, i_* is an isomorphism, as desired. □

Let us now see what happens in the case of the attachment of a 2-cell.

4.4.4 **Proposition**. *Let $f : \mathbb{S}^1 \longrightarrow Y$ be continuous. If $\lambda_f : I \longrightarrow Y$ is the loop given by $\lambda_f(t) = f(\mathrm{e}^{2\pi \mathrm{i}t})$ and $\omega : y_0 \simeq f(1)$ is a path in Y, then the inclusion $i : Y \hookrightarrow Y \cup_f e^2$ induces an epimorphism $i_* : \pi_1(Y, y_0) \longrightarrow \pi_1(Y \cup_f e^2, y_0)$, and its kernel is the normal subgroup N_{α_f} generated by the element $\alpha_f = [\omega\lambda_f\overline{\omega}] \in \pi_1(Y, y_0)$. Therefore,*

$$\pi_1(Y \cup_f e^2, y_0) \cong \pi_1(Y, y_0)/N_{\alpha_f}\,.$$

The group N_{α_f} does not depend on the path ω, since the loop $\mu_f = \omega\lambda_f\overline{\omega}$ that surrounds the cell is contractible in $Y \cup_f e^2$, because it can be contracted over the cell, as shown in Figure 4.8. Before attaching the cell one has $\mu_f \not\simeq 0$, but after doing it, $\mu_f \simeq 0$. Therefore, $i_*(\alpha_f) = [\mu_f] = 1$ in $\pi_1(Y \cup_f e^2, y_0)$. One says that the element $\alpha_f \in \pi_1(Y, y_0)$ *is killed by attaching the* 2*-cell* using the map f.

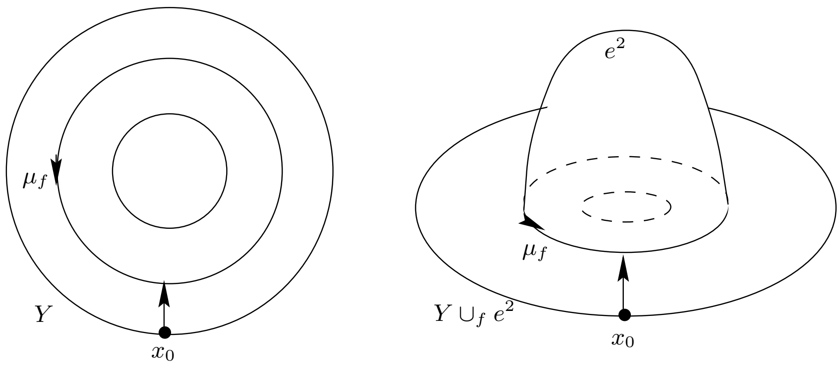


Figure 4.8 Killing a path by attaching a cell

Proof: Using the same notation as in the previous proof, we have that the canonical inclusion $Y \hookrightarrow X_1$ is a homotopy equivalence and that X_2 is contractible. Moreover, the intersection $X_1 \cap X_2 \approx \overset{\circ}{\mathbb{B}}{}^2 - \{0\}$ has the same homotopy type of the circle $\mathbb{S}^1$ and so is not simply connected. By 4.3.22(a) the inclusion $X_1 \hookrightarrow X$ induces an epimorphism on the fundamental group, and so $i_* : \pi_1(Y, y_0) \longrightarrow \pi_1(X, y_0)$ is an epimorphism.

On the other hand, if $z_0 = q(0) = q(1)$, then the loop $\lambda'_f : I \longrightarrow X$ given by $\lambda'_f(t) = q(\frac{1}{2}\mathrm{e}^{2\pi i t})$, which indeed lies inside $X_1 \cap X_2$, generates $\pi_1(X_1 \cap X_2, z_0) \cong \mathbb{Z}$. Also, the deformation retraction of X_2 into Y deforms λ'_f in λ_f. Letting $j : X_1 \hookrightarrow X$ denote the inclusion, we know from 4.3.22(a) that $\ker(j_*)$ is generated as a normal subgroup by the element $[\lambda'_f]$, and so $\ker(i_* : \pi_1(Y, f(1)) \longrightarrow \pi_1(X, f(1)))$ is generated by $[\lambda_f]$, and, as in the previous proof, $\ker(i_* : \pi_1(Y, y_0) \longrightarrow \pi_1(X, y_0))$ is generated by α_f. $\square$

Inductively, it is possible to prove the following result.

4.4.5 **Corollary**. *If the 2-cells $e_1^2, e_2^2, \ldots, e_k^2$ are attached to Y using the maps $f_1, f_2, \ldots, f_k : \mathbb{S}^1 \longrightarrow Y$, then*

$$\pi_1(Y \cup e_1^2 \cup e_2^2 \cup \cdots \cup e_k^2, y_0) \cong \pi_1(Y, y_0)/N_{\{\alpha_{f_1}, \alpha_{f_2}, \ldots, \alpha_{f_k}\}}\,. \qquad \square$$

4.4.6 EXAMPLES.

(a) For any integer $k \geq 1$, let $X_k = \mathbb{S}^1 \cup e^2$, where the cell is attached using the map $g_k : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ of degree k, $g_k(\zeta) = \zeta^k$. If $[\alpha] \in \pi_1(\mathbb{S}^1, 1)$ is the canonical generator, then $\pi_1(X_k, 1) \cong \pi_1(\mathbb{S}^1, 1)/N_{\{\alpha_k\}}$, where $\alpha_k = [\lambda_{g_k}] \in \pi_1(\mathbb{S}^1, 1)$. By 4.2.13, $\alpha_k = \alpha_1^k \in \pi_1(\mathbb{S}^1, 1)$; that is, α_k is the kth power of the canonical generator. Therefore,

$$\pi_1(X_k, 1) \cong \mathbb{Z}/k,$$

that is, this fundamental group is cyclic of order k.

(b) The construction of (a) for $k = 2$ produces $X_2 \approx \mathbb{RP}^2$, that is, the projective plane. Therefore,

$$\pi_1(\mathbb{RP}^2) \cong \mathbb{Z}/2.$$

There are several ways of grasping this fact. If, for instance, we realize $\mathbb{RP}^2$ by identifying antipodal points in the boundary of $\mathbb{B}^2$, then the map $\lambda_1 : I \longrightarrow \mathbb{B}^2$ given by $\lambda_1(t) = \mathrm{e}^{\pi it}$ determines a loop λ' in $\mathbb{RP}^2$ (see Figure 4.9(a)). Since by 4.3.22(a), $\pi_1(\mathbb{S}^1) \longrightarrow \pi_1(\mathbb{RP}^2)$ is an epimorphism, the class $[\lambda']$ generates $\pi_1(\mathbb{RP}^2)$; that is, this group is cyclic. Defining $\lambda_2(t) = \mathrm{e}^{\pi \mathrm{i}(t+1)}$ and $\lambda = \lambda_1\lambda_2$, we have that λ surrounds $\mathbb{B}^2$ once and therefore is contractible. Since λ_1, λ_2, and λ' all determine the same homotopy class in $\pi_1(\mathbb{RP}^2)$, $[\lambda']^2 = 1 \in \pi_1(\mathbb{RP}^2)$, that is, this group is cyclic of order 2.

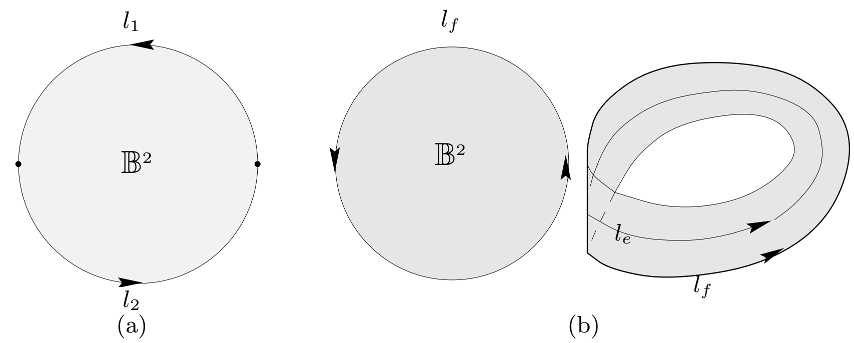


Figure 4.9 The square of the generator of $\pi_1(\mathbb{RP}^2)$ is trivial.

Another way of looking at this is the following. $\mathbb{RP}^2$ is obtained by attaching a 2-cell to the Moebius band M along its boundary, which is homeomorphic to $\mathbb{S}^1$. Since M has the same homotopy type of $\mathbb{S}^1$, the equatorial loop λ_e that surrounds the equator of M once (see 4.2.9(c)) generates $\pi_1(M)$ as an infinite cyclic group. If $f : \mathbb{S}^1 \longrightarrow \partial M \hookrightarrow M$ is a homeomorphism onto the boundary of M, the loop λ_f in $\mathbb{RP}^2 = M \cup_f e^2$ is such that it deforms inside M to the equator to become λ_e^2 (see Figure 4.9(b)). Consequently, $[\lambda_e]^2 = 1 \in \pi_1(\mathbb{RP}^2)$, so we again see that this group is cyclic of order 2.

Considering $\mathbb{RP}^2$ as a quotient of $\mathbb{S}^2$ by identifying antipodal points, we may repeat the construction above. A path $\lambda : I \longrightarrow \mathbb{S}^2$ that uniformly travels along one-half of the equator of the sphere determines in $\mathbb{RP}^2$ a loop μ, generating $\pi_1(\mathbb{RP}^2)$ and whose square comes from λ^2. Since it travels along the whole equator of $\mathbb{S}^2$, the loop λ_2 can be deformed into a constant loop, and so $[\mu]^2 = 1$ in $\pi_1(\mathbb{RP}^2)$.

(c) As we saw in 3.3.18, the *orientable surface of genus g*, S_g, is obtained by attaching a 2-cell to the wedge of $2g$ circles $S^1_{2g} = \mathbb{S}^1_{a_1} \vee \mathbb{S}^1_{b_1} \vee \cdots \vee \mathbb{S}^1_{a_g} \vee \mathbb{S}^1_{b_g}$ with the map $f_g : \mathbb{S}^1 \longrightarrow S^1_{2g}$, such that as the argument travels around the circle counterclockwise, the values of the map first go around $\mathbb{S}^1_{a_1}$ counterclockwise, then $\mathbb{S}^1_{b_1}$ also counterclockwise, then again $\mathbb{S}^1_{a_1}$ but now clockwise, and then $\mathbb{S}^1_{b_1}$ clockwise, and so on, and finishing by going around $\mathbb{S}^1_{b_g}$ clockwise. Then the associated loop $\lambda_g = \lambda_{f_g} : I \longrightarrow S^1_{2g}$ is the loop product $\lambda_{a_1}\lambda_{b_1}\overline{\lambda}_{a_1}\overline{\lambda}_{b_1}\lambda_{a_2}\cdots\overline{\lambda}_{a_g}\overline{\lambda}_{b_g}$, where λ_{a_i} and λ_{b_i} are the canonical loops in $\mathbb{S}^1_{a_i} = \mathbb{S}^1$ and $\mathbb{S}^1_{b_i} = \mathbb{S}^1$, and $\overline{\lambda}_{a_i}$ and $\overline{\lambda}_{b_g}$ are their inverses, $i = 1, \ldots, g$. By 4.4.1, $\pi_1(S^1_{2g})$ is freely generated by the classes $\alpha_i = [\lambda_{a_i}]$, $\beta_i = [\lambda_{b_i}]$.

By 4.4.4, $\pi_1(S_g) \cong \pi_1(S^1_{2g})/N_{\alpha_{f_g}}$. That is,

$$\pi_1(S_g) \cong \underbrace{\mathbb{Z} * \cdots * \mathbb{Z}}_{2g} / N_{\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_g\beta_g\alpha_g^{-1}\beta_g^{-1}}\,,$$

where α_i is the generator of the $(2i-1)$th copy of $\mathbb{Z}$ and β_i of the $2i$th, $i = 1, \ldots, g$. In terms of generators and relations, this fact is usually written as

$$\pi_1(S_g) = \langle \alpha_1, \beta_1, \ldots, \alpha_g, \beta_g \mid \alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_g\beta_g\alpha_g^{-1}\beta_g^{-1}\rangle\,,$$

and one says that this group *has as generators* the elements $\alpha_1, \beta_1, \ldots,$ α_g, β_g subject only to the *relation*

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_g\beta_g\alpha_g^{-1}\beta_g^{-1} = 1\,.$$

(d) Analogously to (c) we can compute the fundamental group of a *nonorientable surface N_g of genus g* defined as the result of attaching a 2-cell to a wedge of g circles $S^1_g = \mathbb{S}^1_{a_1} \vee \cdots \vee \mathbb{S}^1_{a_g}$ but now with the map $f_g : \mathbb{S}^1 \longrightarrow S^1_g$ such that as the argument travels around the circle counterclockwise, the values of the map first go around $\mathbb{S}^1_{a_1}$ counterclockwise, then $\mathbb{S}^1_{a_2}$ also counterclockwise, and so on, and finishing by going around $\mathbb{S}^1_{a_g}$ counterclockwise. Therefore, we now have

$$\pi_1(N_g) \cong \underbrace{\mathbb{Z} * \cdots * \mathbb{Z}}_{g} / N_{\alpha_1^2 \cdots \alpha_g^2} ,$$

where α_i is the generator of the ith copy $\mathbb{Z}$. In terms of generators and relations, one has

$$\pi_1(N_g) = \langle \alpha_1, \ldots, \alpha_g \mid \alpha_1^2 \cdots \alpha_g^2 \rangle ;$$

that is, this group has as generators the elements $\alpha_1, \ldots, \alpha_g$ subject to the one relation $\alpha_1^2 \cdots \alpha_g^2 = 1$.

Using examples (c) and (d) above, we can distinguish surfaces of different genus.

4.4.7 **Corollary**. *No two surfaces in the list*

$$S_0, S_1, S_2, \ldots, N_1, N_2, \ldots$$

have the same homotopy type, and in particular, they are not homeomorphic.

Proof: If the fundamental groups of these surfaces are abelianized, we have

$$\pi_1(S_g)^{\mathrm{ab}} \cong \mathbb{Z}^{2g} , \qquad \pi_1(N_h)^{\mathrm{ab}} \cong \mathbb{Z}^{h-1} \times (\mathbb{Z}/2) .$$

Here $\mathbb{Z}^0$ denotes 0. Since no two of these groups are isomorphic, we have that no two of these surfaces have the same homotopy type. This implies that no two of them are homeomorphic. □

4.4.8 Remark. Given a 0-connected topological space X, one may define its *first homology group* by

$$H_1(X) = \pi_1(X)^{\mathrm{ab}} .$$

This is in fact a theorem, but it may be used as an *ad hoc* definition.

4.4.9 Exercise. Let $x_1, x_2, \ldots, x_k \in \mathbb{R}^2$ be different points.

(a) Prove that $\mathbb{R}^2 - \{x_1, x_2, \ldots, x_k\}$ contains as a (strong) deformation retract a subspace X_k homeomorphic to $\underbrace{\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1}_{k}$.

(b) Prove that

$$\pi_1(\mathbb{R}^2 - \{x_1, \ldots, x_k\}) \cong \underbrace{\mathbb{Z} * \cdots * \mathbb{Z}}_{k} ;$$

that is, it is free in k generators. Deduce that $\mathbb{R}^2 - \{x_1, x_2, \ldots, x_k\}$ and $\mathbb{R}^2 - \{x_1, x_2, \ldots, x_l\}$ are homeomorphic if and only if $k = l$.

4.4.10 EXERCISE. Compute the fundamental groups of the following spaces:

(a) $\mathbb{S}^1 \vee \mathbb{S}^2$, $\mathbb{S}^1 \times \mathbb{RP}^2$, $\mathbb{RP}^2 \vee \mathbb{RP}^2$, $\mathbb{RP}^2 \times \mathbb{RP}^2$.

(b) $\mathbb{R}^3 - C$, where C is the circle $x^2 + y^2 = 1$, $z = 0$.

(c) $(\mathbb{S}^1 \times \mathbb{S}^1) \cup e^2$, where the 2-cell is attached using the map $f(\zeta) = (\zeta^2, \zeta^3)$.

4.4.11 EXERCISE. Prove the theorem on invariance of dimension 1.1.9 for $m = 2$ and $n > 2$; in other words, prove that $\mathbb{R}^2$ and $\mathbb{R}^n$ (resp. $\mathbb{S}^2$ and $\mathbb{S}^n$, $\mathbb{B}^2$ and $\mathbb{B}^n$), $n > 2$, are not homeomorphic. (*Hint:* $\pi_1(\mathbb{R}^2 - \{x\}) \cong \mathbb{Z}$, while $\pi_1(\mathbb{R}^n - \{y\}) = 1$ for $n > 2$. Cf. the proof of 4.4.3, or 4.3.2.)

4.4.12 EXERCISE. Let G be a *finitely presented group*, that is G has finitely many generators $a_1, \ldots, a_k$ and finitely many relations $r_1, \ldots, r_l$; in symbols, $G = \langle a_1, \ldots, a_k \mid r_1, \ldots, r_l \rangle$. Prove that there is a topological space X such that $\pi_1(X) \cong G$. (*Hint:* Take $Y = \bigvee_{j=1}^k \mathbb{S}^1_j$ and kill the relations in $\pi_1(Y)$ by attaching 2-cells as in 4.4.4.)

To finish this chapter, let us recall that every 3-dimensional closed, connected, and orientable manifold is the union of two handle bodies (see 2.3.14). Applying the Seifert–van Kampen theorem to this decomposition, we can compute the fundamental group of any 3-manifold of this kind, at least potentially, since necessarily we need to know in each case, what the corresponding Heegaard decomposition looks like.

4.4.13 EXAMPLE. Take the handle body H_g defined in 2.3.13, with genus $g \geq 1$. By the classification theorem for surfaces 2.2.23 we know that $A_g = \partial H_G \approx S_g$, where S_g is the connected, closed, and orientable surface of genus g. Therefore, $\pi_1(A_g) = \langle \alpha_1, \beta_1, \ldots, \alpha_g, \beta_g \mid \alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1} \cdots \alpha_g\beta_g\alpha_g^{-1}\beta_g^{-1} \rangle$. On the other hand, as indicated in 3.3.16(c), $H_g \simeq \mathbb{S}^1_1 \vee \cdots \vee \mathbb{S}^1_g$, so that $\pi_1(H_g)$ is a free group of rank g. Let $i : A_g \hookrightarrow H_g$ be the inclusion. We may choose the generators of $\pi_1(A_g)$ in such a way that $i_*(\alpha_1), \ldots, i_*(\alpha_g)$ is a set of generators of $\pi_1(H_g)$ and $i_*(\beta_1) = \cdots = i_*(\beta_g) = 1$. If we denote by $q : H_g \sqcup H'_g \longrightarrow M = H_g \cup_\varphi H'_g$, $H_g = H'_g$, the identification along a homeomorphism $\varphi : A_g \longrightarrow A_g$, we have elements $a_\nu = q_*(\alpha_\nu), b_\nu = q_*(\beta_\nu), a'_\nu = q_*\varphi_*(\alpha_\nu), b'_\nu = q_*\varphi_*(\beta_\nu) \in \pi_1(M)$ such that, indeed, $b_1 = \cdots = b_g = b'_1 = \cdots = b'_g = 1$. Moreover, the homeomorphism $\varphi : A_3 \longrightarrow A_3$ is such that $\varphi_*(\alpha_\nu) = r_\nu(\alpha_1, \beta_1, \ldots, \alpha_g, \beta_g)$, $\varphi_*(\beta_\nu) = s_\nu(\alpha_1, \beta_1, \ldots, \alpha_g, \beta_g)$, where r_ν and s_ν are certain words in the generators $\alpha_1, \beta_1, \ldots, \alpha_g, \beta_g$.

If we call $W = q(H_g)$ and $W' = q(H'_g)$, and $T = q(A_g) = q(\varphi(A_g))$, $A_g \subset H_g$, we have that $M = W \cup W'$ and $T = W \cap W' \approx A_g \approx S_g$. By the choice of the generators,

$$\pi_1(W) = \langle a_1, \dots a_g | - \rangle \,,$$

$$\pi_1(W') = \langle a'_1, \dots a'_g | - \rangle \,.$$

Besides, $\pi_1(T)$ has two possible systems of generators:

$$\{x_1, y_1, \dots, x_g, y_g\}, \quad x_\nu = q_*(\alpha_\nu)\,,\ y_\nu = q_*(\beta_\nu)\,,$$

$$\{x'_1, y'_1, \dots, x'_g, y'_g\}, \quad x'_\nu = q_*\varphi_*(\alpha_\nu)\,,\ y'_\nu = q_*\varphi_*(\beta_\nu)\,,$$

$\nu = 1, \dots, g$, that are related to each other via the equations

$$x'_\nu = r_\nu(x_1, y_1, \dots, x_g, y_g)\,, \quad y'_\nu = s_\nu(x_1, y_1, \dots, x_g, y_g)\,.$$

Since the inclusions $j : T \hookrightarrow W$, $j' : T \hookrightarrow W'$ are such that $j_*(x_\nu) = a_\nu$, $j_*(y_\nu) = 1$, $j'_*(x'_\nu) = a'_\nu$, $j'_*(y'_\nu) = 1$, the equations

$$a'_\nu = r_\nu(a_1, 1, \dots, a_g, 1)\,, \quad 1 = b'_\nu = s_\nu(a_1, 1, \dots, a_g, 1)$$

hold. So we have, applying version 4.3.23 of the Seifert–van Kampen thorem, that

$$\pi_1(M) \cong \langle a_1, \dots, a_g, a'_1, \dots, a'_g | a_\nu^{-1} r_\nu(a_1, 1, \dots a_g, 1), s_\nu(a_1, 1, \dots a_g, 1) \rangle \,,$$

and since $a_1, \dots, a_g$ can be expressed in terms of $a'_1, \dots a'_g$, these last elements are enough to generate $\pi_1(M)$, and we obtain

$$\pi_1(M) \cong \langle a'_1, \dots, a'_g \mid s_\nu(a_1, 1, \dots a_g, 1) \text{ para } 1 \leq \nu \leq g \rangle \,.$$

This can be easily interpreted by stating that $\pi_1(M)$ is generated by the equatorial loops a'_ν in $\partial W'$. And given the relation $\varphi_*(\beta_\nu) = s_\nu(\alpha_1, \beta_1, \dots, \alpha_g, \beta_g)$, the relation $s_\nu(a_1, 1, \dots a_g, 1) = 1$ holds, since the meridional loops β_ν and β'_ν are nullhomotopic in M.

4.4.14 Example. It is interesting to see now the result of 4.3.2 in the context of the previous example, when $n = 3$, namely to compute the fundamental group of the 3-sphere be decomposing it as the union of two solid tori, according to proposition 2.3.6. In this case, $\mathbb{S}^3 = H_1 \cup_\varphi H'_1$, where $H_1 = H'_1 = \mathbb{S}^1 \times \mathbb{B}^2$, and $\varphi : \mathbb{S}^1 \times \mathbb{S}^1 \longrightarrow \mathbb{S}^1 \times \mathbb{S}^1$ is such that $\varphi(x, y) = (y, x)$. With the notation above, it results that $a'_1 = r_1(a_1, b_1) = b_1 = 1, 1 = b'_1 = s_1(a_1, b_1) = a_1 \in \pi_1(\mathbb{S}^3)$, and thus $\pi_1(\mathbb{S}^3) = \langle a'_1 \mid a'_1 \rangle$, but obviously, this is the trivial group.

4.4.15 Note. The last year of the nineteenth century Henri Poincaré proposed a conjecture that has puzzled topologists being one of the most powerful motors of research on 3-manifolds during all of the twentieth century, and that all along the century could not be either proved or disproved by showing a counterexample.

In the preceding example we proved again that the 3-sphere is simply connected. Poincaré conjectured that this is the only such 3-manifold:

THE POINCARÉ CONJECTURE. *Every closed, connected, and simply connected* 3-*manifold is homeomorphic to* $\mathbb{S}^3$.

Hempel's book [14] deals with several aspects of this conjecture.

Let us recall that a group is *finitely presented* if it is given by a finite number of generators and a finite number of relations among them. Examples of finitely presented groups are the fundamental groups of the closed surfaces (see 4.4.6(c) and (d)), as well as the fundamental groups of the closed 3-manifolds (see 4.4.13).

It is interesting to observe the following theorem about 4-manifolds, whose proof lies beyond the scope of this book. For its proof we refer the reader to [11].

4.4.16 **Theorem**. *For every finitely presented group G there is a closed* 4-*manifold M such that $\pi_1(M) \cong G$.* □

4.4.17 NOTE. Compare this result with 4.4.12.

4.4.18 NOTE. We remarked in Chapter 2 that it is not possible to classify *all* 4-manifolds. Under this fact lies an algebraic result that states that there is no algorithm that allows to decide whether a finitely presented group is the trivial group or not. This, due to Theorem 4.4.16, makes very complicated to decide whether two 4-manifolds are homeomorphic or not; therefore, it is not clear how to produce an algorithm to classify all 4-manifolds. This explains why it is necessary to ask for Freedman's restriction in his classification theorem 2.3.16 that the 4-manifolds are simply connected.

4.4.19 NOTE. If M is a 4-manifold of the same homotopy type of the 4-sphere $\mathbb{S}^4$, then, since the homology of both coincides, and consequently also the associated bilinear forms are isomorphic, then by Freedman's theorem 2.3.16, M must be homeomorphic to $\mathbb{S}^4$. Therefore, as a corollary of this theorem, the Poincaré conjecture in dimension 4 is established as follows.

4.4.20 **Theorem**. *If a* 4-*dimensional topological manifold M is homotopy equivalent to $\mathbb{S}^4$, then M is homeomorphic to $\mathbb{S}^4$.* □

Chapter 5 Covering maps

In this chapter we shall analyze the concept of covering map. This is an important notion in several branches of mathematics. We shall study it here with respect to its relationship to the fundamental group.

5.1 Definitions and examples

In this section we shall give the definition of a covering map, and we shall analyze some examples.

5.1.1 Example. Consider the *exponential map* $p : \mathbb{R} \longrightarrow \mathbb{S}^1$, given by $p(t) = \mathrm{e}^{2\pi i t}$. One way to visualize this map is shown in Figure 5.1.

Take the open set $U = \mathbb{S}^1 - \{1\}$. Then $p^{-1}(U) = \mathbb{R} - \mathbb{Z}$. Clearly one has $\mathbb{R} - \mathbb{Z} = \sqcup_{n\in\mathbb{Z}}(n, n+1)$ and for each n, the restriction $p|_{(n,n+1)} : (n, n+1) \longrightarrow U$ is a homeomorphism, whose inverse is a branch of the logarithm. More generally:

(i) If $U \subset \mathbb{S}^1$ is any open set different from $\mathbb{S}^1$, then $p^{-1}(U) = \sqcup_{n\in\mathbb{Z}}\widetilde{U}_n$ and the restriction $p|_{\widetilde{U}_n} : \widetilde{U}_n \longrightarrow U$ is a homeomorphism.

It is precisely this property what was used implicitly to define the degree, namely to show that:

(ii) Given a loop $\lambda : I \longrightarrow \mathbb{S}^1$ based on 1, there is a unique path $\widetilde{\lambda} : I \longrightarrow \mathbb{R}$ such that $\widetilde{\lambda}(0) = 0$ and $p \circ \widetilde{\lambda} = \lambda$.

The first property means that p is a covering map. The second property is the so-called unique path lifting property, which all covering maps have.

5.1.2 Definition. Consider a topological space X. A *covering map* over X is a map $p : \widetilde{X} \longrightarrow X$, $\widetilde{X} \neq \emptyset$ such that each point $x \in X$ has a neighborhood U in X satisfying

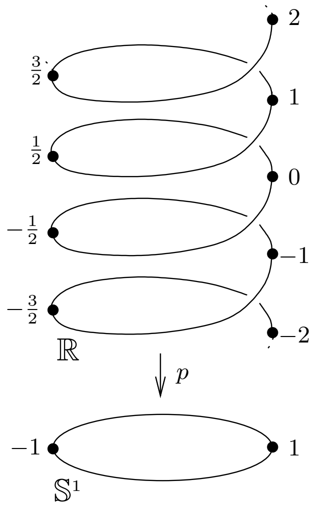


Figure 5.1 The universal covering map of $\mathbb{S}^1$

(a) The inverse image of U, $p^{-1}(U)$, is the disjoint union of open sets $\widetilde{U}_j \subset \widetilde{X}$, $j \in \mathcal{J}$, donde $\mathcal{J}$ is some nonempty set of indexes.

(b) For each $j \in \mathcal{J}$, the restriction $p|_{\widetilde{U}_j} : \widetilde{U}_j \longrightarrow U$ is a homeomorphism.

In particular, by (a), p is surjective. Furthermore, if we assume that X and $\widetilde{X}$ are path-connected, then we shall say that p is a *path-connected covering map*. The space X is called the *base space* of the covering map, the space $\widetilde{X}$ is called the *total space* of the covering map (or *covering space*). For each $x \in X$, the inverse image $p^{-1}(x)$ is called the *fiber* over x of the covering space. The fiber $p^{-1}(x)$ is nonempty, since p is surjective. A neighborhood U which satisfies (a) and (b) is said to be *evenly covered* by p and the sets $\widetilde{U}_j$ are called the *leaves* over U (see Figyre 5.2).

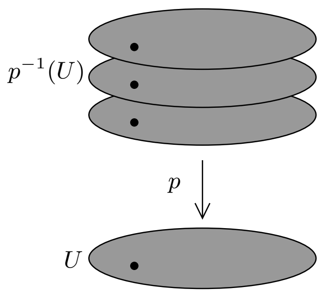


Figure 5.2 A covering map seen locally

We have the following result.

5.1.3 **Proposition**. *Let $p : \widetilde{X} \longrightarrow X$ be a covering map such that X is connected. Take $x, y \in X$. Then the fibers $p^{-1}(x)$ and $p^{-1}(y)$ have the same cardinality.*

Proof: By 5.1.2(a), the set A of all points of X whose fibers have the same cardinality of $p^{-1}(x)$ is open. If $p^{-1}(y)$ had a different cardinality, then the set B of all points whose fibers have a different cardinality from that of $p^{-1}(x)$ would be a nonempty open set complementary to A, thus contradicting the connectedness of X. □

The cardinality of the fibers of a covering map is called the *multiplicity* of the covering map. It might be finite or infinite. If it is finite, say of cardinality n, then we say that we have an *n-fold covering map.*

The next result gathers several properties of a covering map.

5.1.4 **Theorem**. *Let $p : \widetilde{X} \longrightarrow X$ be a covering map. Then the following hold:*

(a) *For each $x \in X$, the fiber $p^{-1}(x)$ is a discrete space.*

(b) *If a connected subspace $\widetilde{A} \subseteq \widetilde{X}$ lies over an evenly covered neighborhood U, namely, it is such that $p(\widetilde{A}) \subset U$, then $\widetilde{A}$ lies inside a leaf, i.e., $\widetilde{A} \subset \widetilde{U}_j$ for some $j \in \mathcal{J}$.*

(c) *If U is an evenly covered neighborhood of x in X and V is another neighborhood of x such that $V \subset U$, then V is evenly covered.*

(d) *The neighborhoods in X which are evenly covered by p form a basis for the topology of X. Furthermore, the leaves which lie over the evenly covered neighborhoods form a basis for the topology of $\widetilde{X}$.*

(e) *p is a local homeomorphism. Furthermore, if X is connected, then p is continuous, surjective and open, so that it is an identification.* □

5.1.5 EXAMPLES. Not every local homeomorphism $f : Y \longrightarrow X$ is a covering map. For instance,

(a) if $Y = \mathbb{R} \sqcup \mathbb{R}/\sim$, where the corresponding negative real numbers of each copy of $\mathbb{R}$ are identified (see 2.1.9), then the natural projection $Y \longrightarrow \mathbb{R}$ is a local homeomorphism, but it is not a covering map, since no neighborhood of 0 is evenly covered by f;

(b) also the map $f : (0,3) \longrightarrow \mathbb{S}^1$ given by $f(t) = \mathrm{e}^{2\pi \mathrm{i} t}$, is a local homeomorphism, which is not a covering map, since no neighborhood of $1 \in \mathbb{S}^1$ is evenly covered by f.

The next result will be useful below.

5.1.6 **Proposition**. *Let X be locally connected, let $p : \widetilde{X} \longrightarrow X$ and $p' : \widetilde{X}' \longrightarrow X$ be covering maps, and let $q : \widetilde{X}' \longrightarrow \widetilde{X}$ be surjective and such that $p \circ q = p'$, namely such that the triangle*

$$\begin{tikzcd} \widetilde{X}' \arrow[rr, "q"] \arrow[dr, "p'"'] & & \widetilde{X} \arrow[dl, "p"] \\ & X & \end{tikzcd}$$

commutes. Thus q is also a covering map.

Proof: Take $\widetilde{x} \in \widetilde{X}$ and let $x = p(\widetilde{x})$. Since p and p' are covering maps, there is a connected $U \in \mathcal{N}_x$ which is evenly covered by both p and p'. Thus, in particular, $p^{-1}(U) = \sqcup_{j \in \mathcal{J}} \widetilde{U}_j$, and $\widetilde{p}|_{\widetilde{U}_j} : \widetilde{U}_j \longrightarrow U$ is a homeomorphism. Let $\widetilde{U} = \widetilde{U}_{j_0}$ be such that $\widetilde{x} \in \widetilde{U}_{j_0}$. Since U is also evenly covered by p', $p'^{-1}(U) = \sqcup_{i \in \mathcal{J}'} \widetilde{U}'_i$. For each $j \in \mathcal{J}$ put $\mathcal{J}'_j = \{i \in \mathcal{J}' \mid q(\widetilde{U}'_i) \subset \widetilde{U}_j$. Then $\mathcal{J}'_j \neq \emptyset$, since each $\widetilde{U}'_i$ is connected and q is surjective. Therefore $q^{-1}(\widetilde{U_j}) = \sqcup_{i \in \mathcal{J}'_j} \widetilde{U}'_i$ and in particular $q^{-1}(\widetilde{U}) = \sqcup_{i \in \mathcal{J}'_{j_0}} \widetilde{U}'_i$. Since for each $i \in \mathcal{J}'_j$, the following diagram

$$\begin{tikzcd} \widetilde{U}'_i \arrow[rr, "q|_{\widetilde{U}'_i}"] \arrow[dr, "p'|_{\widetilde{U}'_i}"'] & & \widetilde{U}_j \arrow[dl, "p|_{\widetilde{U}_j}"] \\ & U\,, & \end{tikzcd}$$

commutes for all j and in particular for j_0, and since $p|_{\widetilde{U}_j}$ as well as $p'|_{\widetilde{U}'_i}$ in the diagram are homeomorphisms, one has that $q|_{\widetilde{U}'_i} : \widetilde{U}'_i \longrightarrow \widetilde{U}_j$ is a homeomorphism. Consequently each $\widetilde{U}_j$ and in particular $\widetilde{U}$, is evenly covered by q. Therefore q is a covering map. $\square$

5.1.7 DEFINITION. Two covering maps $p : \widetilde{X} \longrightarrow X$ and $p' : \widetilde{X}' \longrightarrow X$ over the same space X are said to be *equivalent*, if there exists a homeomorphism $\widetilde{\varphi} : \widetilde{X}' \longrightarrow \widetilde{X}$ such that $p \circ \widetilde{\varphi} = p'$, namely such that the diagram

$$\begin{tikzcd} \widetilde{X}' \arrow[rr, "\widetilde{\varphi}"] \arrow[dr, "p'"'] & & \widetilde{X} \arrow[dl, "p"] \\ & X & \end{tikzcd}$$

commutes. Hence φ is a *fiberwise* homeomorphism, namely, it maps bijectively fibers onto fibers.

5.1.8 Examples.

(a) The homeomorphisms are the one-leaf covering maps (or 1-fold covering maps).

(b) The exponential map defined in Example 5.1.1 $p : \mathbb{R} \longrightarrow \mathbb{S}^1$ is a covering map, whose fiber is (equivalent to) $\mathbb{Z}$.

(c) If F is a discrete space and we define $p : E = F \times X \longrightarrow X$ to be the projection onto X, then p is a covering map. These are the so-called *product* covering maps.

(d) A covering map $p : E \longrightarrow X$ is said to be *trivial* if there is a homeomorphism $\varphi : X \times F \longrightarrow E$ such that $p \circ \varphi = \mathrm{proj}_X$. Such homeomorphism φ is called a *trivialization* of p.

5.1.9 Exercise. Prove that a map $p : E \longrightarrow X$ is a covering map if and only if it has discrete fibers and it is *locally trivial*, namely there is an open cover $\mathcal{U}$ of X such that $p_U = p|_{p^{-1}(U)} : p^{-1}(U) \longrightarrow U$ is a trivial covering map.

There are some constructions that start with covering maps and yield new covering maps.

5.1.10 **Proposition**. *Given covering maps $p_1 : \widetilde{X}_1 \longrightarrow X_1$ and $p_2 : \widetilde{X}_2 \longrightarrow X_2$, the product map $p = p_1 \times p_2 : \widetilde{X}_1 \times \widetilde{X}_2 \longrightarrow X_1 \times X_2$ is a covering map.*

Proof: If $U_1 \subset X_1$ and $U_2 \subset X_2$ are open sets which are evenly covered by p_1 and p_2, respectively, then $U = U_1 \times U_2$ is evenly covered by p. More precisely, if $\widetilde{U}_i$ is a leaf over U_1 and $\widetilde{V}_j$ is a leaf over U_2, then $\widetilde{U}_i \times \widetilde{V}_j$ is a leaf over U. In particular, the fiber of p over the point (x_1, x_2) is $p_1^{-1}(x_1) \times p_2^{-1}(x_2)$. $\square$

The covering map p in the previous proposition is called the *product* of p_1 and p_2.

5.1.11 **Proposition**. *Given covering maps $p_1 : \widetilde{X}_1 \longrightarrow X$ and $p_2 : \widetilde{X}_2 \longrightarrow X$ over the same space, consider $\widetilde{X}_1 \times_X \widetilde{X}_2 = \{(y_1, y_2) \in \widetilde{X}_1 \times \widetilde{X}_2 \mid p_1(y_1) = p_2(y_2)\}$ and $p : \widetilde{X} = \widetilde{X}_1 \times_X \widetilde{X}_2 \longrightarrow X$ defined by $p(y_1, y_2) = p_1(y_1) = p_2(y_2)$. Then p is a covering map.*

Proof: If $U \subset X$ is evenly covered by p_1 and $V \subset X$ is evenly covered by p_2, then $U \cap V$ is evenly covered by p. The fiber of p over $x \in X$ is the product $p_1^{-1}(x) \times p_2^{-1}(x)$. $\square$

The covering map p in the previous proposition is called the *fibered product* of p_1 and p_2 (the space $\widetilde{X}_1 \times_X \widetilde{X}_2$ is also called the *fibered product* of the total spaces).

5.1.12 **Proposition**. *Let $p : \widetilde{X} \longrightarrow X$ be a covering map and let $f : Y \longrightarrow X$ be a continuous map. If $\widetilde{Y} = \{(y, \widetilde{x}) \in Y \times \widetilde{X} \mid f(y) = p(\widetilde{x})\}$ and $q : \widetilde{Y} \longrightarrow Y$ is given by $q(y, \widetilde{x}) = y$, then q is a covering map.*

Proof: If the open set $U \subset X$ is evenly covered by p, then its inverse image $V = f^{-1}(U) \subset Y$ is evenly covered by q. □

The covering map q is called the covering map *induced* by p over f and is usually denoted by $f^*(p) : f^*(\widetilde{X}) \longrightarrow Y$.

5.1.13 Exercise. Show that in the previous construction, the fiber of q over $y \in Y$ is the same as the fiber of p over $f(y)$, namely $q^{-1}(y) = \{y\} \times p^{-1}(f(y))$. In particular, notice that any fiber of p over x is the total space of the covering map induced by p over the inclusion $\{x\} \hookrightarrow X$ and the fibered product $p : \widetilde{X}_1 \times_X \widetilde{X}_2 \longrightarrow X$ is the covering map induced by the product covering map $p_1 \times p_2 : \widetilde{X}_1 \times \widetilde{X}_2 \longrightarrow X \times X$ over the diagonal map $\Delta : X \longrightarrow X \times X$.

5.1.14 Exercise. Let $p : E \longrightarrow X$ and $q : E' \longrightarrow Y$ be covering maps and consider the following commutative diagram:

$$\begin{tikzcd} E' \arrow[r, "\widetilde{f}"] \arrow[d, "q"'] & E \arrow[d, "p"] \\ Y \arrow[r, "f"'] & X \end{tikzcd},$$

where $\widetilde{f}|_{q^{-1}(y)} : q^{-1}(y) \longrightarrow p^{-1}(f(y))$ is a bijection. Show that q is equivalent to the covering map induced by p over f.

5.1.15 Exercise. Check all details of the proofs of the three previous propositions.

5.1.16 Exercise. Let $p : \widetilde{X} \longrightarrow X$ be a covering map and take $Y \subset X$. Show that the restriction $p|_{p^{-1}Y} : p^{-1}Y \longrightarrow Y$ is a covering map, namely the so-called *restriction* of p to Y. Furthermore, show that if $i : Y \hookrightarrow X$ is the inclusion map, then the induced covering map $q = i^*(p) : \widetilde{Y} = i^*(\widetilde{X}) \longrightarrow Y$ is equivalent to the restriction of p to Y.

5.1.17 Exercise. Consider the space $\widetilde{X} = \widetilde{X}_1 \times_X \widetilde{X}_2$ defined in Proposition 5.1.11 and take the (restriction of) the first projection $q : \widetilde{X}_1 \times_X \widetilde{X}_2 \longrightarrow \widetilde{X}_1$. Show that q is a covering map. Indeed, it is the covering map induced by p_2 over p_1.

5.1.18 Exercise. Let $p : \widetilde{X} \longrightarrow X$ be a covering map. Show the following functorial properties of the induced covering maps.

(a) If $f = \mathrm{id}_X$, then $f^*\widetilde{X} \approx \widetilde{X}$, where the homeomorphism is given by the associated map $\widetilde{f} : (x, \widetilde{x}) \mapsto \widetilde{x}$.

(b) If we have maps $f : Y \longrightarrow X$ and $g : Z \longrightarrow Y$, then $(f \circ g)^*(\widetilde{X}) = g^*(f^*\widetilde{X})$.

5.1.19 Exercise. Let $p : \widetilde{X} \longrightarrow X$ and $p' : \widetilde{X}' \longrightarrow X'$ be covering maps. Show that if $\widetilde{f} : \widetilde{X} \longrightarrow widetildeX'$ is a *morphism of covering maps*, that is, there exists a continuous map $f : X \longrightarrow X'$ such that $f \circ p' = p \circ \widetilde{f}$, and for each $x \in X$, the restriction to the fiber $\widetilde{f}_x : p^{-1}(x) \longrightarrow p'^{-1}(f(x))$ is bijective, then $\widetilde{X} \approx f^*\widetilde{X}'$.

Passing to a more geometrical setting, we recall some other examples of covering maps that have been present in the text.

5.1.20 Examples.

(a) Take $n \geq 1$ and consider the n-sphere $\mathbb{S}^n$. Then the quotient space obtained by identifying each pair of antipodal points $x \sim -x$ determines *real projective space* of dimension n, namely $\mathbb{RP}^n = \mathbb{S}^n/\sim$. The quotient map $p : \mathbb{S}^n \longrightarrow \mathbb{RP}^n$ is a 2-fold covering map.

(b) The product of two copies of the covering map $p : \mathbb{R} \longrightarrow \mathbb{S}^1$ of Example 5.1.1 yields a covering map $\mathbb{R}^2 \longrightarrow \mathbb{T}^2$, where $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ is the 2-dimensional torus. The fiber of this covering map over any point is of the form $\mathbb{Z} \times \mathbb{Z}$ and looks like in Figure 5.3.

In the prior example 5.1.20 (a), in the case $n = 1$, the projective space $\mathbb{RP}^1$ is homeomorphic to $\mathbb{S}^1$. In other words, the map $\pi : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ given by $\zeta \mapsto \zeta^2$ is equivalent to $p : \mathbb{S}^1 \longrightarrow \mathbb{RP}^1$. One can see this, by observing that the map π is continuous and surjective from a compact space to a Hausdorff space. Hence it is an identification. Furthermore, it identifies exactly two points if and only if they are antipodal, just as p does. In particular, this shows that there may be nontrivial covering maps, for which the total and the base space are the same space.

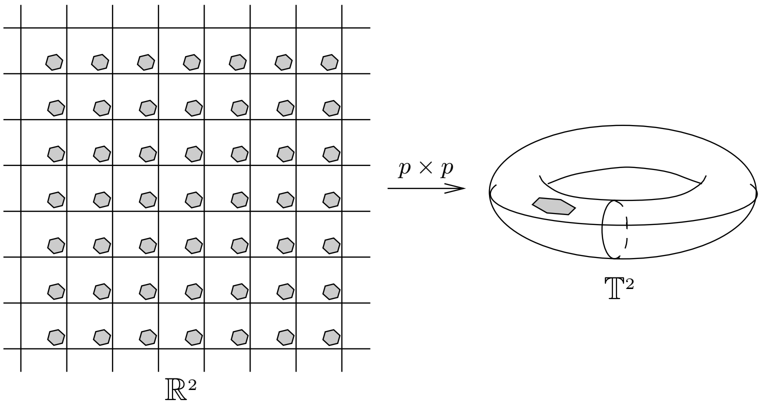


Figure 5.3 The universal covering space of the torus

5.1.21 Exercise. Consider the map $g_k : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ given by $g_k(\zeta) = \zeta^k$. Show that g_k is a k-fold covering map.

5.1.22 Definition. Let G be a (discrete) group. A *(left) action* of G on a space X is a continuous map

$$\mu : G \times X \longrightarrow X\,,$$

where we write gx instead of $\mu(g, x)$, which satisfies

$$1x = x$$
$$g_1(g_2x) = (g_1g_2)x\,.$$

Hence each element $g \in G$ determines a homeomorphism $X \longrightarrow X$, given by $x \mapsto gx$ (and inverse given by $x \mapsto g^{-1}x$) and the equations above mean that the function $G \longrightarrow \mathbf{Homeo}(X)$, determined by the action, where $\mathbf{Homeo}(X)$ is the (topological) group of homeomorphisms of X onto itself, is a group homomorphism.

We say that a group action is *even** if every point $x \in X$ has a neighborhood V such that $V \cap gV = \emptyset$ if and only if $1 \neq g \in G$, where $gV = \{gx \mid x \in V\}$. Hence, for $g_1 \neq g_2 \in G$, one has $g_1V \cap g_2V = \emptyset$. Therefore, if G acts evenly on X and $1 \neq g \in G$, then for every $x \in X$, $x \neq gx$. This means that the action is *free*. Given $x \in X$, the set $Gx = \{gx \mid g \in G\}$ is called the *orbit* of x under the action of G on X. Thus, if one has a free action of G on X, then for each $x \in X$, the mapping $G \longrightarrow X$ given by $g \mapsto gx$ is an embedding. Hence, in this case, each orbit is homeomorphic to the group. Otherwise, this mapping has a "kernel", namely, there is a subgroup G_x such that $gx = x$ if and only if $g \in G_x$. In this case, the the mapping $G \longrightarrow X$ given by $g \mapsto gx$ defines a mapping from the quotient (which is not necessarily a group) $G/G_x \longrightarrow X$ which is an embedding.

*Most authors say that the action is *properly discontinuous*, but we find this designation contradictory, since the action is continuous.

The *orbit space* or *quotient space* of the action of G on X is the quotient space $X/G = X/{\sim}$, where $x_1 \sim x_2$ if and only if $x_2 = gx_1$ for some $g \in G$. The quotient map $q : X \longrightarrow X/G$ is called the *orbit map*.

5.1.23 EXAMPLE. The (additive) group $\mathbb{R}$ acts on the topological space $\mathbb{R}$ by $(g, x) \mapsto g + x$. This action is free, but it is not even (since $\mathbb{R}$ is not discrete). However, the restriction of this action to the subgroup $\mathbb{Z}$ of $\mathbb{R}$ is even, and thus it is free.

5.1.24 EXERCISE. Show that if G is a finite (discrete) group that acts freely on a Hausdorff space X, then the action must be even. (This is the case, for instance, for the *antipodal action* of $\mathbb{Z}_2$ on any sphere $\mathbb{S}^n$, given by $1x = x$ and $(-1)x = -x$ for any $x \in \mathbb{S}^n$.)

Even actions of groups on topological spaces are a source of covering maps. We have the following result.

5.1.25 **Theorem**. *If a group G acts evenly on X, then the orbit map $q : X \longrightarrow X/G$ is a covering map, whose multiplicity is the cardinality of the group G.*

Proof: Take $x \in X$ and a neighborhood V of x such that $g_1V \cap g_2V = \emptyset$ if $g_1 \neq g_2 \in G$. Then $U = q(V) \subset X/G$ is a neighborhood of $q(x)$ which is evenly covered by q. Namely, $q^{-1}(U) = \sqcup_{g \in G} gV \subset X$ and hence each fiber is equivalent as a set to G, since the mapping $G \longrightarrow q^{-1}(q(x)) = \{gx \mid g \in G\}$ given by $g \mapsto gx$ is clearly bijective. Thus the multiplicity of q is the cardinality of G. $\square$

5.1.26 EXAMPLES.

(a) According to 5.1.24, the antipodal action $\mathbb{Z}_2 \times \mathbb{S}^n \longrightarrow \mathbb{S}^n$ is even. Hence the orbit map $\mathbb{S}^n \longrightarrow \mathbb{S}^n/\mathbb{Z}_2$ is a covering map. Since $\mathbb{S}^n/Z_2 = \mathbb{RP}^n$, this covering map is the one mentioned in 5.1.20 (a).

(b) The cyclic group of order k, $\mathbb{Z}_k$ (seen as the group of the kth roots of unity), acts on $\mathbb{S}^1 \subset \mathbb{C}$ as follows. If $a \in \mathbb{Z}_k$ is the canonical generator, namely $a = \mathrm{e}^{2\pi\mathrm{i}/k}$ (the primitive kth root of 1), then $a\zeta = \mathrm{e}^{2\pi\mathrm{i}/k}\zeta$ is given by the product of complex numbers. In other words, one can see $\mathbb{Z}_k$ as a subgroup of $\mathbb{S}^1$ and the action is given by the multiplication in the group $\mathbb{S}^1$. This action is clearly even, and so the orbit map $q : \mathbb{S}^1 \longrightarrow \mathbb{S}^1/\mathbb{Z}_k$ is a covering map. On the other hand, the map $p : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ given by $p(\zeta) = \zeta^k$, is an identification for which $p(\zeta) = p(\xi)$ if and only if $\zeta = b\xi$, where b is a kth root of 1. Hence there is a homeomorphism $\mathbb{S}^1 \longrightarrow \mathbb{S}^1/\mathbb{Z}_k$, up to which p corresponds to q. We have thus shown that *the mapping of degree k, $g_k : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ given by $g_k(\zeta) = \zeta^k$, is a covering map of multiplicity k.* (Cf. comment before 5.1.21.)

(c) Generalizing 5.1.20 (b), we have that the free abelian group in n generators $\mathbb{Z}^n$ acts evenly on $\mathbb{R}^n$ via $(g,x) \mapsto g+x$. Therefore $p : \mathbb{R}^n \longrightarrow \mathbb{R}^n/\mathbb{Z}^n$ is a covering map with countably many leaves, which up to homeomorphism, coincides with the product covering map $\mathbb{R}^n \longrightarrow \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ (n factors), obtained as the product of n copies of the covering map of example 5.1.1. In other words, p maps $(x_1,\dots,x_n)$ to $(\mathrm{e}^{2\pi \mathrm{i} x_1},\dots,\mathrm{e}^{2\pi \mathrm{i} x_n})$. Hence one has the mapping $p(x_1,\dots,x_n) \mapsto (\mathrm{e}^{2\pi \mathrm{i} x_1},\dots,\mathrm{e}^{2\pi \mathrm{i} x_n})$ is a homeomorphism $\mathbb{R}^n/\mathbb{Z}^n \longrightarrow \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ (n factores). The space $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ is called the *n-torus* or n-dimensional torus.

(d) There is an even action of $\mathbb{Z}$ on $\mathbb{R}^2$ given by $(n,(x_1,x_2)) \mapsto (n+x_1,(-1)^n x_2)$. In the associated covering map $p : \mathbb{R}^2 \longrightarrow \mathbb{R}^2/\mathbb{Z}$, the base space is an open Moebius strip, namely without its boundary (see 1.4.11).

(e) Let G be a subgroup of the group of rigid transformations of $\mathbb{R}^n$ (i.e. rotations, translations, and reflections), whose natural action on $\mathbb{R}^n$ is even. Then the orbit map $p : \mathbb{R}^n \longrightarrow \mathbb{R}^n/G$ is a covering map. The base space $\mathbb{R}^n/G$ is called *Euclidean space form.* It is a smooth n-manifold which inherits from $\mathbb{R}^n$ a natural Euclidean geometric structure. Examples (c) and (d) have this form. The group G is called *crystalogrphic group* of $\mathbb{R}^n$.

5.1.27 EXERCISE. Let G be the subgroup of the group of rigid transformations of $\mathbb{R}^2$ generated by the transformations

$$(x_1,x_2) \longmapsto (x_1+1,x_2) \quad \text{and} \quad (x_1,x_2) \longmapsto (-x_1,x_2+1)\,.$$

Show that the action is even and verify that the orbit space $\mathbb{R}^2/G$ is homeomorphic to the Klein bottle.

5.1.28 EXERCISE. Considering the $(2n-1)$-sphere $\mathbb{S}^{2n-1}$ as

$$\mathbb{S}^{2n-1} = \{z = (z_1,\dots,z_n) \in \mathbb{C}^n \mid |z_1|^2 + \cdots + |z_n|^2\}$$

and $\mathbb{Z}/k$ as the multiplicative group of the kth roots of unity in $\mathbb{S}^1$, there is an action given by $(\zeta,z) \mapsto (\zeta z_1,\dots,\zeta z_n)$. Show that this action is even and thus one has a covering map $p : \mathbb{S}^{2n-1} \longrightarrow \mathbb{S}^{2n-1}/\mathbb{Z}_k$. If k is prime, the space $\mathbb{S}^{2n-1}/\mathbb{Z}_k$ is called *lens space*, which we denote by L_k^{2n-1}, where $2n-1$ is the dimension of this manifold. The case $k=2$ corresponds to the real projective spaces of odd dimension, i.e. $L_2^{2n-1} = \mathbb{RP}^{2n-1}$.

5.1.29 EXERCISE. Show that the mapping $\mathbb{T}^2 \longrightarrow \mathbb{T}^2$ given by

$$(\zeta,\xi) \mapsto (\zeta^a \xi^b, \zeta^c \xi^d)\,,$$

with $a,b,c,d \in \mathbb{Z}$, if $m = ad - bc \neq 0$, is a $|m|$-fold covering map.

5.1.30 Exercise. Construct a 2-fold covering map onto the Klein bottle $\mathbb{T}^2 \longrightarrow K$.

5.1.31 Exercise. Recall that the *quaternions* are the elements x of $\mathbb{R}^4$ written as $x = x_0 + \mathrm{i}x_1 + \mathrm{j}x_2 + \mathrm{k}x_3$, where x_0 is short for $(x_0, 0, 0, 0)$ and is called the *real part* of x, and $\mathrm{i} = (0, 1, 0, 0)$, $\mathrm{j} = (0, 0, 1, 0)$, and $\mathrm{k} = (0, 0, 0, 1)$, which are the generators of the *imaginary part* of the quaternions. The set $\mathbb{H}$ of quaternions has a multiplicative structure, generated by

$$\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1 \quad \text{and}$$

$$\mathrm{ij} = \mathrm{k} = -\mathrm{ji}, \mathrm{jk} = \mathrm{i} = -\mathrm{kj}, \mathrm{ki} = \mathrm{j} = -\mathrm{ik}\,.$$

Consider $\mathbb{S}^3 \subset \mathbb{R}^4 = \mathbb{H}$ as the set of quaternions of norm 1. Taking $\mathbb{R}^3$ as the subset of $\mathbb{H}$ consisting of all quaternions of real part 0, show:

(a) If $x \in \mathbb{S}^3$, then the mapping $f_x : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ given by $f_x(y) = xyx^{-1}$, is an orthogonal transformation with determinant $+1$. Hence it restricts to a mapping $f : \mathbb{S}^3 \longrightarrow \mathrm{SO}_3$, such that $x \mapsto f_x$, where SO_3 denotes the group of orthogonal transformations of $\mathbb{R}^3$ with determinant $+1$).

(b) The map $f : \mathbb{S}^3 \longrightarrow \mathrm{SO}_3$ is continuous and surjective (hence an identification) and $f_x = f_{x'}$ if and only if $x = \pm x'$.

(c) $f : \mathbb{S}^3 \longrightarrow \mathrm{SO}_3$ is a 2-fold covering map.

(d) f induces a homeomorphism $\mathbb{RP}^3 \longrightarrow \mathrm{SO}_3$.

5.2 Lifting properties

The fundamental property of the covering maps is the "lifting property", which we shall analyze in this section.

5.2.1 Definition. Let $p : \widetilde{X} \longrightarrow X$ be a covering map and let $f : Y \longrightarrow X$ be continuous. A map $\widetilde{f} : Y \longrightarrow \widetilde{X}$ is said to *lift* f if $p \circ \widetilde{f} = f$. In a diagram:

$$\begin{tikzcd} & \widetilde{X} \arrow[d, "p"] \\ Y \arrow[ur, dashed, "\widetilde{f}"] \arrow[r, "f"'] & X \end{tikzcd}$$

Of special importance in the theory of covering maps is the *path lifting* property, namely the following: Given a path $\omega : I \longrightarrow X$, there is a path *lifting* ω, namely a path $\widetilde{\omega} : I \longrightarrow \widetilde{X}$ such that $p \circ \widetilde{\omega} = \omega$.

In a figure it would look as shown in Figure 5.4, where several paths lifting the same path are shown.

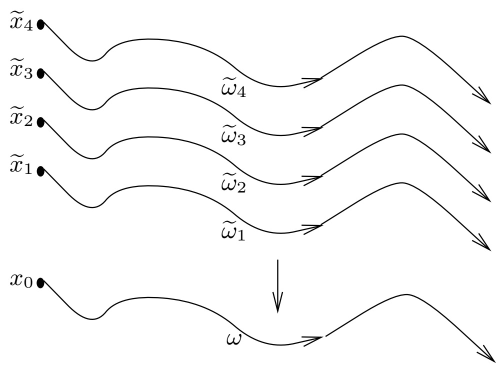


Figure 5.4 In a covering map there are several paths lifting a given path in the base space

A typical problem of the theory is the *lifting problem* which asks about the solution of the following problem:

Given $p : \widetilde{X} \longrightarrow X$ a pointed covering map, i.e. such that $p(\widetilde{x}_0) = x_0$ for some $\widetilde{x}_0 \in \widetilde{X}$ and $x_0 \in X$, and given a map $f : Y \longrightarrow X$ and a point $y_0 \in Y$ such that $f(y_0) = x_0$, does there exist a map $\widetilde{f} : Y \longrightarrow \widetilde{X}$ lifting f, such that $\widetilde{f}(y_0) = \widetilde{x}_0$.

In a sequence of steps, we shall analyze the solutions to the lifting problem. In any case, there is only one solution if it exists, in the sense of the following result.

5.2.2 **Proposition**. *Let $p : \widetilde{X} \longrightarrow X$ be a covering map. If Y is a connected space and $\widetilde{f}, \widetilde{g} : Y \longrightarrow \widetilde{X}$ are continuous maps such that $p \circ \widetilde{f} = p \circ \widetilde{g}$, then $\widetilde{f} = \widetilde{g}$ if and only if there is a point $y \in Y$ for which $\widetilde{f}(y) = \widetilde{g}(y)$.*

Proof: Take $y \in Y$ and let U be a neighborhood of $p\widetilde{f}(y) = p\widetilde{g}(y)$ which is evenly covered by p. If $\widetilde{U}_1$ is the neighborhood of $\widetilde{f}(y)$ and $\widetilde{U}_2$ is the neighborhood $\widetilde{g}(y)$, and they are such that $p : \widetilde{U}_1 \approx U$ and $p : \widetilde{U}_2 \approx U$, then $V = \widetilde{f}^{-1}(\widetilde{U}_1) \cap \widetilde{g}^{-1}(\widetilde{U}_2)$ is a neighborhood of y in Y. If $\widetilde{f}(y) = \widetilde{g}(y)$, then $\widetilde{U}_1 = \widetilde{U}_2$ and $\widetilde{f}(y') = \widetilde{g}(y')$ for every point $y' \in V$. Thus the set on which the two maps $\widetilde{f}$ and $\widetilde{g}$ coincide is open. If on the other hand $\widetilde{f}(y) \neq \widetilde{g}(y)$, then $\widetilde{U}_1 \cap \widetilde{U}_2 = \emptyset$ and hence $\widetilde{f}(y') \neq \widetilde{g}(y')$ for every point $y' \in V$. Hence the set on which $\widetilde{f}$ and $\widetilde{g}$ are different is also open. By the connectedness of Y and the fact that there is a point at which $\widetilde{f}$ and $\widetilde{g}$ coincide, the second open set must be empty. Therefore both maps coincide at every point of Y. □

We shall now start with the analysis of the paths $\widetilde{\omega} : I \longrightarrow \widetilde{X}$ lifting $\omega : I \longrightarrow X$. The fundamental result of the theory of covering maps is the following.

5.2.3 **Theorem**. (unique path-lifting) *Consider a covering map $p : \widetilde{X} \longrightarrow X$. For each path $\omega : I \longrightarrow X$ and for each point $\widetilde{x}$ such that $p(\widetilde{x}) = \omega(0)$, there is a unique path $\widetilde{\omega} : I \longrightarrow \widetilde{X}$ lifting ω, such that $\widetilde{\omega}(0) = \widetilde{x}$. We denote this lifting path by $L(\omega, \widetilde{x})$. Furthermore, if ω_0 and ω_1 are paths in X such that $\omega_0 \simeq \omega_1 \operatorname{rel} \partial I$ and $\widetilde{x}$ is a point such that $p(\widetilde{x}) = \omega_0(0) = \omega_1(0)$, then $L(\omega_0, \widetilde{x}) \simeq L(\omega_1, \widetilde{x}) \operatorname{rel} \partial I$ in $\widetilde{X}$. In particular, the ends of both lifting paths coincide, namely $L(\omega_0, \widetilde{x})(1) = L(\omega_1, \widetilde{x})(1)$.*

Before passing to the proof of this theorem, consider the following assertion, which will be necessary for the last part of the proof.

5.2.4 **Lemma**. *Let $p : \widetilde{X} \longrightarrow X$ be a covering map. Then for each continuous map $H : I^2 \longrightarrow X$ and for each point $\widetilde{x}$ on the fiber over $H(0,0)$, there is a unique map $\widetilde{H} : I^2 \longrightarrow \widetilde{X}$, such that $p \circ \widetilde{H} = H$ and $\widetilde{H}(0,0) = \widetilde{x}$.*

Proof: The uniqueness follows immediately from 5.2.2. We show the existence. Since I is compact, there is a sufficiently fine partition $0 = t_0 < t_1 < \cdots < t_n = 1$ of I so that each square $Q_{ij} = [t_{i-1}, t_i] \times [t_{j-1}, t_j]$ is mapped by H into a neighborhood U_{ij} which is evenly covered by p (this follows by taking the partition so fine that the diameter of each square is smaller than the Lebesgue number of the cover of $I \times I$ defined by $\{f^{-1}U \mid U \subseteq X \text{ is evenly covered by } p\}$. For a leaf which is not yet determined, $\widetilde{U}_{ij}$ over U_{ij}, let $p_{ij} : \widetilde{U}_{ij} \longrightarrow U_{ij}$ be the homeomorphism given by $p_{ij} = p|_{\widetilde{U}_{ij}}$, and let $\widetilde{H}_{ij} : Q_{ij} \longrightarrow \widetilde{X}$ be given by $\widetilde{H}(s,t) = p_{ij}^{-1} H(s,t)$. We shall see that it is possible to choose the leaves $\widetilde{U}_{ij}$ in such a way that the partial mappings $\widetilde{H}_{ij}$ define a continuous map $\widetilde{H}$ as desired.

Let first $\widetilde{U}_{11}$ be the leaf that contains $\widetilde{x}$. A consequence of this will be that $\widetilde{H}(0,0) = \widetilde{x}$. Since the set $\widetilde{H}_{11}(Q_{11} \cap Q_{21})$ lies over U_{21}, there is a (unique) leave $\widetilde{U}_{21}$ which contains it. If we choose it, then $\widetilde{H}_{11}$ and $\widetilde{H}_{21}$ will coincide in $Q_{11} \cap Q_{21}$. Analogously, leaves $U_{31}, \ldots, U_{n1}$ are successively chosen and with them the desired map $\widetilde{H}$ on the row $Q_{11} \cup \cdots \cup Q_{n1}$ is obtained. To do the corresponding on the second row, one first chooses $\widetilde{U}_{12}$ to be the leaf that contains $\widetilde{H}_{11}(Q_{12} \cap Q_{22})$. In the next step, we have two choices: either to take the leaf $\widetilde{U}_{22}$ which contains $\widetilde{H}_{12}(Q_{12} \cap Q_{22})$, or the leaf that contains $\widetilde{H}_{21}(Q_{21} \cap Q_{22})$. However there is no ambiguity, since either leaf must contain the point $\widetilde{H}_{12}(t_1, t_1) = \widetilde{H}_{21}(t_1, t_1)$, and thus both must be the same. This way, after choosing $\widetilde{U}_{22}$, one has that $\widetilde{H}_{12} = \widetilde{H}_{22}$

and $\widetilde{H}_{2\,1} = \widetilde{H}_{2\,2}$ in the intersection of their domains. Successively, one constructs $\widetilde{H}$ over the second row, and similarly on the following rows. □

We may now use this lemma to prove Theorem 5.2.3.

Proof of 5.2.3*:* Given the path $\omega : I \longrightarrow X$, take the homotopy $H : I \times I \longrightarrow X$ given by $H(s,t) = \omega(s)$. Then $H(0,0) = \omega(0)$ and if $p(\widetilde{x}) = \omega(0)$, then by the previous lemma, there is a homotopy $\widetilde{H} : I \times I \longrightarrow \widetilde{X}$ such that $\widetilde{H}(0,0) = \widetilde{x}$ and $p \circ \widetilde{H} = H$. Therefore the path $\widetilde{\omega} : I \longrightarrow \widetilde{X}$ given by $\widetilde{\omega}(s) = \widetilde{H}(s,0)$, is a path such that $\widetilde{\omega}(0) = \widetilde{x}$ and $p \circ \widetilde{\omega} = \omega$. That is, $\widetilde{\omega}$ lifts ω starting at $\widetilde{x}$ and by the uniqueness, we know that it is unique.

Finally if $\omega_0, \omega_1 : I \longrightarrow X$ are paths such that $\omega_0 \simeq \omega_1 \text{ rel } \partial I$, and $H : I^2 \longrightarrow X$ is a corresponding homotopy, then by 5.2.4, we know that there is a homotopy $\widetilde{H} : I^2 \longrightarrow \widetilde{X}$ such that $p \circ \widetilde{H} = H$ and $\widetilde{H}(0,0) = \widetilde{x}$. But the path $t \mapsto \widetilde{H}(0,t)$ lies over the fiber of $\omega_0(0) = \omega_1(0)$. Since this fiber is discrete, the path must be constant. Analogously, the path $t \mapsto \widetilde{H}(1,t)$ must be constant. Hence $\widetilde{H}(0,t) = \widetilde{x}$, $t \in I$, and $\widetilde{H}(1,t) = \widetilde{y}$ for some fixed point $\widetilde{y} \in \widetilde{X}$ on the fiber over $\omega_0(1) = \omega_1(1)$ and for all $t \in I$. On the other hand, the path $s \mapsto \widetilde{H}(s,0)$ lifts ω_0 starting at $\widetilde{x}$. Hence, by the uniqueness of the lifting paths, which follows from 5.2.2, $\widetilde{H}(s,0) = \widetilde{\omega}_0(s)$. Analogously $\widetilde{H}(s,1) = \widetilde{\omega}_1(s)$. This means that the lifting $\widetilde{H}$ of H is a homotopy $\widetilde{\omega}_0 \simeq \widetilde{\omega}_1 \text{ rel } \partial I$, as desired. □

Given a covering map $p : \widetilde{X} \longrightarrow X$, the unique path-lifting theorem yields to each path $\omega : I \longrightarrow X$ and each point $\widetilde{x}$ on the fiber over $\omega(0)$ a lifting $L(\omega, \widetilde{x})$, namely it defines a function

$$L : X^I \times_X \widetilde{X} = \{(\omega, \widetilde{x}) \in X^I \times \widetilde{X} \mid \omega(0) = p(\widetilde{x})\} \longrightarrow \widetilde{X}^I\,,$$

which we call *lifting-function* of the covering map p.

5.2.5 EXERCISE. Given a covering map $p : \widetilde{X} \longrightarrow X$, show that its lifting-function $L : X^I \times_X \widetilde{X} \longrightarrow \widetilde{X}^I$ is continuous if X^I and $\widetilde{X}^I$ are endowed with the compact-open topology, and the domain of L has the relative topology induced by the product topology.

The following corollary of the unique path-lifting theorem summarizes the fundamental properties of the path $L(\omega, \widetilde{x})$, which are convenient to have at hand in the applications of the theorem. The proof consists of immediate applications of the existence and uniqueness of the lifting paths and is left as an *exercise*.

5.2.6 **Corollary**. *Let $p : \widetilde{X} \longrightarrow X$ be a covering map and let $L : X^I \times_X \widetilde{X} \longrightarrow \widetilde{X}^I$ be its path lifting function, then the following hold.*

(a) *The path $L(\omega, \widetilde{x})$ is uniquely determined by the conditions $p \circ L(\omega, \widetilde{x}) = \omega$ and $L(\omega, \widetilde{x})(0) = \widetilde{x}$.*

(b) *If λ is a nullhomotopic loop, then $L(\lambda, \widetilde{x})$ is a nullhomotopic loop.*

(c) *The set $\{L(\omega, \widetilde{x}) \mid \widetilde{x} \in p^{-1}(\omega(0))\}$ consists of all lifting paths of ω. Therefore they are as many as the cardinality of the fiber $p^{-1}(\omega(0))$ and hence as the multiplicity of the covering map.*

(d) *If ω and σ are connectable paths in X, then*

$$L(\omega\sigma, \widetilde{x}) = L(\omega, \widetilde{x})L(\sigma, \widetilde{y}) \, ,$$

where $\widetilde{y}$ is the destination of the lifting path $L(\omega, \widetilde{x})$, which lies in the fiber of the origin $\sigma(0)$ of σ. Furthermore, $\overline{L(\omega, \widetilde{x})} = L(\overline{\omega}, \widetilde{z})$, where $\widetilde{z}$ is the destination of $L(\omega, \widetilde{x})$. □

5.2.7 Exercise. Let $p : \widetilde{X} \longrightarrow X$ be a covering map and let $\widetilde{x}_0 \in \widetilde{X}$ and $x_0 \in X$ be base points such that $p(\widetilde{x}_0) = x_0$. Let Y be a simply connected and locally path connected space with base point y_0. Show that every pointed continuous map $f : Y \longrightarrow X$ has a unique *lifting map* $\widetilde{f} : Y \longrightarrow \widetilde{X}$, such that $\widetilde{f}(y_0) = \widetilde{x}_0$ and $p \circ \widetilde{f} = f$. (*Hint:* For each $y \in Y$ take $\sigma : y_0 \simeq y$ and define $\widetilde{f}(y) = L(f \circ \sigma, \widetilde{x}_0)(1)$.)

5.2.8 Note. The statement of the previous exercise is not true in general if the space Y is not simply connected, namely, if one takes the covering map $p : \mathbb{R} \longrightarrow \mathbb{S}^1$ of 5.1.1, one may take $Y = \mathbb{S}^1$ and $f = \mathrm{id}_{\mathbb{S}^1}$, then there is no map lifting p, since otherwise, if $s : \mathbb{S}^1 \longrightarrow \mathbb{R}$ were such that $p \circ s = \mathrm{id}_{\mathbb{S}^1}$, then we would get a contradiction. To see this, one would have the following commutative diagram of spaces

$$\begin{tikzcd}
& \mathbb{R} \arrow[dr, "p"] & \\
\mathbb{S}^1 \arrow[ur, "s"] \arrow[rr, "\mathrm{id}_{\mathbb{S}^1}"'] & & \mathbb{S}^1 \, ,
\end{tikzcd}$$

which would induce the following commutative diagram of fundamental groups

$$\begin{tikzcd}
& \pi_1(\mathbb{R}, 0) \arrow[dr, "p_*"] & \\
\pi_1(\mathbb{S}^1, 1) \arrow[ur, "s_*"] \arrow[rr, "1_{\pi_1(\mathbb{S}^1,1)}"'] & & \pi_1(\mathbb{S}^1, 1) \, .
\end{tikzcd}$$

Since $\pi_1(\mathbb{S}^1) \cong \mathbb{Z}$ and $\pi_1(\mathbb{R}) = 1$, this would mean that the identity homomorphism of $\mathbb{Z}$ factors through the trivial group, and this is impossible.

The following exercises are an application of Exercise 5.2.7.

5.2.9 Exercise. Let $p : \widetilde{X} \longrightarrow X$ be a covering map, such that its total space $\widetilde{X}$ is contractible. Show that if Y is simply connected and locally path connected, then every continuous map $f : Y \longrightarrow X$ is nullhomotopic.

5.2.10 Exercise. Show that if $n > 1$, then one has

$$[\mathbb{S}^n, \mathbb{S}^1] = 0\,, \quad [\mathbb{S}^n, \mathbb{S}^1 \times \mathbb{S}^1] = 0 \quad \text{y} \quad [\mathbb{RP}^n, \mathbb{S}^1] = 0\,,$$

where $[X, Y]$ stands for the set of homotopy classes of the maps $X \longrightarrow Y$.

As we saw in Remark 5.2.8, given a covering map, it is not always possible to find a map lifting a given map into the base space to the total space. What is the most general condition so that given a covering map $p : \widetilde{X} \longrightarrow X$ and a map $f : Y \longrightarrow X$, there exists a map *lifting* f, namely a map $\widetilde{f} : Y \longrightarrow \widetilde{X}$ such that $p \circ \widetilde{f} = f$?

If such a lifting map exists, then the following diagram commutes:

$$\begin{tikzcd}
 & \pi_1(\widetilde{X}, \widetilde{x}_0) \arrow[d, "p_*"] \\
\pi_1(Y, y_0) \arrow[ur, "\widetilde{f}_*"] \arrow[r, "f_*"'] & \pi_1(X, x_0)\,.
\end{tikzcd}$$

This implies in particular that $f_*(\pi_1(Y, y_0)) \subset p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$, namely, this is a necessary condition for the existence of the lifting map. We shall see that it is also sufficient.

5.2.11 **Theorem**. (Map lifting) *Let $p : \widetilde{X} \longrightarrow X$ be a covering map and let $\widetilde{x}_0 \in \widetilde{X}$ and $x_0 \in X$ be base points such that $p(\widetilde{x}_0) = x_0$. Let Y be a connected and locally path connected space with base point y_0 and let $f : Y \longrightarrow X$ be continuous such that $f(y_0) = x_0$, then the following are equivalent:*

(a) *There is a map $\widetilde{f} : Y \longrightarrow \widetilde{X}$ such that $\widetilde{f}(y_0) = \widetilde{x}_0$ and $p \circ \widetilde{f} = f$.*

(b) $f_*(\pi_1(Y, y_0)) \subset p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$.

If (a) *and thus also* (b) *hold, then the map $\widetilde{f}$ is unique.*

Proof: We already saw that (a) $\Longrightarrow$ (b). If we now assume (b), we shall explain how to construct $\widetilde{f}$. Take $y \in Y$ and let $\sigma : y_0 \simeq y$ be a path. Define $\widetilde{f}(y) = L(f \circ \sigma, \widetilde{x}_0)(1)$. We shall prove that this definition does not depend on the choice of the path and that it determines a continuous map. Since $p(L(f\circ\sigma,\widetilde{x}_0)(1)) = f(\sigma(1)) = f(y)$, $\widetilde{f}$ should be the desired lifting map.

To show that $\widetilde{f}(y)$ does not depend on the path σ, assume that $\gamma : y_0 \simeq y$ is another path. Then the loop $\sigma\overline{\gamma}$ represents an element of the fundamental group $\pi_1(Y,y_0)$. By (b), there is a loop $\widetilde{\lambda}$ in $\widetilde{X}$ based at $\widetilde{x}_0$, such that $f_*([\sigma\overline{\gamma}]) = p_*([\widetilde{\lambda}])$. Hence $f\circ\sigma \simeq (p\circ\widetilde{\lambda})(f\circ\gamma)$ rel ∂I in X. By the lifting theorem 5.2.3 and its corollary 5.2.6(d), one has $L(f\circ\sigma,\widetilde{x}_0) \simeq L((p\circ\widetilde{\lambda})(f\circ\gamma),\widetilde{x}_0) = \widetilde{\lambda}L(f\circ\gamma,\widetilde{x}_0)$ rel ∂I. Hence, in particular, these paths have the same end points, namely

$$L(f\circ\sigma,\widetilde{x}_0)(1) = (\widetilde{\lambda}L(f\circ\gamma,\widetilde{x}_0))(1) = L(f\circ\gamma,\widetilde{x}_0))(1)\,,$$

and this shows that $\widetilde{f}(y)$ is well defined and, in fact, it shows also that $\widetilde{f}(y_0) = \widetilde{x}_0$, since the constant path lifts to the constant path.

Now we have to verify that the defined function $\widetilde{f} : Y \longrightarrow \widetilde{X}$ is continuous. To do it, we shall make use of the local-connectedness assumption. Let us take $y \in Y$ and let $\widetilde{U}$ be the leaf over some evenly-covered neighborhood of $f(y)$, on which $\widetilde{f}(y)$ lies. Since f is continuous, $f^{-1}U$ is a neighborhood of y in Y. Let $V \subset f^{-1}U$ be a path-connected neighborhood of y, take $y' \in V$ and let $\gamma : y_0 \simeq y$ be a path in Y and let $\mu : y \simeq y'$ be a path in V. Then $\widetilde{f}(y') = L(f\circ(\sigma\mu),\widetilde{x})(1) = L(f\circ\mu,\widetilde{f}(y))(1)$. But since $\widetilde{U}$ is the leaf on which $\widetilde{f}(y)$ lies, and given that $f\circ\mu$ is a path in U, the path $L(f\circ\mu,\widetilde{f}(y))$ and in particular its destination $\widetilde{f}(y')$ lie on $\widetilde{U}$. We have thus shown that $\widetilde{f}(V) \subset \widetilde{U}$ and since any neighborhood of $\widetilde{f}(y)$ contains one like $\widetilde{U}$, this shows the continuity of $\widetilde{f}$.

The uniqueness of $\widetilde{f}$ is a consequence of 5.2.2. □

Let $p : \widetilde{X} \longrightarrow X$ be a covering map and let $\widetilde{x}_0 \in \widetilde{X}$ and $x_0 \in X$ be base points such that $p(\widetilde{x}_0) = x_0$. If $\widetilde{\lambda}_0, \widetilde{\lambda}_1 : I \longrightarrow \widetilde{X}$ are loops based at $\widetilde{x}_0$ such that $p_*([\widetilde{\lambda}_0]) = [p\circ\widetilde{\lambda}_0] = [p\circ\widetilde{\lambda}_1] = p_*([\widetilde{\lambda}_1])$. Then $p\circ\widetilde{\lambda}_0 \simeq p\circ\widetilde{\lambda}_1$ rel ∂I. From the lifting lemma 5.2.3 one obtains that $\widetilde{\lambda}_0 \simeq \widetilde{\lambda}_1$ rel ∂I, that is $[\widetilde{\lambda}_1] = [\widetilde{\lambda}_2]$. We have shown part of the following.

5.2.12 **Theorem** and Definition. *Let $p : \widetilde{X} \longrightarrow X$ be a covering map and let $\widetilde{x}_0 \in \widetilde{X}$ and $x_0 \in X$ be base points such that $p(\widetilde{x}_0) = x_0$. Then the homomorphism induced by p between the fundamental groups*

$$p_* : \pi_1(\widetilde{X},\widetilde{x}_0) \longrightarrow \pi_1(X,x_0)$$

is a monomorphism. The image $p_*(\pi_1(\widetilde{X},\widetilde{x}_0)) \subset \pi_1(X,x_0)$ consists precisely of the classes $[\lambda] \in \pi_1(X,x_0)$ such that the lifting path $L(\lambda,\widetilde{x}_0)$ is a loop. This subgroup is called *characteristic subgroup of the covering map $p : \widetilde{X} \longrightarrow X$.*

Proof: It is enough to prove the second statement. If $[\lambda] \in p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$, then $[\lambda] = [p \circ \widetilde{\lambda}]$ for some loop $\widetilde{\lambda}$ in $\widetilde{X}$ with base point $\widetilde{x}_0$. Therefore $L(\lambda, \widetilde{x}_0) \simeq L(p \circ \widetilde{\lambda}, \widetilde{x}_0) = \widetilde{\lambda}$ rel ∂I, namely $L(\lambda, \widetilde{x}_0)$ is a loop, since $\widetilde{\lambda}$ is a loop.

Conversely, if $\widetilde{\lambda} = L(\lambda, \widetilde{x}_0)$ is a loop, then $[\lambda] = [p \circ \widetilde{\lambda}] = p_*([\widetilde{\lambda}])$, namely $[\lambda] \in p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$. □

An interesting consequence of the previous theorem is the following.

5.2.13 **Corollary**. *Let X be a connected and locally path-connected space and let $p : \widetilde{X} \longrightarrow X$ and $p' : \widetilde{X}' \longrightarrow X$ be covering maps such that their total spaces $\widetilde{X}$ and $\widetilde{X}'$ are connected. If $\widetilde{x}_0 \in \widetilde{X}$, $\widetilde{x}'_0 \in \widetilde{X}'$ and $x_0 \in X$ are base points such that $p(\widetilde{x}_0) = x_0 = p'(\widetilde{x}'_0)$, then both covering maps are equivalent if and only if*

$$p_*(\pi_1(\widetilde{X}, \widetilde{x}_0)) = p'_*(\pi_1(\widetilde{X}', \widetilde{x}'_0)),$$

or, equivalently, if both have the same characteristic subgroup.

Proof: An equivalence $\widetilde{\varphi} : \widetilde{X}' \longrightarrow \widetilde{X}$ such that $p \circ \widetilde{\varphi} = p'$, is a lifting of p', which is a homeomorphism. Clearly, if this homeomorphism exists, then the images of the fundamental groups of both total spaces must coincide. Conversely, if if these subgroups coincide, since p and p' are covering maps, then applying 5.2.12 to both, one obtains maps $\widetilde{\varphi} : \widetilde{X}' \longrightarrow \widetilde{X}$ lifting p', and $\widetilde{\varphi}' : \widetilde{X} \longrightarrow \widetilde{X}'$ lifting p. Furthermore, the composites $\varphi' \circ \varphi : \widetilde{X}' \longrightarrow \widetilde{X}'$ and $\varphi \circ \varphi' : \widetilde{X} \longrightarrow \widetilde{X}$ are maps lifting p' to $\widetilde{X}'$ and lifting p to $\widetilde{X}$, which fix $\widetilde{x}'_0$ and $\widetilde{x}_0$, respectively, since $\mathrm{id}_{\widetilde{X}'}$ and $\mathrm{id}_{\widetilde{X}}$ are also maps lifting p' to $\widetilde{X}'$ and lifting p to $\widetilde{X}$. Then, by the uniqueness of the lifting maps, one has $\varphi \circ \varphi' = \mathrm{id}_{\widetilde{X}}$ and $\varphi' \circ \varphi = \mathrm{id}_{\widetilde{X}'}$. □

The characteristic subgroup of a covering map depends on the base point $\widetilde{x}_0 \in p^{-1}(x_0)$. In the following theorem we shall study this dependence.

5.2.14 **Theorem**. *Let $p : \widetilde{X} \longrightarrow X$ be a covering map and take a base point $x_0 \in X$ and points $\widetilde{x}_0, \widetilde{x}'_0 \in p^{-1}(x_0)$. Then the following hold:*

(a) *If $\widetilde{\omega} : \widetilde{x}_0 \simeq \widetilde{x}'_0$ is a path in $\widetilde{X}$, and $\alpha = [p \circ \widetilde{\omega}] \in \pi_1(X, x_0)$, then $p_*(\pi_1(\widetilde{X}, \widetilde{x}_0)) = \alpha p_*(\pi_1(\widetilde{X}, \widetilde{x}'_0))\alpha^{-1}$. Hence, if $\widetilde{X}$ is path connected, then two characteristic subgroups of p are always conjugate.*

(b) *If $H \subset \pi_1(X, x_0)$ is a conjugate group of the characteristic subgroup $p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$, then H is a characteristic subgroup, namely, $H = p_*(\pi_1(\widetilde{X}, \widetilde{x}'_0))$, for some point $\widetilde{x}'_0 \in p^{-1}(x_0)$.*

Proof: (a) By 4.1.21, the elements of $\pi_1(\widetilde{X}, \widetilde{x}_0)$ are of the form $[\widetilde{\omega}]\gamma[\widetilde{\omega}]^{-1}$, where γ varies in $\pi_1(\widetilde{X}, \widetilde{x}'_0)$. Applying p_*, the statement follows.

(b) Take the subgroup $H = \alpha^{-1}\pi_1(\widetilde{X}, \widetilde{x}_0)\alpha$ for some element $\alpha = [\omega] \in \pi_1(X, x_0)$. If $\widetilde{\omega} = L(\omega, \widetilde{x}_0)$, take $\widetilde{x}'_0 = \widetilde{\omega}(1)$. Then $H = p_*(\pi_1(\widetilde{X}, \widetilde{x}'_0))$. □

As a consequence of the previous results, we have that the characteristic subgroups of a covering map, whose total space is path connected, build up the family of all conjugate subgroups of a certain fixed subgroup, that is, the family $\mathcal{C}(\widetilde{X}, p) = \{p_*(\pi_1(\widetilde{X}, \widetilde{x}_0)) \mid \widetilde{x}_0 \in p^{-1}(x_0)\}$ builds up a complete conjugacy class of subgroups of $\pi_1(X, x_0)$, which we call *characteristic conjugacy class* of the covering map $p : \widetilde{X} \longrightarrow X$. Its relevance is shown in the next result, which is somehow a reformulation of 5.2.13.

5.2.15 **Proposition**. *Let X be a connected and locally path-connected space and let $p : \widetilde{X} \longrightarrow X$ and $p' : \widetilde{X}' \longrightarrow X$ be covering maps, such that their total spaces $\widetilde{X}$ and $\widetilde{X}'$ are connected. Both covering maps are equivalent if and only if their characteristic conjugacy classes $\mathcal{C}(\widetilde{X}, p)$ and $\mathcal{C}(\widetilde{X}', p')$ coincide.*

Proof: If p and p' are equivalent, it is clear that their characteristic conjugacy classes coincide. Conversely, if $\mathcal{C}(\widetilde{X}, p)$ and $\mathcal{C}(\widetilde{X}', p')$ coincide, then $p_*(\pi_1(\widetilde{X}, \widetilde{x}_0)) \in \mathcal{C}(\widetilde{X}', p')$. Therefore
$$p_*(\pi_1(\widetilde{X}, \widetilde{x}_0)) = p_*(\pi_1(\widetilde{X}', \widetilde{x}'_0))$$
and thus, by 5.2.13, p and p' are equivalent. □

Given a covering map $p : \widetilde{X} \longrightarrow X$, let $\widetilde{x}_0, \widetilde{x}_1, \dots \in p^{-1}(x_0)$ be all the points of the fiber, whose number is the multiplicity of p. For $i = 0, 1, \dots$, let $\widetilde{\omega}_i : \widetilde{x}_0 \simeq \widetilde{x}_i$ be a path in $\widetilde{X}$. Then $\alpha_i = [p \circ \widetilde{\omega}_i] \in \pi_1(X, x_0)$ and the cosets
$$\alpha_0(p_*\pi_1(\widetilde{X}, \widetilde{x}_0))\,,\ \alpha_1(p_*\pi_1(\widetilde{X}, \widetilde{x}_0))\,, \dots$$
in the group $\pi_1(X, x_0)$ are all different cosets of the subgroup $p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$ in $\pi_1(X, x_0)$. The number of such cosets is the *index* of the subgroup $p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$ in $\pi_1(X, x_0)$, namely $[\pi_1(X, x_0) : p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))]$. We have shown the next result.

5.2.16 **Theorem**. *The multiplicity of a path connected covering map $p : \widetilde{X} \longrightarrow X$ is the index of its characteristic subgroup in the fundamental group of the base space, namely, it is $n = [\pi_1(X, x_0) : p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))]$ (n can be infinite).* □

5.2.17 Example. For the covering map $p : \mathbb{R} \longrightarrow \mathbb{S}^1$ given by $p(t) = \mathrm{e}^{2\pi \mathrm{i} t}$, one has $\pi_1(\mathbb{S}^1, 1) \cong \mathbb{Z}$ and $\pi_1(\mathbb{R}, 0) = 1$. Then the characteristic subgroup is trivial.

The characteristic conjugacy class $\mathcal{C}(\mathbb{R},p)$ consists of only one element, and the multiplicity of p, which is infinite, coincides with $[\mathbb{Z}:1]$, which is the cardinality of $\mathbb{Z}$.

On the other hand, if $p_n : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ is the map of degree n, given by $\zeta \mapsto \zeta^n$, then $p_{n*} : \pi_1(\mathbb{S}^1, 1) \longrightarrow \pi_1(\mathbb{S}^1, 1)$ corresponds to the homomorphism $\mu_n : \mathbb{Z} \longrightarrow \mathbb{Z}$, given by $\mu_n(k) = nk$. Hence the characteristic subgroup is $n\mathbb{Z}$ and the multiplicity of p_n is $[\mathbb{Z} : n\mathbb{Z}] = n$.

5.2.18 Exercise. Take $\widetilde{X} = \mathbb{R}\times\mathbb{Z}\cup\mathbb{Z}\times\mathbb{R} \subset \mathbb{R}^2$ and $X = \mathbb{S}^1 \vee \mathbb{S}^1 \subset \mathbb{S}^1 \times \mathbb{S}^1$. Let $p : \widetilde{X} \longrightarrow X$ be given by $(s,n) \mapsto (\mathrm{e}^{2\pi i s}, 1)$ and $(m,t) \mapsto (1, \mathrm{e}^{2\pi i t})$, where $m, n \in \mathbb{Z}$ and $s, t \in \mathbb{R}$. Show the following:

(a) $p : \widetilde{X} \longrightarrow X$ is a covering map of infinite multiplicity.

(b) $\pi_1(\widetilde{X}, (0,0))$ is a free group of infinite rank (i.e., with an infinite number of generators).

(c) The characteristic subgroup $p_*(\pi_1(\widetilde{X}, (0,0)))$ is the commutator of $\pi_1(X, x_0) = \langle a, b \mid - \rangle$, namely of the free group with two generators a, b.

(d) This commutator is free of infinite rank. For each natural number n there is a free subgroup of $\langle a, b \mid - \rangle$ of rank n.

5.3 Universal covering maps

Under the adequate assumptions on X it is possible to obtain any covering map over X as the quotient of a "universal" covering map. In this section we shall see how to construct such covering map, we shall study its properties, and we shall analyze the consequences of its existence.

5.3.1 Definition. A covering map $p : \widetilde{X} \longrightarrow X$ is called *universal* if the total space $\widetilde{X}$ is simply connected. The total space $\widetilde{X}$ is called *universal covering space.*

5.3.2 **Proposition**. *Let $p : \widetilde{X} \longrightarrow X$ be a covering map. Then the following are equivalent:*

(a) *$p : \widetilde{X} \longrightarrow X$ is universal.*

(b) *The characteristic conjugacy class $\mathcal{C}(\widetilde{X}, p)$ consists only of the trivial subgroup $1 \subset \pi_1(X, x_0)$.*

(c) *No loop in X based on x_0, unless it is nullhomotopic, lifts to a loop in $\widetilde{X}$.* □

Assume that X is locally path connected. By 5.2.15, any two universal covering maps over X are equivalent. Hence, there is no essential ambiguity if one talks about *the universal covering map* of X, when it exists. By 5.2.16, its multiplicity is the order of the group $\pi_1(X, x_0)$. By the lifting theorem 5.2.11, if $p' : \widetilde{X}' \longrightarrow X$ is any covering map, then there is a map $q : \widetilde{X} \longrightarrow \widetilde{X}'$ lifting p. We have the following.

5.3.3 **Proposition**. *Let X be a connected and locally path connected space. Assume that $p : \widetilde{X} \longrightarrow X$ is the universal covering map and let $p' : \widetilde{X}' \longrightarrow X$ be any path connected covering map. Then there is a unique covering map $q : \widetilde{X} \longrightarrow \widetilde{X}'$ such that $p' \circ q = p$,* that is, p is *initial* among all covering maps. It is from this fact that it is called universal covering map.

Proof: We have already mentioned how to construct q such that $p' \circ q = p$. By 5.1.6, it is enough to prove that q is surjective, to show that it is a covering map.

Take a point $\widetilde{x}' \in \widetilde{X}'$ and let $x = p'(\widetilde{x}') \in X$. There is a point $\widetilde{x} \in \widetilde{X}$ such that $p(\widetilde{x}) = x$. Take $\widetilde{x}'_0 = q(\widetilde{x})$. Since $\widetilde{X}'$ is path connected, there is a path $\widetilde{\omega}' : \widetilde{x}'_0 \simeq \widetilde{x}'$. Then $\omega = p' \circ \widetilde{\omega}'$ is a loop in X based at x. Since p is a covering map, there is a path $\widetilde{\omega}$ lifting ω to $\widetilde{X}$ with starting point $\widetilde{x}$. Then the path $q \circ \widetilde{\omega}$ lifts ω to $\widetilde{X}'$ with starting point $\widetilde{x}_0$. Therefore, by the uniqueness of the lifting paths, the path $q \circ \widetilde{\omega}$ coincides with the given path $\widetilde{\omega}'$. Hence, in particular, $\widetilde{x}' = \widetilde{\omega}'(1) = q(\widetilde{\omega}(1))$. Therefore q is surjective. Figure 5.5 depicts the proof. □

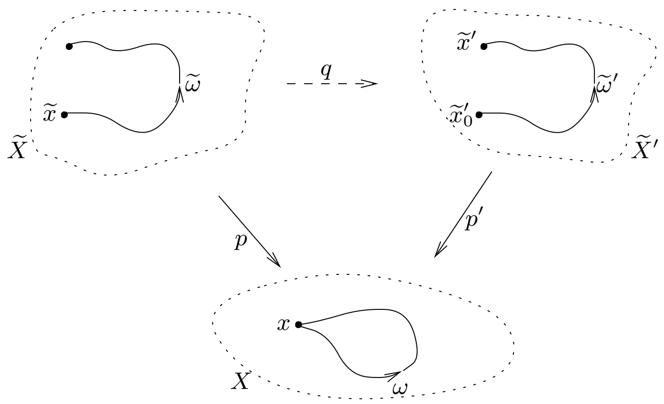


Figure 5.5 The universal covering map is initial among all covering maps

5.3.4 **Proposition**. *Let $p : \widetilde{X} \longrightarrow X$ be the universal covering map and let $p' : \widetilde{X}' \longrightarrow X$ be any other covering map. Then there is a unique covering map $q : \widetilde{X} \longrightarrow \widetilde{X}'$, such that $p' \circ q = p$,* namely, *p* es *initial among all covering maps*

Proof: We already mentioned how to construct q such that $p' \circ q = p$. By 5.1.6, q is a covering map. □

5.3.5 Examples. The universal covering spaces of the circle $\mathbb{S}^1$ and of the torus $\mathbb{S}^1 \times \mathbb{S}^1$ are $\mathbb{R}$ and $\mathbb{R}^2$, respectively. If $n \geq 2$, the sphere $\mathbb{S}^n$ is simply connected, and consequently the covering map $\mathbb{S}^n \longrightarrow \mathbb{RP}^n$ is universal. Since this covering map has two leaves, the fundamental group of $\mathbb{RP}^n$ has two elements. Hence we have the following result.

5.3.6 **Proposition**. *Take $n \geq 2$. Then the fundamental group of the n-dimensional projective space $\mathbb{RP}^n$ is cyclic of order 2, namely*

$$\pi_1(\mathbb{RP}^n) \cong \mathbb{Z}_2 \,.$$

□

In what follows we shall analyze the conditions under which a topological space X admits a universal covering map $p : \widetilde{X} \longrightarrow X$. Most interesting spaces, which play an important role in different branches of mathematics, satisfy these conditions, so that they have a universal covering map. This is a reason why the theory of covering maps has successful applications in other branches.

Before stating the required definitions, assume that $p : \widetilde{X} \longrightarrow X$ is a universal covering map and that X is path connected y locally path connected. Given $x \in X$, there is a path connected neighborhood U of x, which is evenly covered by p. Therefore, if λ is a loop in U, this loop lifts to a loop $\widetilde{\lambda}$ to $\widetilde{X}$. Now, since $\widetilde{X}$ is simply connected, the loop $\widetilde{\lambda}$ is nullhomotopic. Consequently, $\lambda = p \circ \widetilde{\lambda}$ is nullhomotopic in X. This establishes a necessary condition on X in order to admit a universal covering map. We have the next.

5.3.7 Definition.

(a) A topological space X is said to be *semilocally simply connected* if each point $x \in X$ has a neighborhood U, such that each loop in U is nullhomotopic in X.

(b) A topological space X is said to be *sufficiently connected* if X is path connected, locally path connected and semilocally simply connected.

As we already saw, the condition that X is semilocally simply connected is necessary in order that X has a universal covering map. We shall see that a sufficient condition in order for a space X to have a universal covering map, is that it is sufficiently connected.

Notice that a sufficient condition in order for a space X to be semilocally simply connected, is the condition that it is *locally simply connected*, namely, that each point of X has a simply connected neighborhood. For instance, a contractible neighborhood. Therefore, all manifolds are semilocally simply connected.

5.3.8 EXERCISE. For each natural number n, let $C_n \subset \mathbb{R}^2$ be the circle with center at $(\frac{1}{2^n}, 0)$ and radius $\frac{1}{2^n}$, and define $C = \cup_n C_n$. This space is usually called the *hawaiian earring*.

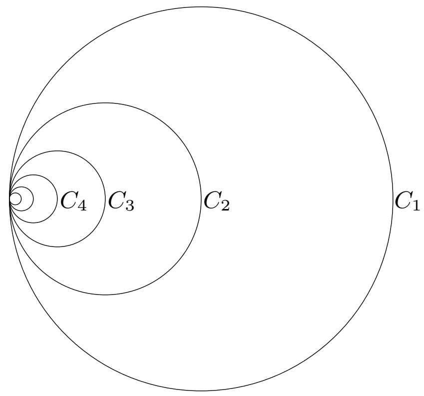


Figure 5.6 The hawaiian earring

(a) Show that C is connected and locally path connected, but it is not semilocally simply connected.

(b) Let $X \subset \mathbb{R}^3$ be the cone over C, that is, the union of all line segments joining a point in C with the point $(0, 0, 1)$. Show that X is semilocally simply connected, but it is not locally simply connected.

5.3.9 **Theorem**. (Existence of the universal covering map) *If X is a sufficiently connected space, then X admits a universal covering map $p : \widetilde{X} \longrightarrow X$.*

Proof: The idea of the proof is the following. Assume that we already have the a universal covering map $p : \widetilde{X} \longrightarrow X$. Cover X with a collection $\{U_j \mid j \in \mathcal{J}\}$ of evenly covered open sets. Each inverse image $p^{-1}(U_j)$ consists of as many homeomorphic open sets, as there are elements in the fundamental group $\pi_1(X, x_0)$.

In other words, $p^{-1}(U_j)$ is homeomorphic to $U_j \times \pi_1(X, x_0)$, where we consider the group $\pi_1(X, x_0)$ as a discrete space. The homeomorphisms $U_j \times \pi_1(X, x_0) \longrightarrow p^{-1}(U_j)$ are assembled together in such a way that one may define an identification

$$Y = \sqcup_j U_j \times \pi_1(X, x_0) \longrightarrow \widetilde{X}\,.$$

Namely, we take the leaves $U_j \times \{\alpha\}$, for each $\alpha \in \pi_1(X, x_0)$, over each U_j separately, and then we glue them conveniently. Let us now proceed with the proof.

As a **first step**, assume that each open set U_j is such that every loop in it is nullhomotopic in X. Then we fix a path in X, $\mu_j : I \longrightarrow X$ starting at x_0 and ending at some point in U_j, in such a way that the following holds:

(a) If $x_0 \in U_j$, then μ_j is the constant path c_{x_0}.

If $x \in U_i \cap U_j$, take $g_{ij}(x) = [\mu_i \omega_i \overline{\omega}_j \overline{\mu}_j] \in \pi_1(X, x_0)$, where ω_i and ω_j are paths in U_i or U_j from $\mu_i(1)$ and $\mu_j(1)$, respectively, to x. Since any loop in U_i or U_j is nullhomotopic in X, the element $g_{ij}(x)$ does not depend on the choices made, and the following conditions hold:

(b1) If $x \in U_i$, then $g_{ii}(x) = 1$,

(b2) If $x \in U_i \cap U_j$, then $g_{ij}(x) = g_{ji}(x)^{-1}$

(b3) If $x \in U_i \cap U_j \cap U_k$, then $g_{ij}(x) g_{jk}(x) = g_{ik}(x)$.

These are called *cocycle conditions* (see [2]).

(c) If $W \subset U_i \cap U_j$ and W is path connected, then $g_{ij}(x) = g_{ij}(y)$ for any x, $y \in W$.

For the **second step**, take the product $X \times \pi_1(X, x_0) \times \mathcal{J}$, where $\pi_1(X, x_0)$ and $\mathcal{J}$ have the discrete topology, and consider the subspace Y of the triples (x, α, j), where $x \in U_j$. Then the space Y is the disjoint union of the open sets $U_j \times \pi_1(X, x_0) \times \{j\}$ (as we already have mentioned in the introduction of this proof). We declare two triples (x, α, j) and (y, β, i) as equivalent, if $x = y$ and $\beta = g_{ij}(x)\alpha$. By (b), this is an equivalence relation $\sim$.

Let $\widetilde{X} = Y/\sim$ be the quotient space under this relation, and let $q : Y \longrightarrow \widetilde{X}$ be the identification.

(d) If $V \subset U_j$ is open in X, then $q(V \times \{\alpha\} \times \{j\})$ is open in $\widetilde{X}$ for all α.

To verify this, let us see that the intersection $q^{-1}q(V\times\{\alpha\}\times\{j\})\cap U_i\times\{\beta\}\times\{i\})$ is open in Y for all β and all i. Indeed, this intersection consists of the triples (y,β,i) with $y\in V\cap U_i$ and $\beta=g_{ij}(y)\alpha$. If $W\subset V\cap U_i$ is a path connected neighborhood of y in X, then by (c), $W\times\{\beta\}\times\{i\})$ also lies in the intersection. Hence it is open.

For the **third step** of the proof let us take the projection $\mathrm{proj}_X : Y\longrightarrow X$ given by $(x,\alpha,j)\mapsto x$ and define $p:\widetilde{X}\longrightarrow X$, such that $p\circ q=\mathrm{proj}_X$. Since p is well defined and q is an identification, then p is continuous. By (d), $\widetilde{U}_{j,\alpha}=q(U_j\times\{\alpha\}\times\{j\})$ is open in $\widetilde{X}$ and since $g_{jj}(x)=1$, one has $\widetilde{U}_{j,\alpha}\cap\widetilde{U}_{j,\beta}=\emptyset$ if $\alpha\neq\beta$. Furthermore, it is clear that $p^{-1}(U_j)$ is the union of the sets $\widetilde{U}_{j,\alpha}$, with $\alpha\in\pi_1(X,x_0)$. Thus, for each α, $p|_{\widetilde{U}_{j,\alpha}}:\widetilde{U}_{j,\alpha}\longrightarrow U_j$ is continuous, bijective, and by (c), it is open. Consequently, $p|_{\widetilde{U}_{j,\alpha}}$ is a homeomorphism. Consequently U_j is evenly covered by p, and so p is a covering map.

We still have to check that $\widetilde{X}$ is simply connected. This will be the **fourth step**. Let $\lambda:I\longrightarrow X$ be a loop based at x_0 and let $q(x_0,\alpha,j)$ be a point in the fiber over x_0, where $x_0\in U_j$. We have the following:

(e) The path lifting $\widetilde{\lambda}$ that starts at $q(x_0,\alpha,j)$, has $q(x_0,[\lambda]^{-1}\alpha,j)$ as destination.

From this it follows that any two points $q(x_0,\alpha,j)$ and $q(x_0,\beta,j)$ can be joined by a path, namely by the image of a representative ω of $\alpha\beta^{-1}$. From this we may deduce that $\widetilde{X}$ is path connected. Furthermore, a loop λ lifts to a loop if and only if $\alpha=[\lambda]^{-1}\alpha$ for some α, since $g_{jj}(x_0)=1$, that is, if and only if $[\lambda]=1$. In other words, the only loops that lift to loops are the nullhomotopic loops. This proves that $p:\widetilde{X}\longrightarrow X$ is universal. Thus it is enough to prove (e).

To do it, take a partition $0=t_0<t_1<\cdots<t_n=1$ and choose sets $U_{j_1},U_{j_2},\ldots,U_{j_n}$, among all sets U_j defined in the first step, in such a way that $\lambda([t_{\nu-1},t_\nu])\subset U_{j_\nu}$, for $\nu=1,2,\ldots,n$. Put $x_\nu=\lambda(t_\nu)$ and let $\widetilde{\lambda}:I\longrightarrow\widetilde{X}$ be such that

$$\widetilde{\lambda}(t)=\begin{cases} q(\lambda(t),\alpha,1) & \text{for } t_0\le t\le t_1,\\ q(\lambda(t),g_{21}(x_1)\alpha,2) & \text{for } t_1\le t\le t_2,\\ q(\lambda(t),g_{32}(x_2)g_{21}(x_1)\alpha,3) & \text{for } t_2\le t\le t_3,\\ \qquad\vdots & \\ q(\lambda(t),g_{n,n-1}(x_{n-1})\cdots g_{21}(x_1)\alpha,n) & \text{for } t_{n-1}\le t\le t_n.\end{cases}$$

This is a continuous path continua, such that $p\circ\widetilde{\lambda}=\lambda$ and $\widetilde{\lambda}(0)=q(x_0,\alpha,1)$, that is, it is the desired path lifting λ. From the definition of the cocycles g_{ij}, one may conclude from the first part that $g_{n,n-1}(x_{n-1})\cdots g_{21}(x_1)=[\lambda]^{-1}$. Thus $\widetilde{\lambda}(1)=q(x_0,[\lambda]^{-1}\alpha,1)$, since $g_{1n}(x_0)=1$. This proves (e). □

5.3.10 EXERCISE. Let S be a closed surface different from $\mathbb{S}^2$ and $\mathbb{RP}^2$. Then $\pi_1(S)$ is an infinite group. Therefore, its universal covering map has infinite multiplicity. The universal covering space $\widetilde{S}$ is homeomorphic to $\mathbb{R}^2$.

5.3.11 EXERCISE. In the product $\mathbb{B}^2 \times \{1, 2, \ldots, n\}$ (where $\{1, 2, \ldots, n\}$ is discrete) take the equivalence relation $(x, i) \sim (x, j)$ given by $x \in \mathbb{S}^1$ and $1 \leq i, j \leq n$. Show that the quotient space $\widetilde{X}$ obtained is the universal covering space of $X = \mathbb{S}^1 \cup_n e^2$. Which is the covering map?

5.3.12 EXAMPLE. Take $X = \mathbb{S}^1 \vee \mathbb{S}^1$, as shown in Figure 5.7.

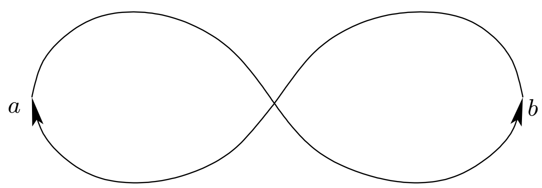


Figure 5.7 The wedge of two circles $\mathbb{S}^1 \vee \mathbb{S}^1$

As we know, $\pi_1(X, x_0)$ is a free group in two generators, that is $\mathbb{Z} * \mathbb{Z}$. Its universal covering space $\widetilde{X}$ can be constructed as an *infinite tree*, namely as the picture shown in Figure 5.8, where a part of the tree is depicted.

The line segments in the tree, with decreasing size, are to be considered as equally long intervals. They are drawn so to avoid overlapping. The figure shows how to identify one with the others, namely with infinitely many copies of the interval I. Each line segment is mapped onto one of the copies of the circle in X: the horizontal line segments are all mapped onto one, while the vertical line segments are mapped onto the other circle. The mapping restricted to each segment is the usual exponential map $I \longrightarrow \mathbb{S}^1$, given by $s \mapsto \mathrm{e}^{2\pi \mathrm{i} s}$.

5.3.13 EXERCISE. Let X be the wedge $\mathbb{S}^1_1 \vee \mathbb{S}^1_2$ of two copies of the circle, and let $\widetilde{X}$ be as described in the previous example. Take $p : \widetilde{X} \longrightarrow X$ be such that its restriction to each copy of I is the exponential map $s \mapsto \mathrm{e}^{2\pi \mathrm{i} s}$, either onto the copy $\mathbb{S}^1_1$ if the line segment is horizontal, or onto the copy $\mathbb{S}^1_2$ if the line segment is vertical.

(a) Show that $\widetilde{X}$ is simply connected and that p is the universal covering map of X.

(b) Show that the free group $\mathbb{Z} * \mathbb{Z}$ acts evenly on $\widetilde{X}$. (*Hint*: The generator 1_1 of the first copy of $\mathbb{Z}$ acts by shifting each horizontal line segment to the next

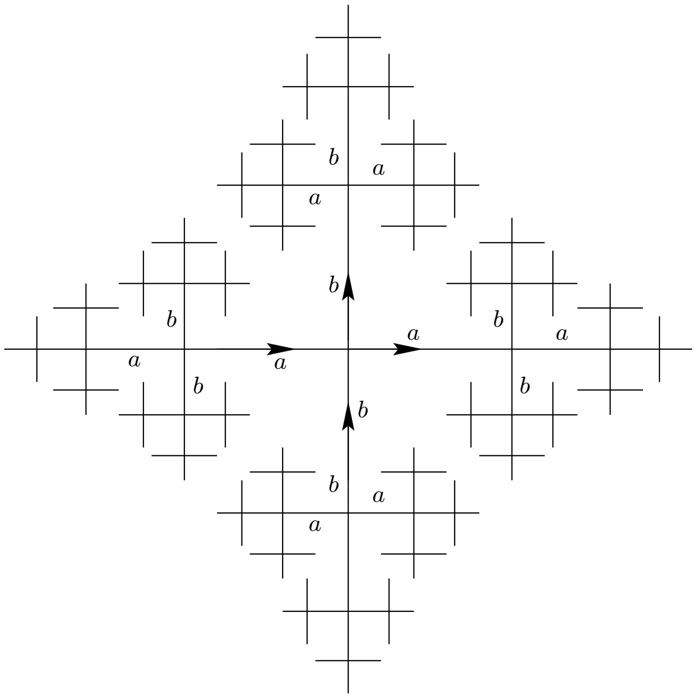


Figure 5.8 The universal covering space of $\mathbb{S}^1 \vee \mathbb{S}^1$

horizontal line segment on the right, and with it also the vertical ones. On the other hand, the generator 1_2 of the second copy of $\mathbb{Z}$ does the same thing, but vertically upwards.) Deduce from the former, that the fundamental group of X is the given free group.

(c) Generalize the previous construction to a wedge of k copies of the circle. (*Hint*: Take k linearly independent directions instead of 2, namely, build the corresponding tree as if it were in $\mathbb{R}^k$.)

5.3.14 Exercise. Show that the universal covering space of $X = \mathbb{R}^2 - 0$ can be realized as the open right half-plane $\widetilde{X} = \{(r, \theta) \in \mathbb{R}^2 \mid r > 0, -\pi < \theta < \pi\}$ with the map given in polar coordinates

$$p(r, \theta) = (r\cos(\theta), r\sin(\theta))\,.$$

Show furthermore, that it can also be realized as all the complex plane $\widetilde{X}' = \mathbb{C}$, with the exponential map $z \mapsto \mathrm{e}^z$. Determine an equivalence between both covering maps.

Consider a connected locally path-connected space X and let $\mathcal{C}_X$ by the set of conjugacy classes of all subgroups $H \subset \pi_1(X, x_0)$. Given any path-connected covering map $p : E \longrightarrow X$, just before Proposition 5.2.15 we defined the characteristic conjugacy class $\mathcal{C}(E, p)$ as the set $\{p_*(\pi_1(E, e_i)) \subset \pi_1(X, x_0) \mid e_i \in p^{-1}(x_0)\}$, which is a complete set of conjugacy classes of a subgroup of $\pi_1(X, x_0)$. By 5.2.15, we have that the assignment $[p] \mapsto \mathcal{C}(E, p)$ gives a well defined injective mapping, where $[p]$ denotes the class of all covering maps which are equivalent to p. This proves, among other things, that the equivalence classes of all connected covering maps over the space X constitute a set, which we denote by $\mathcal{C}ov(X)$. Thus we have a well-defined injective function

$$\Xi : \mathcal{C}ov(X) \longrightarrow \mathcal{C}_X$$

given by $\Xi[p] = \mathcal{C}(E, p)$. Indeed we have the following classification theorem, which asserts that covering maps over a space X are classified, up to equivalence, by the conjugacy classes of subgroups $H \subset \pi_1(X, x_0)$.

5.3.15 **Theorem**. *Let X be a sufficiently connected space. Then the function*

$$\Xi : \mathcal{C}ov(X) \longrightarrow \mathcal{C}_X$$

is a bijection.

Proof: We only have to prove that given any subgroup $H \subset \pi_1(X, x_0)$, there is a connected covering map $p : E \longrightarrow X$ such that the subgroup $p_*(\pi_1(E, e_0)) \subset \pi_1(X, x_0)$, where $p(e_0) = x_0$, is conjugate to H. We sketch the construction of p. It follows the same steps as the construction of the universal covering map over $\widetilde{X} \longrightarrow X$ given in the proof of Theorem 5.3.9.

The only change is to choose the identification space $Y \longrightarrow E$ adequately. Then follow the same steps.

In particular, in the second step, take the product $X \times (\pi_1(X, x_0)/H) \times \mathcal{I}$. Then define Y as the subset of triples $(x, \alpha H, j)$, where $x \in U_j$ and αH is a coset of H in $\pi_1(X, x_0)/H$. Now declare two triples $(x, \alpha H, j)$ and $(y, \beta H, i)$ as equivalent, if $x = y$ and $\alpha H = g_{ij}(x)\beta H$ or, equivalently, if $\alpha^{-1} g_{ij}(x)\beta \in H$. Define E as $Y/\sim$.

The rest of the proof is essentially the same. Just the fourth step deserves some argumentation. Indeed we have to prove that the characteristic subgroup of p is conjugate to H. To do this we apply 5.2.12. We shall show that a loop λ such that $[\lambda] \in H$ is characterized by the fact that it lifts to a loop $\widetilde{\lambda}$. Thus by 5.2.12 the characteristic subgroup of p must be H.

Take a loop λ in X based at x_0. Then, similarly to (e) in the proof of 5.3.9, we may prove the next:

- The path lifting $\widetilde{\lambda}$ that starts at $q(x_0, \alpha H, j)$, has $q(x_0, [\lambda]^{-1}\alpha H, j)$ as destination.

This states, in other words that a loop λ lifts to a loop $\widetilde{\lambda}$ if and only if $[\lambda] \in \alpha H \alpha^{-1}$.

5.3.16 Exercise. Verify all details of the previous proof sketch. In particular, for the last part of the proof, namely that the characteristic conjugacy class of the constructed covering map $p : E \longrightarrow X$ is the conjugacy class of H in $\pi_1(X, x_0)$.

5.4 Deck transformations

Covering maps codify information in many different forms. In the previous section we saw that the fundamental group of the base space is involved in the fiber. Indeed, the number of leaves of an arbitrary path connected covering map $p : \widetilde{X} \longrightarrow X$ is the index of the subgroup $p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$ in the fundamental group of the base space $\pi_1(X, x_0)$. In particular, when p is the universal covering map, this number of leaves is the order of the fundamental group of the base space.

In this section we shall again extract information from the covering maps, but in a different way.

5.4.1 Definition. A *deck transformation* of a covering map $p : \widetilde{X} \longrightarrow X$ is a fiberwise homeomorphism $f : \widetilde{X} \longrightarrow \widetilde{X}$, namely such that $p \circ f = p$. Hence f maps each fiber $p^{-1}(x)$ onto itself. Under composition, the deck transformations constitute a group $\mathcal{D}(\widetilde{X}, p)$, which is simply called *group of deck transformations*.

Next examples show the type of information kept inside $\mathcal{D}(\widetilde{X}, p)$.

5.4.2 Examples.

(a) The covering map $p : \mathbb{R} \longrightarrow \mathbb{S}^1$, $p(t) = \mathrm{e}^{2\pi i t}$ (5.1.1) has as the only possible deck transformations the homeomorphisms $\mathbb{R} \longrightarrow \mathbb{R}$ given by $x \mapsto x + n$, $n \in \mathbb{Z}$. Thus the group of deck transformations $\mathcal{D}(\mathbb{R}, p)$ is in this case $\mathbb{Z}$.

(b) In the case of the torus $p \times p : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{S}^1 \times \mathbb{S}^1$, the deck transformations are the homeomorphisms $\mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R} \times \mathbb{R}$ given by $(x, y) \mapsto (x+m, y+n)$, $m, n \in \mathbb{Z}$. Therefore in this case the group of deck transformations $\mathcal{D}(\mathbb{R} \times \mathbb{R}, p \times p)$ is isomorphic to $\mathbb{Z} \times \mathbb{Z}$.

(c) If we take the covering map $g_n : \mathbb{S}^1 \longrightarrow \mathbb{S}^1$ of degree n given by $g_n(\zeta) = \zeta^n$, then the deck transformations are the rotations of $\mathbb{S}^1$ by angles that are a multiple of $2\pi/n$. Hence the group of deck transformations $\mathcal{D}(\mathbb{S}^1, g_n)$ is isomorphic to the cyclic group $\mathbb{Z}_n$ of order n.

(d) The covering map $p : \mathbb{S}^n \longrightarrow \mathbb{RP}^n$ of 5.1.20 (a) has as the only deck transformations the identity $\mathrm{id}_{\mathbb{S}^n}$ and the antipodal map $-\mathrm{id}_{\mathbb{S}^n}$. Thus the group of deck transformations $\mathcal{D}(\mathbb{S}^n, p)$ is $\mathbb{Z}_2$.

(e) It is obvious that for the identical covering map $\mathrm{id}_X : X \longrightarrow X$, the group of deck transformations $\mathcal{D}(X, \mathrm{id}_X)$ consists only of the identity id_X, so that it is the trivial group 1.

(f) If G acts evenly on $\widetilde{X}$, then $f : \widetilde{X} \longrightarrow \widetilde{X}$ is a deck transformation for the associated covering map $p : \widetilde{X} \longrightarrow \widetilde{X}/G$, if it is a homeomorphism f such that the following is a commutative diagram

$$\begin{tikzcd}
\widetilde{X} \arrow[rr, "f"] \arrow[dr, "p"'] & & \widetilde{X} \arrow[dl, "p"] \\
& \widetilde{X}\,. &
\end{tikzcd}$$

Namely if for each $\widetilde{x} \in \widetilde{X}$, $p(f(\widetilde{x})) = p(\widetilde{x})$. This means that the orbit of $\widetilde{x}$ coincides with the orbit of $f(\widetilde{x})$. This is the case if and only if, for some element $\widetilde{x} \in \widetilde{X}$ and some $g \in G$, one has $p(\widetilde{x}) = g\widetilde{x}$. The maps $g, f : \widetilde{X} \longrightarrow \widetilde{X}$ both lift the orbit map $\widetilde{X} \longrightarrow \widetilde{X}/G$ and coincide in a point. Hence, by the unique map lifting theorem 5.2.2, both maps g and f are equal. This means that the group of deck transformations $\mathcal{D}(\widetilde{X}, p)$ of the orbit covering map of a group G acting evenly on a space $\widetilde{X}$, coincides with the group G.

If we complete what we proved in (f), then we have that indeed any path connected covering map is the orbit map of an even action of some group on the total space. We have the following.

5.4.3 **Theorem**. *Let $p : \widetilde{X} \longrightarrow X$ be a path connected covering map. The group of deck transformations $\mathcal{D} = \mathcal{D}(\widetilde{X}, p)$ acts evenly on $\widetilde{X}$, in such a way that in particular no deck transformation which is different from $\mathrm{id}_{\widetilde{X}}$ has fixed points, and two deck transformations that coincide in one point are the same. Furthermore, if the fundamental group $\pi_1(X)$ is abelian, then the orbit map $\widetilde{X} \longrightarrow \widetilde{X}/\mathcal{D}$ is a equivalent, as a covering map, to p.*

Proof: Let $\widetilde{x} \in \widetilde{X}$ be an arbitrary point and let $\widetilde{U}$ be the leaf that contains $\widetilde{x}$ over an evenly covered neighborhood U of $x = p(\widetilde{x})$. It is now enough to check that if $f \neq \mathrm{id}_{\widetilde{X}}$ is a deck transformation, then $\widetilde{U} \cap f(\widetilde{U}) = \emptyset$. This is true, since otherwise one would have a point $\widetilde{y} \in \widetilde{U}$ such that $f(\widetilde{y}) \in \widetilde{U}$. Due to the fact that $p(\widetilde{y}) = pf(\widetilde{y})$ and that $p|_{\widetilde{U}}$ is injective, then $\widetilde{y} = f(\widetilde{y})$. Again by 5.2.2, this can only be true if $f = \mathrm{id}_{\widetilde{X}}$.

To prove the second part of the theorem, observe first that if $f \in \mathcal{D}$, then $p(f(\widetilde{x})) = p(\widetilde{x})$. Therefore there is $\varphi : \widetilde{X}/\mathcal{D} \longrightarrow X$ such that the diagram

$$\begin{tikzcd} & \widetilde{X} \arrow[dl, "q"'] \arrow[dr, "p"] & \\ \widetilde{X}/\mathcal{D} \arrow[rr, dashed, "\varphi"'] & & X \end{tikzcd}$$

commutes. Clearly φ is surjective, since p is surjective. In fact, it is a covering map, hence it is open, so it is enough to verify that it is injective.

To see that φ is injective, take two points $\widetilde{x}_0, \widetilde{x}_1 \in \widetilde{X}$ such that $p(\widetilde{x}_0) = p(\widetilde{x}_1) = x$. We shall construct a deck transformation $f : \widetilde{X} \longrightarrow \widetilde{X}$ such that $f(\widetilde{x}_0) = \widetilde{x}_1$. Then both points will represent the same orbit in the orbit space $\widetilde{X}/\mathcal{D}$. To do this, let $\widetilde{\omega} : \widetilde{x}_0 \simeq \widetilde{x}_1$ be a path in $\widetilde{X}$ and let $\omega = p \circ \widetilde{\omega}$ be the induced loop in X. If $\alpha = [\omega] \in \pi_1(X, x)$, then $\alpha p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))\alpha^{-1} = p_*(\pi_1(\widetilde{X}, \widetilde{x}_1))$. But since $\pi_1(X, x)$ is abelian, we have $\alpha p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))\alpha^{-1} = p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$. Therefore, by the lifting theorem 5.2.11, the lifting problem

$$\begin{tikzcd} & \widetilde{X} \arrow[d, "p"] \\ \widetilde{X} \arrow[ur, dashed, "f"] \arrow[r, "p"'] & X, \end{tikzcd}$$

such that $f(\widetilde{x}_0) = \widetilde{x}_1$, has a solution. As in the proof of 5.2.15, f is a homeomorphism. Hence it is a deck transformation as desired. □

5.4.4 NOTE. In the last part of the last theorem on requires that $\pi_1(X)$ is abelian. If we omit this hypothesis, then the map $\varphi : \widetilde{X}/\mathcal{D} \longrightarrow X$ exists, is continuous and surjective, but not necessarily injective. It is an *exercise* to search for examples of covering maps for which φ is not injective.

Since each covering map is determined, up to isomorphism, by its characteristic conjugacy class $\mathcal{C}(\widetilde{X}, p)$ (5.2.15), there must be a relation between this class and the group $\mathcal{D}(\widetilde{X}, p)$. Recall that given a group G and a subgroup $H \subset G$, the *normalizer* $N_G(H)$ of H in G, is the maximal subgroup of G that contains H as a normal subgroup. It is given by

$$N_G(H) = \{g \in G \mid g^{-1}Hg = H\}.$$

If $p : \widetilde{X} \longrightarrow X$ is a covering map, let

$$\mathcal{N}(\widetilde{X}, \widetilde{x}_0, p) = N_{\pi_1(X,x_0)}(p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))).$$

We obtain a more general result than the last part of the previous theorem.

5.4.5 **Theorem**. *Let $p : \widetilde{X} \longrightarrow X$ be a path-connected covering map, whose base space X is locally path connected, and take a point $\widetilde{x}_0 \in \widetilde{X}$ over $x_0 \in X$. Then for each $\alpha = [\omega] \in \mathcal{N}(\widetilde{X}, \widetilde{x}_0, p) \subset \pi_1(X, x_0)$, there exists a unique deck transformation $f_\alpha : \widetilde{X} \longrightarrow \widetilde{X}$ given by $f_\alpha(\widetilde{x}_0) = L(\omega, \widetilde{x}_0)(1)$. The assignment $\alpha \mapsto f_\alpha$ determines a group epimorphism $\Phi : \mathcal{N}(\widetilde{X}, \widetilde{x}_0, p) \longrightarrow \mathcal{D}(\widetilde{X}, p)$ such that* $\ker(\Phi) = p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$*, so that it determines an isomorphism (equally denoted by)* $\Phi : \mathcal{N}(\widetilde{X}, \widetilde{x}_0, p)/p_*(\pi_1(\widetilde{X}, \widetilde{x}_0)) \longrightarrow \mathcal{D}(\widetilde{X}, p)$*. In other words, $\mathcal{D}(\widetilde{X}, p)$ is isomorphic to the quotient group NH/H, for a subgroup $H \subset \pi_1(X, x_0)$ which belongs to the class $\mathcal{C}(\widetilde{X}, p)$.*

Proof: Put $\widetilde{x}'_0 = L(\omega, \widetilde{x}_0)(1)$. Since $\alpha = [\omega]$ lies in the normalizer, by 5.2.14(a),

$$\begin{aligned} p_*(\pi_1(\widetilde{X}, \widetilde{x}_0)) &= \alpha^{-1} p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))\alpha \\ &= p_*(\pi_1(\widetilde{X}, \widetilde{x}'_0)) . \end{aligned}$$

Therefore, by 5.2.11, since X and consequently $\widetilde{X}$ are locally path connected, it is possible to solve the lifting problem

$$\begin{tikzcd} & \widetilde{X} \arrow[d, "p"] \\ \widetilde{X} \arrow[r, "p"'] \arrow[ur, dashed, "\widetilde{f}"] & X , \end{tikzcd}$$

in such a way that $\widetilde{f}(\widetilde{x}_0) = \widetilde{x}'_0$. As in the proof of 5.2.15, one has that $\widetilde{f}$ is a homeomorphism and hence a deck transformation. By 5.4.3, $\widetilde{f}$ is unique, and by 5.2.3, $\widetilde{f}$ depends only from the homotopy class $\alpha = [\omega]$. Put $\Phi(\alpha) = \widetilde{f}$. With this is Φ defined.

Let us see now that Φ es un homeomorphism. To see this, take $\beta = [\sigma] \in \mathcal{N}(\widetilde{X}, \widetilde{x}_0, p)$ and $\Phi(\beta) = \widetilde{g}$. Then by 5.2.6, one has

$$\begin{aligned} L(\sigma\omega, \widetilde{x}_0)(1) &= (L(\sigma, \widetilde{x}_0)L(\omega, \widetilde{g}(\widetilde{x}_0)))(1) \\ &= L(\omega, \widetilde{g}(\widetilde{x}_0))(1) \\ &= \widetilde{g}(L(\omega, \widetilde{x}_0)(1)) \\ &= \widetilde{g}\widetilde{f}(\widetilde{x}_0) . \end{aligned}$$

Therefore, the deck transformations $\Phi([\sigma\omega]) = \Phi(\beta\alpha)$ and $\widetilde{g} \circ \widetilde{f} = \Phi(\beta)\Phi(\alpha)$ coincide on $\widetilde{x}_0$ and hence they are the same.

The fact that $\ker(\Phi) = p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$ is immediately deduced from 5.2.12 and 5.4.3. We still have to check that Φ is surjective. To see this take $\widetilde{f} \in \mathcal{D}(\widetilde{X}, p)$ and let $\widetilde{\omega} : \widetilde{x}_0 \simeq \widetilde{f}(\widetilde{x}_0)$ and let $\alpha = [p \circ \widetilde{\omega}] \in \pi_1(X, x_0)$. From 5.2.14(a) we obtain

$$\begin{aligned} \alpha^{-1} p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))\alpha &= p_*(\pi_1(\widetilde{X}, \widetilde{f}(\widetilde{x}_0))) \\ &= p_*\widetilde{f}_*(\pi_1(\widetilde{X}, \widetilde{x}_0)) \\ &= p_*(\pi_1(\widetilde{X}, \widetilde{x}_0)) . \end{aligned}$$

Consequently, α lies in the normalizer of $p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$ and from the definition of Φ and from 5.4.3 one gets $\Phi(\alpha) = \widetilde{f}$. □

As a particular case of the former theorem, for the case of the universal covering map, we obtain the following.

5.4.6 **Corollary**. *If $p : \widetilde{X} \longrightarrow X$ is the universal covering map of a locally path-connected space X and $\widetilde{x}_0$ is such that $p(\widetilde{x}_0) = x_0$, then for each $\alpha = [\omega] \in \pi_1(X, x_0)$ there is a unique deck transformation $f_\alpha : \widetilde{X} \longrightarrow \widetilde{X}$ such that $f_\alpha(\widetilde{x}_0) = L(\omega, \widetilde{x}_0)(1)$. The function $\Phi : \pi_1(X, x_0) \longrightarrow \mathcal{D}(\widetilde{X}, p)$ is a group isomorphism.* □

Of course, the previous corollary allows us, in many cases, to compute the fundamental group, as one can see in the following.

5.4.7 Examples.

(a) For the universal covering map $p : \mathbb{R} \longrightarrow \mathbb{S}^1$, one has that $\mathcal{D}(\widetilde{X}, p)$ consists of the translations in $\mathbb{R}$, given by adding integers, so that $\mathcal{D}(\widetilde{X}, p) \cong \mathbb{Z} \cong \pi_1(\mathbb{S}^1)$.

(b) For the universal covering maps $\mathbb{S}^n \longrightarrow \mathbb{RP}^n$ of the real projective spaces, with $n > 1$, the only deck transformations $f : \mathbb{S}^n \longrightarrow \mathbb{S}^n$ are $f = \mathrm{id}_{\mathbb{S}^n}$ or $f = -\mathrm{id}_{\mathbb{S}^n}$, as one can easily verify. Hence $\mathcal{D}(\widetilde{X}, p) \cong \mathbb{Z}_2 \cong \pi_1(\mathbb{RP}^n)$.

(c) If G acts evenly on a simply connected space Y, then $Y \longrightarrow Y/G$ is the universal covering map, so that, by 5.4.2(f), $\mathcal{D}(Y, p) = G$. Hence $\pi_1(Y/G) \cong G$.

5.4.8 Exercise. Show that for a path connected covering map $p : \widetilde{X} \longrightarrow X$ the following statements are equivalent:

(a) For each pair of points $\widetilde{x}_0, \widetilde{x}_1 \in p^{-1}(x)$, the equality $p_*\pi_1(\widetilde{X}, \widetilde{x}_0) = p_*\pi_1(\widetilde{X}, \widetilde{x}_1)$ holds. Namely, the conjugacy class $\mathcal{C}(\widetilde{X}, p)$ consists of one subgroup of $\pi_1(X, x)$.

(b) For each $\widetilde{x} \in p^{-1}(x)$, $p_*\pi_1(\widetilde{X}, \widetilde{x}) \subset \pi_1(X, x)$ is a normal subgroup.

(c) If a loop ω in X based at x has a loop that lifts it, then any other path lifting ω is a loop.

Show that if one and, hence, all these conditions hold for some point $x \in X$, then they hold for any other point in X. Such a covering map is called *regular*.

Notice that any universal covering map is regular, as is also any covering map de X if $\pi_1(X)$ is abelian (as we saw in 5.4.3). Furthermore, every 2-fold covering map is regular (since every subgroup of index 2 is always normal).

5.4.9 Exercise. Show that given a regular covering map $p : \widetilde{X} \longrightarrow X$, there is an isomorphism
$$\Phi : \pi_1(X,x)/p_*\pi_1(\widetilde{X},\widetilde{x}) \longrightarrow \mathcal{D}(\widetilde{X},p)\,.$$

5.4.10 Exercise. If $p : \widetilde{X} \longrightarrow X$ is an n-fold covering map, and $\{\widetilde{x}_1, \widetilde{x}_2, \dots, \widetilde{x}_n\} = p^{-1}(x)$, show that then the following are equivalent.

(a) There is a deck transformation of $\widetilde{X}$ such that $\widetilde{x}_1 \mapsto \widetilde{x}_2 \mapsto \dots \mapsto \widetilde{x}_n \mapsto \widetilde{x}_1$.

(b) The group $p_*\pi_1(\widetilde{X},\widetilde{x}) \subset \pi_1(X,x)$ is a normal subgroup such that the group $\pi_1(X,x)/p_*\pi_1(\widetilde{X},\widetilde{x})$ is cyclic of order n.

5.5 Classification of covering maps over paracompact spaces

The goal of this section is to classify covering maps whose base space is paracompact, by means of classifying spaces. These classifying spaces will be configuration spaces. This result is extracted from [2]. Before starting, we recall what paracompact spaces are. We follow [21].

5.5.1 Definition. A topological space X is *paracompact* if it is Hausdorff and satisfies the following condition:

(PC) Every open cover of X admits a locally finite open refinement.

5.5.2 Definition. Let X be a topological space and let $f : X \longrightarrow \mathbb{R}$ be continuous. The *support* of f is the closed set
$$\operatorname{supp}(f) = \overline{\{x \in X \mid f(x) \neq 0\}}\,.$$

5.5.3 Definition. Let $\{\eta_\lambda : X \longrightarrow \mathbb{R}\}_{\lambda\in\Lambda}$ be a family of continuous functions. The family is called a *partition of unity* if

(a) $\eta_\lambda(x) \geq 0$ for all $x \in X$.

(b) The collection of supports $\{\text{supp}(\eta_\lambda)\}_{\lambda\in\Lambda}$ is locally finite, namely, for all $x \in X$ a neighborhood V of x exists such that $V \cap \text{sop}(\eta_\lambda) \neq \emptyset$ only for finitely many elements $\lambda \in \Lambda$. In this case, $\sum_{\lambda\in\Lambda} \eta_\lambda(x)$ is well defined, since it is a finite sum, and it is continuous since the sum remains finite inside a neighborhood of each point.

(c) $\sum_{\lambda\in\Lambda} \eta_\lambda(x) = 1$ for all $x \in X$.

Thus the family of supports $\{\text{supp}(\eta_\lambda)\}$ is a locally finite closed cover of the space X.

5.5.4 DEFINITION. Let $\{U_\lambda\}_{\lambda\in\Lambda}$ be an open cover of the topological space X. We say that a partition of unity $\{\eta_\lambda\}_{\lambda\in\Lambda}$ en X *is subordinate* to the cover if for every $\lambda \in \Lambda$, $\text{supp}(\eta_\lambda) \subset U_\lambda$.

The following characterization is what we shall need below. It is proved in [21, 9.5.23].

5.5.5 **Theorem**. *Let X be a Hausdorff space. Then X is paracompact if and only if every open cover of X has a subordinate partition of unity.*

We now pass to give the results for establishing the desired classification theorem. We prove a general result about covering maps.

5.5.6 **Lemma**. *Assume that $p : E \longrightarrow X \times I$ is a covering map whose restrictions to $X \times [0,a]$ and to $X \times [a,1]$ are trivial for some $a \in I$. Then $p : E \longrightarrow X \times I$ itself is trivial.*

Proof: By assumption we have homeomorphisms $\varphi_1 : (X \times [0,a]) \times F \longrightarrow p^{-1}(X \times [0,a])$ and $\varphi_2 : (X \times [a,1]) \times F \longrightarrow p^{-1}(X \times [a,1])$ over the respective base spaces. These homeomorphisms induce a map

$$(X \times \{a\}) \times F \xrightarrow{\varphi_1|} p^{-1}(X \times \{a\}) \xrightarrow{\varphi_2|^{-1}} (X \times \{a\}) \times F$$

of the form $(x,a,y) \mapsto (x,a,g(x)(y))$, where $g : X \longrightarrow \text{Bij}(F)$ is locally constant and $\text{Bij}(F)$ is the group of bijections of F, and $\varphi_1|$ and $\varphi_2|$ are the appropriate restrictions.

Now we define $\varphi : (X \times I) \times F \longrightarrow E$ by

$$\varphi(x,t,y) = \begin{cases} \varphi_1(x,t,y) & \text{if } t \leq a\,, \\ \varphi_2(x,t,g(x)(y)) & \text{if } t \geq a\,. \end{cases}$$

Then φ is the desired trivialization. □

5.5.7 **Lemma**. *Let $p : E \longrightarrow X \times I$ be a covering map. Then there is an open cover $\mathcal{U}$ of X such that $p|_{p^{-1}(U\times I)} : p^{-1}(U \times I) \longrightarrow U \times I$ is trivial for every $U \in \mathcal{U}$.*

Proof: Take $x \in X$. Then for each $t \in I$ there are neighborhoods U_t of x in X and V_t of t in I such that $p^{-1}(U_t \times V_t) \longrightarrow U_t \times V_t$ is trivial. Since I is compact, there is a finite subcover $\{V_{t_r} \mid r = 1, \ldots, m\}$ of the cover $\{V_t \mid t \in I\}$. If we set $U_x = \bigcap_{r=1}^m U_{t_r}$, and choose $0 = s_0 < s_1 < \cdots < s_n = 1$ such that the differences $s_i - s_{i-1}$ for $i = 1, \ldots, n$ are less than a Lebesgue number of the cover $\{V_{t_r}\}$, then the restricted covering map $p^{-1}(U_x \times [s_{i-1}, s_i]) \longrightarrow U_x \times [s_{i-1}, s_i]$ is trivial. And so by iterating and using Lemma 5.5.6 we have that $p^{-1}(U_x \times I) \longrightarrow U_x \times I$ is trivial as well. If we repeat this construction for every $x \in X$ we obtain an open cover $\{U_x\}$ of X such that each $p^{-1}(U_x \times I) \longrightarrow U_x \times I$ is trivial. □

5.5.8 **Proposition**. *Let $p : E \longrightarrow X \times I$ be a covering map, where X is a paracompact space. Let $r : X \times I \longrightarrow X \times I$ be the retraction given by $r(x,t) = (x,1)$ for $(x,t) \in X \times I$. Then there is a commutative diagram*

$$\begin{array}{ccc} E & \xrightarrow{f} & E \\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle p} \\ X \times I & \xrightarrow[r]{} & X \times I \end{array}$$

such that f restricted to each fiber is a bijection. Consequently, there is a homeomorphism $\varphi : E \longrightarrow r^*E = \{(x,t,e) \in X \times I \times E \mid p(e) = (x,1)\}$.

Such a pair of maps (f, r) is a *morphism* of covering maps.

Proof: Using 5.5.7 and the paracompactness of X there is a locally finite open cover $\{U_\lambda\}_{\lambda\in\Lambda}$ of X together with a subordinate partition of unity $\{\eta_\lambda\}_{\lambda\in\Lambda}$ (see Theorem 5.5.5) such that $p^{-1}(U_\lambda \times I) \longrightarrow U_\lambda \times I$ is trivial. For each $\lambda \in \Lambda$, define $\mu_\lambda : X \longrightarrow I$ by

$$\mu_\lambda(x) = \frac{\eta_\lambda}{\max\{\eta_{\lambda'}(x) \mid \lambda' \in \Lambda\}}\,.$$

Due to the fact that only a finite number of the $\eta_{\lambda'}(x)$ are nonzero, the function $\max\{\eta_{\lambda'}(x) \mid \lambda' \in \Lambda\}$ is continuous and nonzero. Therefore, μ_λ is continuous, has support in U_λ and for each $x \in X$ it satisfies $\max\{\mu_\lambda(x)\} = 1$.

Let $\varphi_\lambda : U_\lambda \times I \times F \longrightarrow p^{-1}(U_\lambda \times I)$ for each $\lambda \in \Lambda$ denote a local trivialization. For each $\lambda \in \Lambda$ we then define a morphism of covering maps

$$\begin{array}{ccc} E & \xrightarrow{f_\lambda} & E \\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle p} \\ X \times I & \xrightarrow[r_\lambda]{} & X \times I \end{array}$$

by setting, in the base space, $r_\lambda(x,t) = (x, \max\{\mu_\lambda(x), t\})$ for $(x,t) \in X \times I$ and by setting, in the total space, f_λ to be the identity outside of $p^{-1}(U_\lambda \times I)$ and by setting $f_\lambda(\varphi_\lambda(x,t,v)) = \varphi_\lambda(x, \max\{\mu_\lambda(x), t\}, v)$ inside of $p^{-1}(U_\lambda \times I)$. Now let us choose a well-ordering $\prec$ on Λ. By local finiteness we have that for each x in X there is a neighborhood W_x of x such that $W_x \cap U_\lambda$ is nonempty only for finitely many $\lambda \in \Lambda$, say for $\lambda \in \Lambda_x = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ with $\lambda_1 \prec \lambda_2 \prec \cdots \prec \lambda_m$. We now define $r : X \times I \longrightarrow X \times I$ by

$$r = r_{\lambda_m} \circ r_{\lambda_{m-1}} \circ \cdots \circ r_{\lambda_1}$$

and we define $f : E \longrightarrow E$ by

$$f|_{p^{-1}(W_x \times I)} = f_{\lambda_m} \circ f_{\lambda_{m-1}} \circ \cdots \circ f_{\lambda_1}\,.$$

Since r_λ on $W_x \times I$ and f_λ on $p^{-1}(W_x \times I)$ are the identity if $\lambda \notin \Lambda_x$, we can view r and f as infinite composites of maps almost all (i.e. all except finitely many) of which are the identity in a neighborhood of any point. Since each f_λ is a bijection on every fiber, the composite f also is a bijection on every fiber. $\square$

5.5.9 **Theorem**. *Let $p' : E' \longrightarrow X'$ be a covering map and let X be a paracompact space. Assume that we have two homotopic maps $f, g : X \longrightarrow X'$. Then we a homeomorphism $\varphi : f^*E' \approx g^*E'$ such that $q \circ \varphi = p$, where $p(x, e') = x = q(x, e')$.*

Proof: Let $F : X \times I \longrightarrow X'$ be a homotopy from f to g. Also let $i_n u : X \longrightarrow X \times I$ be the inclusion $i_\nu(x) = (x, \nu)$, $\nu = 0, 1$. It then follows that $f = F \circ i_0$ and $g = F \circ i_1$.

Let $r : X \times I \longrightarrow X \times I$ be the retraction defined by $r(x,t) = (x, 1)$. Then by applying 5.1.18, 5.5.8, and 5.1.19 we have that $f^*E' = (F \circ i_0)^*E' \approx i_0^*(F^*E') \approx i_0^*(r^*(F^*E')) \approx (r \circ i_0)^*(F^*E') = i_1^*(F^*E') \approx (F \circ i_1)^*E' = g^*E'$, where we have used $r \circ i_0 = i_1$. $\square$

5.5.10 Definition. Let X be a topological space. We define its *nth configuration space* $F_n(X)$ by

$$F_n(X) = \{(x_1, x_2, \ldots, x_n) \in X^n \mid x_i \neq x_j \text{ for } i \neq j\}\,.$$

If Σ_n denotes the *symmetric* (or *permutation*) *group* of the set $\{1, 2, \ldots, n\}$, then there is a right free action of this group on $F_n(X)$ given by

$$(x_1, x_2, \ldots, x_n)\sigma = (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})\,, \quad x_i \in X\,.$$

If X is a Hausdorff space, then by 5.1.24 the action is even. Hence it is free and by 5.1.25 the quotient map $q_n : F_n(X) \longrightarrow F_n(X)/\Sigma_n$ is a covering map of multiplicity

$n!$. There is also an *n-fold covering map*, that is, a covering map of multiplicity n, $\pi_n : E_n(X) \longrightarrow F_n(X)/\Sigma_n$ associated to $F_n(X)$ defined as follows. The total space is given by $E_n(X) = \{(C,x) \in (F_n(X)/\Sigma_n) \times X \mid x \in C\}$ and the projection is given by $\pi_n(C,x) = C$.

We shall consider only the case $X = \mathbb{R}^k$, where $1 \leq k \leq \infty$. It can be shown that the space $F_n(\mathbb{R}^\infty)$ is contractible (*exercise*).

5.5.11 EXERCISE. Consider two covering maps $p : E \longrightarrow X$ and $p' : E' \longrightarrow X$ over the same base space, such that there is a morphism (φ, id_X) from p to p', that is, if φ is a continuous map such that $p' \circ \varphi = p$ and for all $x \in X$, $\varphi|_{p^{-1}(x)} : p^{-1}(x) \longrightarrow p^{-1}(x)$ is bijective. Show that this morphism is an equivalence of covering maps, i.e. show that φ is a homeomorphism.

5.5.12 **Lemma**. *$p : E \longrightarrow X$ and $q : E' \longrightarrow X'$ be covering maps. Assume that there are maps $F : E \longrightarrow E'$ and $f : X \longrightarrow X'$ such that $q \circ F = f \circ p$ and for every $x \in X$, $F|_{p^{-1}(x)} : p^{-1}(x) \longrightarrow q^{-1}(f(x))$ is bijective. Then $p : E \longrightarrow X$ is equivalent to the covering map $q' : f^*E' \longrightarrow X$ induced from q by f.*

Proof: Consider the pullback diagram

$$\begin{tikzcd} f^*E' \arrow[r, "\widetilde{f}"] \arrow[d, "q'"'] & E' \arrow[d, "q"] \\ X \arrow[r, "f"'] & X' . \end{tikzcd}$$

Define $\varphi : E \longrightarrow f^*E'$ by $\varphi(e) = (p(e), F(e))$. Then fiberwise φ coincides with F so that it is a bijection of the fibers and thus it is an equivalence. $\square$

5.5.13 DEFINITION. Let $p : E \longrightarrow X$ be an n-fold covering map. A *Gauss map* for p is a map $g : E \longrightarrow \mathbb{R}^k$, $0 \leq k \leq \infty$, such that $g|_{p^{-1}(x)} : p^{-1}(x) \longrightarrow \mathbb{R}^k$ is injective for each $x \in X$.

5.5.14 **Proposition**. *Let $p : E \longrightarrow X$ be an n-fold covering map. Then there is a Gauss map for p if and only if there is a map $f : X \longrightarrow F_n(\mathbb{R}^k)/\Sigma_n$ such that p is equivalent to the pullback $q : f^*E_n(\mathbb{R}^k) \longrightarrow X$. The map f is called a* classifying map *of p.*

Proof: First assume that g is a Gauss map for p and define $f : X \longrightarrow F_n(\mathbb{R}^k)/\Sigma_n$ as follows. For each $x \in X$ choose a bijection $h : \overline{n} \longrightarrow p^{-1}(x)$, where $\overline{n} = \{1, 2, \ldots, n\}$. Since $g \circ h : \overline{n} \longrightarrow \mathbb{R}^k$ is injective, put

$$f(x) = \pi_n(gh(1), gh(2), \ldots, gh(n)) .$$

This is well defined, since given any other bijection $h' : \overline{n} \longrightarrow p^{-1}(x)$, the composite $\sigma = h'^{-1} \circ h$ is a permutation that belongs to Σ_n and

$$(gh'(1), gh'(2), \ldots, gh'(n))\sigma = (gh(1), gh(2), \ldots, gh(n)).$$

To see that f is continuous, take an open cover $\mathcal{U}$ of X such that each open set $U \in \mathcal{U}$ is evenly covered by p. Then we have a homeomorphism $\varphi_U : p^{-1}U \longrightarrow U \times \overline{n}$. For each $x \in U$ we have that the composite

$$p^{-1}(x) \xrightarrow{\varphi_U|} U \times \overline{n} \xrightarrow{\text{proj}} \overline{n}$$

is a bijection and $f(x) = \pi_n(g((\text{proj} \circ \varphi_U)^{-1}(1)), \ldots, g((\text{proj} \circ \varphi_U)^{-1}(n)))$. Now define $F : E \longrightarrow E_n(\mathbb{R}^k)$ by $F(e) = (fp(e), g(e))$ and thus get a commutative diagram

$$\begin{tikzcd} E \arrow[r, "F"] \arrow[d, "p"'] & E_n(\mathbb{R}^k) \arrow[d, "\pi_n"] \\ X \arrow[r, "f"'] & F_n(\mathbb{R}^k)/\Sigma_n . \end{tikzcd}$$

Since F is a bijection on fibers, by Lemma 5.5.12, $f^*E_n(\mathbb{R}^k) \approx E$.

Conversely, let $h : E \longrightarrow f^*E_n(\mathbb{R}^k)$ be an equivalence of covering maps. Then $g : E \longrightarrow \mathbb{R}^k$ given by the composite

$$\begin{tikzcd} f^*E_n(\mathbb{R}^k) \arrow[r, "\widetilde{f}"] & E_n(\mathbb{R}^k) \arrow[r, hook] & (F_n(\mathbb{R}^k)/\Sigma_n) \times \mathbb{R}^k \arrow[d, "\text{proj}"] \\ E \arrow[u, "h"] \arrow[rr, dashed, "g"'] & & \mathbb{R}^k \end{tikzcd}$$

is clearly a Gauss map. $\square$

5.5.15 Exercise. Let $p : E \longrightarrow X$ be an n-fold covering map.

(a) Show that the construction above establishes a one-to-one correspondence between the set of morphisms of the form

$$\begin{tikzcd} E \arrow[r, "\widetilde{f}"] \arrow[d, "p"'] & E_n(\mathbb{R}^k) \arrow[d, "\pi_n"] \\ X \arrow[r, "f"'] & F_n(\mathbb{R}^k)/\Sigma_n \end{tikzcd}$$

and the set of Gauss maps $g : E \longrightarrow \mathbb{R}^k$.

(b) Show that if $G : E \times I \longrightarrow \mathbb{R}^k$ is a homotopy such that for each $t \in I$, the map $G_t : E \longrightarrow \mathbb{R}^k$ given by $G_t(e) = G(e,t)$, is a Gauss map, then we can use the construction above to obtain a morphism of covering maps

$$\begin{tikzcd} E \times I \arrow[r, "\widetilde{F}"] \arrow[d, "p\times\mathrm{id}"'] & E_n(\mathbb{R}^k) \arrow[d, "\pi_n"] \\ X \times I \arrow[r, "F"'] & F_n(\mathbb{R}^k)/\Sigma_n \end{tikzcd}$$

with the property that if $f_\nu : X \longrightarrow F_n(\mathbb{R}^k)/\Sigma_n$ is the function associated to G_ν for $\nu = 0, 1$, then F is a homotopy between f_0 and f_1.

Now we prove that every n-fold covering map $p : E \longrightarrow X$, $n = 1, 2, \cdots$, over a paracompact space has a Gauss map. We need the following result.

5.5.16 **Lemma**. *Let $p : E \longrightarrow X$ be an n-fold covering map over a paracompact space X. Then there exists a countable open cover $\mathcal{W} = \{W_n \mid n \in \mathbb{N}\}$ of X such that $p|_{p^{-1}W_n} : p^{-1}W_n \longrightarrow W_n$ is trivial for all $n \in \mathbb{N}$.*

Proof: Let $\mathcal{U} = \{U_\lambda \mid \lambda \in \Lambda\}$ be an open cover of X such that the restriction $p|_{p^{-1}U_\lambda} : p^{-1}U_\lambda \longrightarrow U_\lambda$ is trivial for all $\lambda \in \Lambda$. Since X is paracompact, there is a partition of unity $\{\eta_\lambda \mid \lambda \in \Lambda\}$ subordinate to $\mathcal{U}$. For each $x \in X$ define $S(x)$ to be the finite set of those $\lambda \in \Lambda$ for which $\eta_\lambda(x)$ is nonzero. Also, for each finite subset $S \subset \Lambda$, set $W(S) = \{x \in X \mid \eta_\lambda(x) > \eta_\mu(x) \text{ whenever } \lambda \in S \text{ and } \mu \notin S\}$.

We claim that $W(S) \subset X$ is open. Indeed, the set $B_{\lambda,\mu} = \{x \in X \mid \eta_\lambda(x) < \eta_\mu(x)\}$ is clearly open, since $B_{\lambda,\mu} = (\eta_\mu - \eta_\lambda)^{-1}(0,1]$. Now for any given point $x_0 \in W(S)$, there is a neighborhood V_{x_0} of x_0 such that η_λ is different from zero in V_{x_0} only for $\lambda = \mu_1 \, dots, \mu_r$ for some finite r. Put $N = \bigcap_{\lambda\in S}(B_{\lambda,\mu_1}, \ldots, B_{\lambda,\mu_r})$, which is open since it is a finite intersection of open sets. Then $x_0 \in N \cap V_{x_0} \subseteq W(S)$ and hence $W(S)$ is open.

If S and T are two different subsets of Λ, each one containing m elements, the $W(S) \cap W(T) = \emptyset$, since otherwise an element $\lambda \in S$ would exist such that $\lambda \notin T$ and an element $\mu \in T$ would exist such that $\mu \notin S$. Hence $x \in W(S) \cap W(T)$ would imply $\eta_\lambda(x) > \eta_\mu(x)$ and $\eta_\mu(x) > \eta_\lambda(x)$, which is a flagrant contradiction.

Now define $W_n = \bigcup\{W(S(x)) \mid |S(x)| = n\}$, where $|\cdot|$ denotes the cardinality of a set.

If $\lambda \in S(x)$, then $W(S(x)) \subset \eta_\lambda^{-1}(0,1] \subset U_\lambda$, and therefore we have that that the restricted covering map $p|_{p^{-1}W(S(x))} : p^{-1}W(S(x)) \longrightarrow W(S(x))$ is trivial. Since for each n the open set W_n is a disjoint union of open sets of the form $W(S(x))$, it follows that $p|_{p^{-1}W_n} : p^{-1}W_n \longrightarrow W_n$ is trivial, as desired. $\square$

5.5.17 **Proposition**. *Every n-fold covering map $p : E \longrightarrow X$ over a paracompact space X has a Gauss map.*

Proof: Since X is paracompact, by Lemma 5.5.16 there is a countable trivializing open cover $\mathcal{W} = \{W_i \mid i \in \mathbb{N}\}$ of X. Let $\varphi_i : p^{-1}W_i \longrightarrow W_i \times \overline{n}$ be a trivialization and let $\{\eta_i\}$ be a partition of unity subordinate to $\mathcal{W}$. For each $i \in \mathbb{N}$ define $g_i : E \longrightarrow \mathbb{R}$ by

$$g_i(e) = \begin{cases} \eta_i(p(e)) \cdot \operatorname{proj} \varphi_i(e) & \text{if } e \in p^{-1}W_i\,, \\ 0 & \text{if } e \notin p^{-1}W_i\,, \end{cases}$$

where $\operatorname{proj}; W_i \times \overline{n} \longrightarrow \overline{n} \subset \mathbb{R}$ is the projection.

Now define $g : E \longrightarrow \mathbb{R}^\infty$ by $g(e) = (g_1(e), g_2(e), \dots, g_i(e), \dots)$. □

5.5.18 DEFINITION. If X is a paracompact space, then denote by $\mathcal{C}_n(X)$ the set of equivalence classes of n-fold covering maps over X.

By Propositions 5.5.14 and 5.5.17 we have the following result, which is the classification theorem of n-fold covering maps.

5.5.19 **Theorem**. *Let X be a paracompact space. Then there is a bijection*

$$[X, F_n(\mathbb{R}^\infty)/\Sigma_n] \longrightarrow \mathcal{C}_n(X)$$

given by $[f] \mapsto [f^* E_n(\mathbb{R}^\infty)]$.

Proof: By Theorem 5.5.9, the function is well defined. Propositions 5.5.14 and 5.5.17 show that the function is surjective.

To see that the function is injective, consider $\mathbb{R}^\infty_1 = \{(t_i) \in \mathbb{R}^\infty \mid t_{2j} = 0,\ j = 1, 2, 3, \dots\}$ and $\mathbb{R}^\infty_2 = \{(t_i) \in \mathbb{R}^\infty \mid t_{2j+1} = 0,\ j = 0, 1, 2, \dots\}$, so that $\mathbb{R}^\infty = \mathbb{R}^\infty_1 \oplus \mathbb{R}^\infty_2$. Now define homotopies $h^1, h^2 : \mathbb{R}^\infty \times I \longrightarrow \mathbb{R}^\infty$ by

$$\begin{aligned} h^1((t_1, t_2, t_3, \dots), t) &= (1-t)(t_1, t_2, t_3, \dots) + t((t_1, 0, t_2, 0, t_3, 0, \dots)\,, \\ h^2((t_1, t_2, t_3, \dots), t) &= (1-t)(t_1, t_2, t_3, \dots) + t((0, t_1, 0, t_2, 0, t_3, \dots)\,, \end{aligned}$$

where $(t_1, t_2, t_3, \dots) \in \mathbb{R}^\infty$ and $t \in I$. These homotopies start with the identity and end with maps which we denote by

$$h^1_1 : \mathbb{R}^\infty \longrightarrow \mathbb{R}^\infty_1 \subset \mathbb{R}^\infty \quad \text{and} \quad h^1_2 : \mathbb{R}^\infty \longrightarrow \mathbb{R}^\infty_2 \subset \mathbb{R}^\infty\,.$$

The composites $h^1_\nu \circ p_2 : E_n(\mathbb{R}^\infty) \longrightarrow \mathbb{R}^\infty$ for $\nu = 1, 2$ are Gauss maps, where $p_2 : E_n(\mathbb{R}^\infty) \longrightarrow \mathbb{R}^\infty$ is the projection on the second coordinate. By 5.5.15(a), these maps induce two morphisms of covering maps, namely,

$$\begin{tikzcd} E_n(\mathbb{R}^\infty) \arrow[r, "\widetilde{\varphi}_\nu"] \arrow[d, "\pi_n"'] & E_n(\mathbb{R}^\infty) \arrow[d, "\pi_n"] \\ F_n(\mathbb{R}^\infty)/\Sigma_n \arrow[r, "\varphi_\nu"'] & F_n(\mathbb{R}^\infty)/\Sigma_n\,, \end{tikzcd} \qquad \nu = 1, 2\,.$$

The composites $h^\nu \circ (p_2 \times \mathrm{id}) : E_n(\mathbb{R}^\infty) \times I \longrightarrow \mathbb{R}^\infty$ for $\nu = 1, 2$ are homotopies that start with p_2, since $h^\nu(q \times \mathrm{id})(e, 0) = h^\nu(p_2(e), 0) = p_2(e)$ for $e \in E_n(\mathbb{R}^\infty)$, and that end with $h_1^\nu \circ p_2$. Furthermore, the restrictions of these homotopies to the slices at each fixed $t \in I$ are Gauss maps. Using 5.5.15(b) we then have that φ_ν for $\nu = 1, 2$ is homotopic to the map induced by p_2, which is obviously the identity. Hence we have shown that $\varphi_\nu \simeq \mathrm{id}$ for $\nu = 1, 2$.

We are now in position to show that the function is injective. Assume that we are given $f_\nu : X \longrightarrow F_n(\mathbb{R}^\infty)/\Sigma_n$ for $\nu = 1, 2$ such that $f_1^* E_n(\mathbb{R}^\infty) \approx f_2^* E_n(\mathbb{R}^\infty)$. To prove injectivity we must show that f_1 and f_2 are homotopic.

Denote by E the space $f_1^* E_n(\mathbb{R}^\infty)$ and use the isomorphism above to obtain two morphisms of covering maps

$$\begin{tikzcd} E \arrow[r, "\widetilde{f}_\nu"] \arrow[d] & E_n(\mathbb{R}^\infty) \arrow[d] \\ X \arrow[r, "f_\nu"'] & F_n(\mathbb{R}^\infty)/\Sigma_n\,, \end{tikzcd} \qquad \nu = 1, 2\,.$$

Let $g_\nu : E \longrightarrow \mathbb{R}^\infty$ for $\nu = 1, 2$ be the associated Gauss map, namely $g_\nu = p_2 \circ \widetilde{f}_\nu$.

Consider the composites $h^\nu \circ g_\nu : E \longrightarrow \mathbb{R}^\infty$ for $\nu = 1, 2$. They are Gauss maps, and according to 5.5.15(a), they induce morphisms of covering maps of the form

$$\begin{tikzcd} E \arrow[r, "\widetilde{f}_\nu"] \arrow[d] & E_n(\mathbb{R}^\infty) \arrow[r, "\widetilde{\varphi}_n u"] \arrow[d] & E_n(\mathbb{R}^\infty) \arrow[d] \\ X \arrow[r, "f_\nu"'] & F_n(\mathbb{R}^\infty)/\Sigma_n \arrow[r, "\varphi_\nu"'] & F_n(\mathbb{R}^\infty)/\Sigma_n\,, \end{tikzcd} \qquad \nu = 1, 2\,.$$

We then define $G : E \times I \longrightarrow \mathbb{R}^\infty$ by $G(e, t) = (1 - t)h_1^1 g_1(e) + t h_1^2 g_2(e)$ for $(e, t) \in E \times I$. This is a homotopy between $h_1^1 \circ g_1$ and $h_1^2 \circ g_2$, and since $h_1^1(\mathbb{R}^\infty)$ and $h_1^2(\mathbb{R}^\infty)$ have no points in common with the exception of 0, it follows that G_t is a Gauss map for each $t \in I$. Therefore, using 5.5.15(b) we have that $\varphi_1 \circ f_1 \simeq \varphi_2 \circ f_2$. But we already know that $\varphi_\nu \simeq \mathrm{id}$ for $\nu = 1, 2$. So $f_1 \simeq f_2$ as desired. □

Chapter 6 Knots and links

In this chapter we shall introduce a very interesting branch of topology, namely the branch called knot theory. It handles topological objects, which in some sense are as simple, as 1-manifolds are. But they are not studied as topological spaces, but it rather studies the very different ways in which they are embedded as subspaces of $\mathbb{R}^3$. Knot theory has had its own development in the sense that many of its methods have been specially designed for studying these embeddings. However it uses many aspects of algebraic topology. Its is worth noticing that knots have found astonishing applications in other branches of mathematics, as well as in other disciplines like physics, chemistry and biology (see [10, Cap.4], [20]).

6.1 1-manifolds and knots

In Chapter 2, Section 2.3 we proved that the only connected manifolds of dimension 1 are, up to homeomorphism, the circle $\mathbb{S}^1$ and the real line $\mathbb{R}$. This last hat little topological interest; on the other hand, the circle is richer, as we already saw in 3.2. Notwithstanding, it is interesting to study not only the circle as a topological space, but how one can embed it into $\mathbb{R}^3$.

It is important to notice that the circle cannot be embedded into $\mathbb{R}$, as the Borsuk-Ulam theorem in dimension 1 shows (see 3.2.39). On the other hand, the Jordan curve theorem (see 3.2.29) implies that the only embedding of the circle into $\mathbb{R}^2$ up to homeomorphism is the canonical one, namely, for any embedding there is a homeomorphism of $\mathbb{R}^2$ onto itself that sends the the given embedding to the canonical inclusion (*exercise*). For $\mathbb{R}^n$ with $n \geq 4$ the same happens, namely, for any embedding of the circle into $\mathbb{R}^n$ there is a homeomorphism of $\mathbb{R}^n$ onto itself that sends the given embedding to the canonical embedding into the plane determined by the first two coordinates, as one can verify using arguments of differential topology (transversality). Therefore the only interesting embeddings are those of the circle into $\mathbb{R}^3$.

The next is a first provisional definition of the concept of a knot. Below we shall describe the concept of a tame knot, which will be our object of study in what follows.

6.1.1 Definition. A *knot* is a subspace K of $\mathbb{R}^3$ which is homeomorphic to $\mathbb{S}^1$.

Figure 6.1 shows five knots, which carry a special name. The necessity to draw them on the plane forces us to do it using their *regular projections*, which codify somehow the embedding of the knot into the space. The (connected) portion of the diagram that goes from one undercrossing to the next will be called *arc* or *strand*

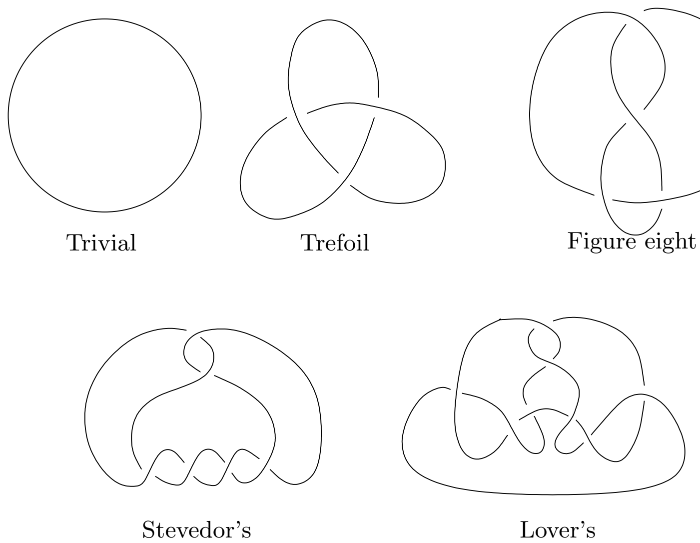


Figure 6.1 Knots

The first of them, the *trivial knot* or "*unknot*", is represented by the unit circle in the plane. If the reader ties each of these five knots with a rope, whose ends are joined thereafter, she/he will be convinced that it is impossible to convert one to each other, unless we could let the rope cross itself or, of course, if we loosen the ends. The next definition might be the natural way of stating the equivalence of knots.

6.1.2 Definition. Two knots K_0 and K_1 are *isomorphic* if there is a homeomorphism $\varphi : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ such that $\varphi(K_0) = K_1$.

For instance, Figure 6.2 represents a trivial knot, that is, a knot which is isomorphic to circle in the plane $\mathbb{R}^2 \subset \mathbb{R}^3$, even though it does not seem plausible at first sight. This means that playing around with the rope adequately, we may unknot it. This is equivalent to saying that there is a homeomorphism of the space that maps this knot onto the circle.

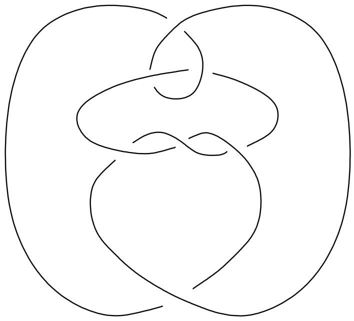
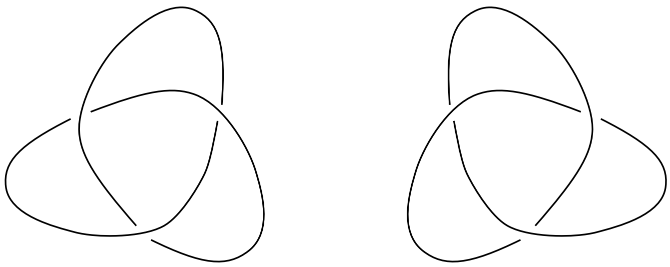


Figure 6.2 Nasty trivial knot

However, the definition includes among all possible homeomorphisms of the space a reflection on a plane, which would transform the trefoil knot into its mirror image. Both are shown in Figure 6.3. But if we tray to play around with the rope and try to transform one into the other, we shall never be able to do it.



Left-handed trefoil knot Right-handed trefoil knot

Figure 6.3 Mirror knots

If we look for a definition which involves the sliding of portions of the knot to convert it to another, we might have difficulties, since if we "tighten" the knot enough, we would obtain the trivial knot, as Figure 6.4 suggests.

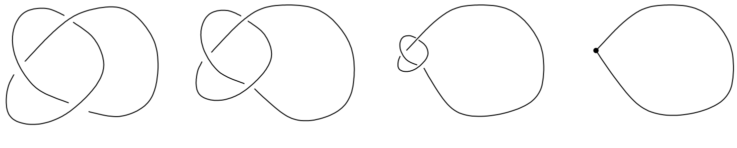


Figure 6.4 Forbidden deformation

To avoid this situation, the deformation of the knot must be such that it deforms neighboring points of the Euclidean space without collapsing them.

Recall that a homeomorphism $\varphi : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ is *isotopic* to the identity if there is an *isotopy* $H : f \simeq g$, namely a homotopy such that (i) for each $t \in I$, $H_t : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ is a homeomorphism, and (ii) $H_0 = \varphi$ and $H_1 = \mathrm{id}_{\mathbb{R}^3}$. If $K \subset \mathbb{R}^3$ is a knot, then $K_t = H_{1-t}(K) \subset \mathbb{R}^3$ is a knot for each t. Moreover $K_0 = K$ and as t varies from 0 to 1 K deforms inside $\mathbb{R}^3$. We say that such a homeomorphism φ *preserves the orientation.*

6.1.3 Definition. Two knots K_0 and K_1 are *equivalent* if there is an orientation preserving homeomorphism $\varphi : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$.

We shall study here the so-called *tame* knots, namely knots which are equivalent to *polygonal* knots,, i.e., knots which are built up by a finite number of straight line segments. The opposite to tame are the *wild* knots, an example of which is a knot that has an infinite sequence of knottings, which are smaller and smaller as shown in Figure 6.5.

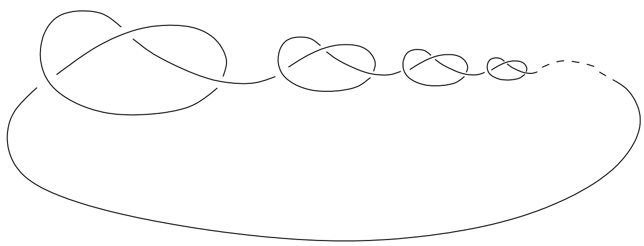


Figure 6.5 Wild knot

The figures of knots that we have shown, as we already said, imply a certain code to be represented with no ambiguity in the plane. This allows us to work with them adequately. These drawings represent a projection of the knot into the plane, in such a way that (i) it has only finitely many crossings, (ii) there are at most two portions of the knot meet at each crossing, and (iii) the crossings are transverse, namely, they build a positive angle. Such a projection is called *regular.*

The following result is true, as one easily sees intuitively. The proof involves small deformations of the knot, which correct possible defects in their projection, but we shall omit it.

6.1.4 **Theorem**. *Every tame knot is equivalent to one that has a regular projection on the plane.* □

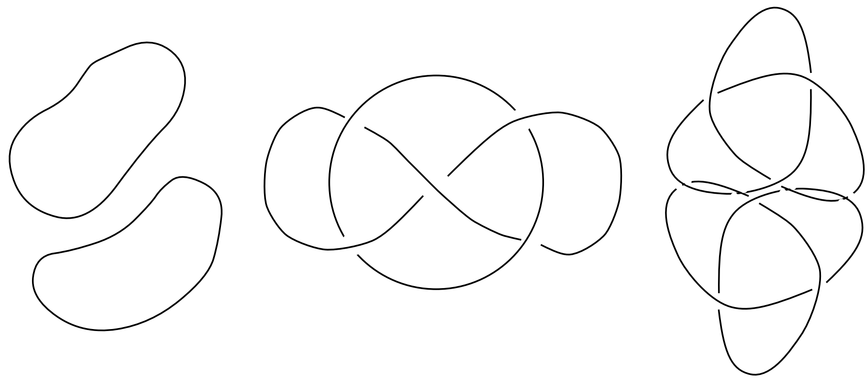
In what follows, every time we refer to a knot we shall always mean a tame knot. We shall see also that it is possible and convenient to generalize the concept of knot.

6.1.5 Definition. A *link* is a disjoint union of a finite family of knots in $\mathbb{R}^3$. That is, knots that might be intertwined, but they do not intersect.

Similarly to a knot, a link has *regular projection*, in which at most two components cross transversely. Corresponding to 6.1.4, we have the following.

6.1.6 **Theorem**. *Every link is equivalent to one which has a regular projection on the plane.* □

Figure 6.6 shows (regular projections of) several links.



(a) (b) (c)

Figure 6.6 Links

6.2 Reidemeister moves

The regular projections of equivalent knots may be quite different, as we already saw in Figure 6.2, which shows a projection which looks quite differently from the canonical projection of the trivial knot, or as one can see in Figure 6.7, which shows two different projections of the figure eight knot.

In this section we shall see how to play around with the regular projections of a knot or a link, in order to compare them. We shall make changes around one, two, or three crossings, each of a particular configuration, without modifying the rest of the projection diagram.

6.2.1 Definition. The *Reidemeister moves* in the regular projection diagram of a knot or link are described as follows.

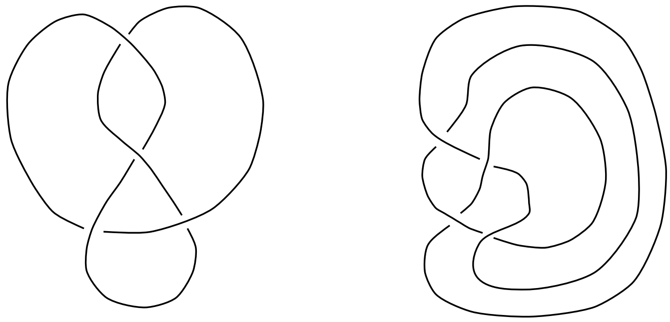


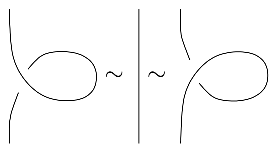
Figure 6.7 Dos proyecciones del knot de la figure eight

Type I. Twist and untwist the strand in either direction.

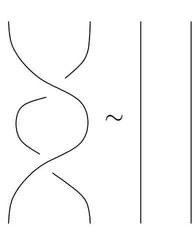
Type II. Move one loop completely over another.

Type III. Move a strand completely over or under a crossing.

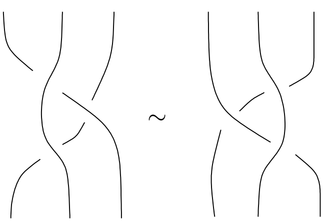
They are shown in Figure 6.8.



Type I



Type II



Type III

Figure 6.8 Reidemeister moves

Two regular projections are said to be *equivalent* if it is possible to pass from one to the other after applying a finite sequence of Reidemeister moves of types I, II, and III and an orientation preserving homeomorphism of the plane (namely, a homeomorphism isotopic to the identity).

The following theorem is due to Reidemeister. We shall omit its proof, but we refer the reader to [1] for a proof.

6.2.2 **Theorem**. *Two knots or links are equivalent if and only if all its regular projections are equivalent.* □

6.2.3 Example. The projections K_1 and K_2 shown in Figure 6.9 seem two represent to rather different knots. indeed, during a good part of the twentieth century everybody thought that they were different. It was not until 1970 that the north American lawyer K. A. Perko proved that it is possible to transform K_1 into K_2 after quite a large number of Reidemeister moves. This pair of diagrams is known as *Perko pair knots* (see Figure 6.9).

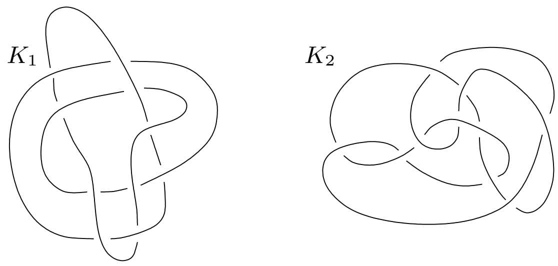


Figure 6.9 Perko pair knots

6.2.4 Note. The number of Reidemeister moves needed to transform a projection of a knot into another, can be in general rather huge. In [13] it is proven that there is a positive constant c such that $2^{c\cdot n}$ Reidemeister moves are needed to transform a a given regular projection of the trivial knot, like for example the one shown in Figure 6.2, into a circle, where n is the number of crossings of the given projection. The order of magnitude of c is 10^{11}. So, to have an idea of this number, let us recall that the age of the universe is around $5 \cdot 10^{12}$ days or less than $5 \cdot 10^{17}$ seconds.* According to this estimate, the number of Reidemeister moves needed to unknot a regular projection of the trivial knot with 7 crossings would have the order of $2^{10^{11}\cdot 7}$ or approximately $10^{10^{10}\cdot 21}$, that is, very many orders of magnitude the age of the universe in seconds (or even in nanoseconds).

The Reidemeister theorem 6.2.2 transforms the problem of studying knots and links to the problem of analyzing their regular projections. From here on, we shall not distinguish between knot or a link and its regular projection, and we shall work exclusively with regular projections.

*The age of the universe is estimated nowadays in around $13.7 \cdot 10^9$ years.

6.3 Knots and colors

In some sense, knots are more complicated than surfaces. The problem of distinguishing among them involves the way they are embedded in space. It was not until the twenties that one could prove the existence of nontrivial knots, namely that the knots really exist. Obviously, it was somehow clear that the trefoil knot was not trivial, but there was no proof for that. Even the use of Reidemeister moves is not a simple business for calculation, as we already said in Note 6.2.4. The first person to prove that the trefoil knot is not trivial was presumably Reidemeister. He did it as follows. He discovered a property which is not at all obvious, whose formulation indicates the possibility of coloring each strand (arc) of a regular projection of a knot with one of three different colors, in such a way that at least two of the three colors are used, and that at each crossing either all three colors are used for all three arcs that come together, or just one color for all of them. In other words, in no crossing may one use two colors only. For instance, if the trefoil knot, in its simplest representation, as in Figure 6.1, which consists of three arcs, each arc is colored with one of the three different colors, then the rules are fulfilled. Then we say that the trefoil knot is *tricolorable*. Reidemeister proved that this property is invariant under the Reidemeister moves (see Theorem 6.3.6). However, the circle, since it has no crossings, does not have this property, so it cannot be equivalent to the trefoil knot.

With the goal to face with more generality the question of distinguish different knots, we shall introduce a procedure that generalizes the tricolorability. It was introduced by R. Fox around 1960. We shall call it the *color game* and the rules are as follows.

Take a wheel with an odd number of radii, which are uniformly distributed. To each radius assign a label (which we can call “color”). For instance we can label the radii by

$$r = \text{red}\,, \quad g = \text{green}\,, \quad y = \text{yellow}\,, \quad p = \text{purple}\,, \quad \text{and} \quad b = \text{blue}\,,$$

as shown in Figure 6.10 for the case $n = 5$.

6.3.1 Definition. We define the *color game* as follows. Using the n different colors of the wheel, we shall paint (label) each of the arcs of the knot projection according to the following rules:

(a) One must use at least two colors.

(b) At every crossing, either all arcs have the same color, or the color of the overcrossing arc must be the average color (or the bisectrix) with respect to the

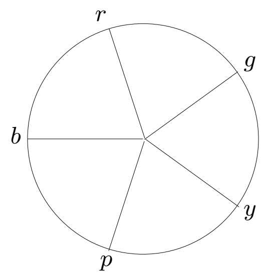


Figure 6.10 Color wheel for $n = 5$

color wheel, of the two colors of the arcs corresponding to the undercrossing. (Since n is odd, the two colors that correspond to the undercrossing arcs determine uniquely the color of the overcrossing arc.)

A coloring that fulfills the rules (a) and (b) will be called *n-admissible coloring*. These rules are illustrated in Figure 6.11.

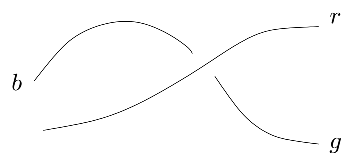


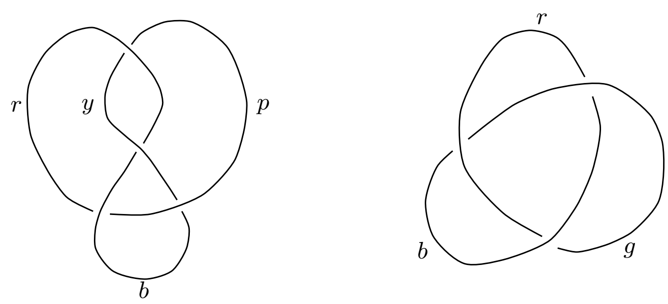
Figure 6.11 Admissible coloring

It is worth noticing again that the use of an odd number of colors guarantees that *given any two colors, there is a unique color which corresponds to the angle bisectrix*, (if n were even, ten there might be either two bisectrices or none).

6.3.2 Example. The *figure eight knot*, K_8, shown in Figure 6.12(a) is colored using the five colors of the wheel shown in Figure 6.11, so that the game rules are fulfilled. However it is impossible to color the *trefoil knot* T of Figure 6.12(b) with five colors according to the game rules.

On the other hand, if instead of using the five-color wheel, we use the three-color wheel, as in Figure 6.13, then it is possible to color the trefoil knot, as we saw already at the start of this section. We can easily check that it is impossible to color the figure eight knot using three colors.

6.3.3 Definition. Given a regular projection of a knot, we define its *chromatic number* to be the minimal odd number n greater than 1 such that the projection



(a) (b)

Figure 6.12 Colorings

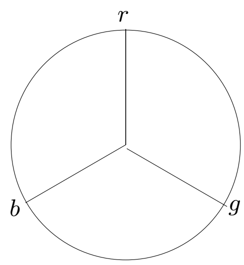


Figure 6.13 Color wheel for $n = 3$

admits an n-coloring, according to the rules 6.3.1(a) and (b). If the projection does not admit any coloring, then we define the chromatic number to be 1 (like, for instance, any projection of the trivial knot).

6.3.4 EXERCISE. Show that a regular projection of a knot is n-colorable if and only if its arcs can be labeled with the integers $0, 1, 2, \ldots, n-1$, such that the following rules are fulfilled:

(a) At least two numbers are used.

(b) At a given crossing, if the overcrossing arc has the label k and the two arcs of the undercrossing have the labels l and m, then $n|2k - l - m$. In other words, $2k \equiv l + m\,(n)$.

(*Hint*: Notice that in order that the congruences have a unique solution, n must be odd. Notice too, that the cyclic group with n elements can be realized as a subgroup of $\mathbb{S}^1 \subset \mathbb{C}$, by taking the nth roots of unity.)

6.3.5 EXERCISE. Take the projection of the *Ochiai knot* given in Figure 6.14. Show that its chromatic number is 1, namely, it does not admit any n-coloring for all $n > 1$. (*Hint*: Use the description of the previous exercise and solve the congruences as diophantine equations, i.e. equations in $\mathbb{Z}$.)

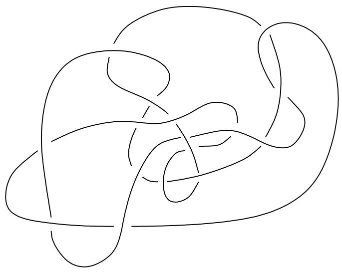


Figure 6.14 knot de Ochiai

The use of the word *invariant* for the chromatic number is justified by the following result, which allows to apply it not only to a knot projection, but to the knot itself.

6.3.6 **Theorem**. *If two regular projections are equivalent, then they have the same chromatic number.*

Proof: We must show that if some projection admits an n-coloring for some n, then any modification of it using any one of the Reidemeister moves still admits an n-coloring.

For the **type I move**, we have that if a certain coloring around a crossing is admissible, as seen in Figure 6.15, then $a = b$, so that all three colors are the same, say a.

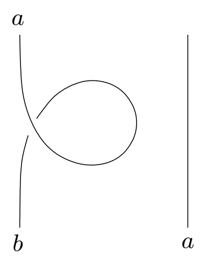


Figure 6.15 Admissible coloring and the type I move

For the **type II move**, notice that if we assume that in Figure 6.16 the coloring on the left with colors a, b, c and d, is admissible, then a is the bisectrix of b and c and it is also a bisectrix of c and d. This is only possible if $b = d$. Therefore the coloring on the right must be admissible.

For the **type III move** we have that if the coloring given on the left of Figure 6.17 using the colors a, b, c, d, e and f, is admissible, then on the right the colors a, b, c, d and e must be correct according to the rules. One must only check if it is

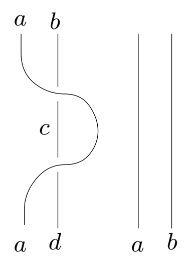


Figure 6.16 Admissible coloring and the type II move

possible to find in the color wheel a color x, such that the coloring is admissible. But it is indeed always possible to take color x in such a way that b is the bisectrix of the angle formed by c and x. Now it is routine to verify that a is the bisectrix of the angle formed by d and x.

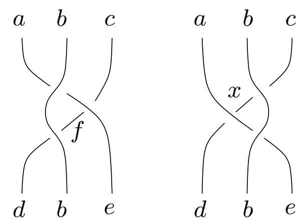


Figure 6.17 Admissible coloring and the type II move

□

6.3.7 EXERCISE. Restate the previous proof using the description of the n-coloring given in Exercise 6.3.4.

By the previous theorem we have that the number of colors required for coloring certain knot projection is an *invariant* of the knot. Namely, it only depends on the knot as such, and not on any of its regular projections. We shall call this invariant, namely, the minimal number of colors required for coloring one of its regular projections, or 1, the *chromatic number* of the knot.

6.3.8 EXAMPLE. The figure eight knot has chromatic number 5, and not 3. The trefoil knot has chromatic number 3, and not 5. From this, we conclude that *the figure eight knot and the trefoil knot are not equivalent.* It is impossible to modify one, without cutting and pasting, to obtain the other.

Figure 6.18 shows the trefoil knot and its mirror image. This poses the question, if this two knots are equivalent. Maybe the daily experience allows us to say that they are not, but our color game is not enough to tell the two knots apart. Namely,

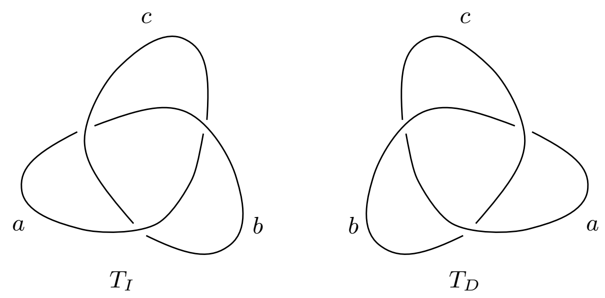


Figure 6.18 Admissible colorings for the left- and the right-handed trefoil knots

if one of the trefoil knots is admissibly colored with the three colors a, b, c, then when we put it in front of a mirror, we obtain automatically an admissible coloring for the other: hence both have the same chromatic number equal to 3. In fact, we have the next result.

6.3.9 **Proposition**. *If the chromatic number of a knot is n, then the chromatic number of its mirror image is also n.* □

6.3.10 Exercise. Compute the chromatic number of each of the knots shown in Figure 6.1.

6.4 Knots, links, and polynomials

The chromatic number introduced in the previous section does not distinguish between a knot and its mirror image. A knot is said to be an *amphicheiral knot* if it is equivalent to its mirror image. Are all knots amphicheiral knots?

We shall now introduce a finer invariant, than the chromatic number, which will be a polynomial ([17] or [18]).†

6.4.1 Definition. The *Kauffman bracket* assigns to a projection of a knot or link K a polynomial in the indeterminates x, y and d with integral coefficients, denoted by $[K] \in \mathbb{Z}[x, y, d]$, according to the following recurrence rules:

(a) $\left[\rangle\!\!\langle \right] = x\left[\asymp \right] + y\left[\rangle\ \langle \right]$

This rule states that we can replace the bracket of the original projection as follows: if we are walking in an "overcrossing" of a knot projection, then we can go left and multiply the resulting bracket by x, and we can go right

†I am grateful to Michael Barot and to Francisco González Acuña for their valuable comments regarding this section.

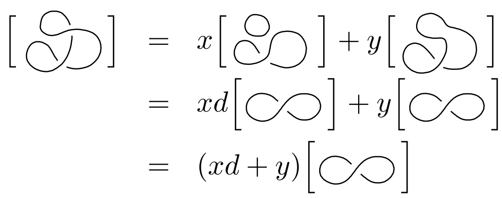
and multiply the resulting bracket by y, and then add both results. This way we reduce the computation of the bracket of some knot projection to the computation of the brackets of two projections each with one crossing less.

Example:

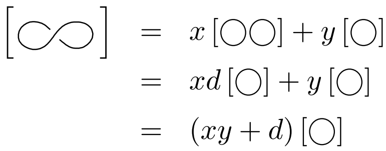
(b) $[K \sqcup \bigcirc] = d[K]$.

This rule establishes that if a link has an unlinked component, which is a trivial knot, then we can eliminate it by multiplying the resulting bracket by d.

Example:

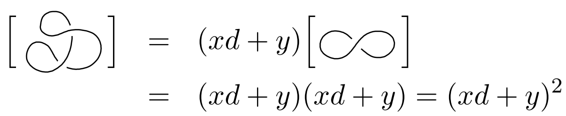


and



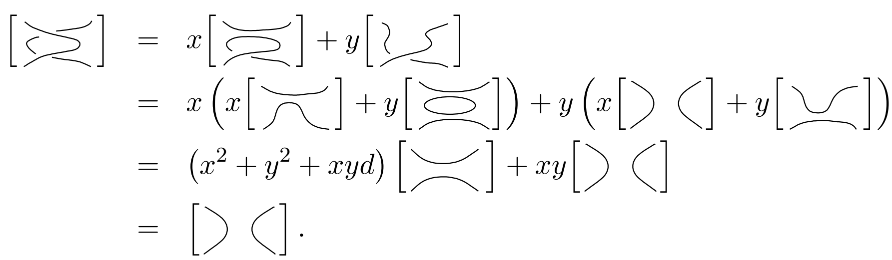
(c) $[\bigcirc] = 1$.

Finally, this rule states that the bracket of the projection of the trivial knot is the trivial polynomial 1. Thus we have that



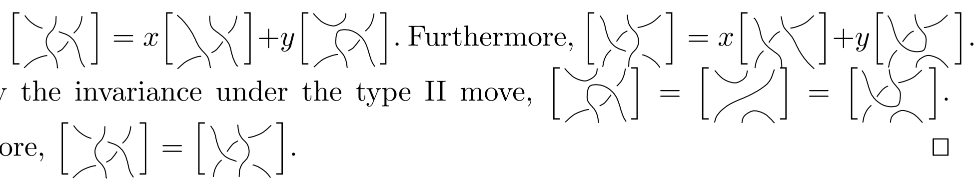
6.4.2 **Lemma**. *The Kauffman bracket* [ ] *is invariant under the type* II *Reidemeister move if and only if* $xy = 1$ *and* $d = -x^2 - x^{-2}$.

Proof: Consider the following series of equalities:

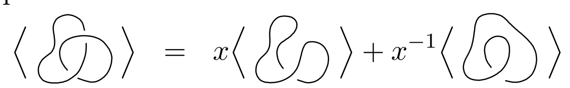


The last of them holds if and only if $xy = 1$, namely if and only if $y = x^{-1}$ and $x^2 + y^2 + xyd = 0$, i.e., if $d = -x^2 - x^{-2}$. $\square$

6.4.3 **Corollary**. *If* $xy = 1$ *and* $d = -x^2 - x^{-2}$*, then the Kauffman bracket* $[\quad]$ *is invariant under the type* III *Reidemeister move .*

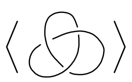
Proof:  Furthermore, But by the invariance under the type II move, Therefore, $\square$

Denote by $\langle K\rangle \in \mathbb{Z}[x, x^{-1}]$ the Kauffman bracket $[K]$ for the case $xy = 1$ and $d = -x^2 - x^{-2}$ and call it the *fine Kauffman bracket*. Hence $\langle K\rangle$ is a *Laurent* polynomial, namely a polynomial in positive and negative powers of x. Now we can finish the computation of the bracket of the trefoil knot. We still have to compute



$$\begin{aligned} &= x\left(-x^3\right) + x^{-1}\left(x\langle\bigcirc\rangle + x^{-1}\langle\bigcirc\bigcirc\rangle\right) \\ &= -x^4 + 1 + x^{-2}\left(-x^2 - x^{-2}\right) \\ &= -x^4 - x^{-4}\,. \end{aligned}$$

Hence we have for the *left-handed* trefoil knot that



$$\begin{aligned} &= x\left(x^6\right) + x^{-1}\left(-x^4 - x^{-4}\right) \\ &= -x^{-5} - x^3 + x^7\,. \end{aligned}$$

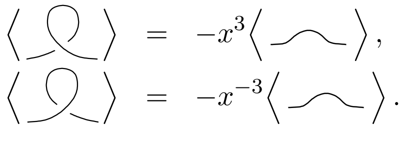
If we take the mirror image of the projection of a knot, the only thing that changes is the "orientation", namely "left" changes to "right" and viceversa. Thus rule 6.4.1(a) implies that in the bracket of the mirror image, the powers of x are exchanged with those of y. But since we convene that $y = x^{-1}$, then the powers of x in the fine bracket of the mirror image have opposite signs to those of the fine bracket of the original projection. Therefore, for the projection of the *right-handed* trefoil knot we have that

$$= x^{-7} - x^{-3} - x^5\,.$$

Consequently, both projections have different polynomials.

As we have already seen, if one modifies the projection diagram K of a knot using Reidemeister moves of the types II or III, then its polynomial $\langle K\rangle$ does not change. However, the type I move does alter $\langle K\rangle$. Namely,

(6.4.4)



Hence the fine Kauffman bracket is not an invariant of knots nor links. In order to correct this misbehavior, we need the following.

6.4.5 DEFINITION. Given a knot or link projection diagram K, we define the *wreath* $w(K)$ of a knot or link projection as follows. To each knot-component of the link, assign an orientation and count how many positive crossings and how many negative crossings there are in the whole diagram, according to Figure 6.19.

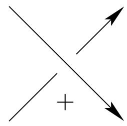


Figure 6.19 Oriented crossings

Then $w(K)$ is the number of positive crossings minus the number of negative crossings. Observe that in the case of a knot, if one takes the opposite orientation, then the both directions in a crossing are changed. Hence the orientation of the crossing remains the same, and thus *if K is the projection diagram of a knot, then the wreath $w(K)$ does not depend on the orientation chosen.*

6.4.6 EXERCISE. To each knot in Figure 6.1 assign an orientation and compute the wreath in each case.

6.4.7 EXERCISE. Put an orientation on each component of the Whitehead link and of the Borromean rings shown in Figure 6.20, and compute the corresponding wreaths.

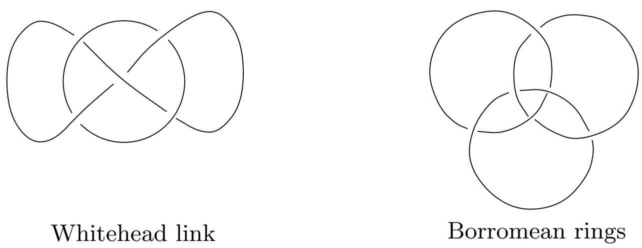


Figure 6.20 Links

It is an easy *exercise* to prove the following statement.

6.4.8 **Proposition**. *The wreath of the projection diagram of a knot is invariant under the Reidemeister moves* II *and* III. □

As we already said, it is a simple *exercise* to prove the next.

6.4.9 **Proposition**. *The wreath of the projection diagram of a knot is independent of the given orientation.* □

6.4.10 EXAMPLE. As one may appreciate in Figure 6.21, the left-handed trefoil knot T_L has wreath $+3$ and the right-handed trefoil knot T_R, has wreath -3.

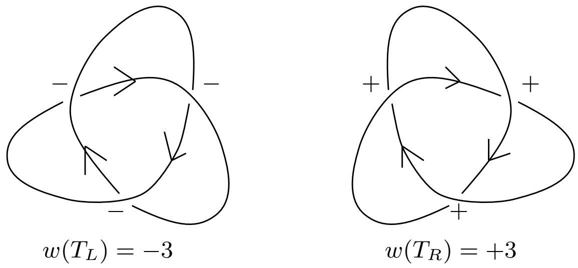


Figure 6.21 The wreath of the left- and right-handed trefoil knots

6.4.11 DEFINITION. Given a projection diagram of a link K, there is a polynomial defined by

$$f_K(x) = (-x^3)^{-w(K)}\langle K\rangle\,.$$

6.4.12 **Theorem**. *The polynomial $f_K(x)$ of the projection diagram K of a knot or a link, is invariant under the type* I, II, *or* III *Reidemeister move. Therefore it is a knot or link invariant.*

Proof: By (6.4.4), getting rid of a loop in K corresponds to multiplying $\langle K\rangle$ by x^3 or x^{-3}, according to the orientation of the loop. Since this modification of the diagram reduces the wreath in 1 or -1, the final effect in the polynomial $\langle K\rangle$ is that it remains invariant under a type I Reidemeister move. By 6.4.8, f_K es invariante under type II or III Reidemeister moves. Hence it remains invariant under all three moves and therefore, it is an invariant of the knot or link itself. □

In particular, for the left-handed trefoil knot, we have

$$f_{T_L}(x) = (-x^3)^3(-x^{-5} - x^3 + x^7) = x^4 + x^{12} - x^{16}\,;$$

while fort the right-handed trefoil knot, we have

$$f_{T_R}(x) = (-x^3)^{-3}(x^{-7} - x^{-3} - x^5) = -x^{-16} + x^{-12} + x^{-4}\,.$$

Therefore, we have shown the following.

6.4.13 **Proposition**. *The left-handed trefoil knot T_L and the right-handed trefoil knot T_R are not equivalent, namely, the trefoil knots are not amphicheiral.* They are *cheiral* (or *chiral*). □

It can be shown, that for any knot K the polynomial $f_K(x)$ has powers that are always multiple of 4. Therefore, we can simplify it by replacing x by $t^{-1/4}$. This way, we obtain the so-called *Jones polynomial* of K, namely $V_K(t) = f_K(t^{-1/4})$, which now is a polynomial in powers of the indeterminate t, (see [16]). In particular, for the trefoil knots we have

$$\begin{aligned} V_{T_L}(t) &= -t^{-4} + t^{-3} + t^{-1}\,, \\ V_{T_R}(t) &= t + t^3 - t^4\,. \end{aligned}$$

One can show that the value m obtained by evaluating the Jones polynomial, when $t = -1$ and defining $m = |V_K(-1)|$, is divisible by the chromatic number n of the given knot K, which we introduced in Section 6.3. In the case of the trefoil knots, $|V_{T_I}(-1)| = 3 = |V_{T_D}(-1)|$. It is an interesting *exercise* to compute the Jones polynomial of the figure eight knot K_8, to obtain

$$V_{K_8}(t) = t^{-2} - t^{-1} + 1 - t^1 + t^2\,.$$

One can now verify that $|V_{K_8}(-1)| = 5$, which is the chromatic number of K_8.

6.4.14 Exercise. Compute the Jones polynomials of the stevedor's and the true lovers' knots shown in Figure 6.1, as well as of the Ochiai knot shown in Figure 6.14. In each of the three cases, compute also $n = |V_K(-1)|$ and confirm that n is odd and that one can color the corresponding knot with n colors, according to the rules 6.3.1.

6.4.15 Definition. Given the regular projections of two knots K and K', one can define a new removing a small portion of an arc in each of the projections and then gluing the ends of the remaining diagrams, as shown in Figure 6.22. The new knot is called the *connected sum* of K and K' and is denoted by $K\#K'$. Of course, in order to realize this operation, we have to assume that the original diagrams do not overlap, and that the arcs chosen to be removed are "outside". If a knot is the connected sum of two other nontrivial knots, then the knot is said to be a *composed knot.* The connected sum of a knot K with a trivial knot is again the same knot K. If a knot is not composed, that is, it is not the connected sum of any two nontrivial knots, then it is said to be a *prime knot.*

Notice that the definition of the connected sum of knots is in some sense a special case of the connected sum of manifolds 2.1.32, given in Chapter 2.

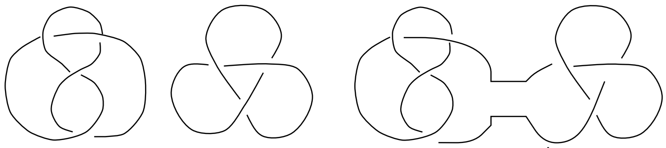


Figure 6.22 Connected sum of knots

6.4.16 Exercise.

(a) Show that the fine Kauffman bracket of a connected sum of two (regular diagrams of) knots is the product of their brackets of each summand, namely

$$\langle K\#K'\rangle = \langle K\rangle\langle K'\rangle\,.$$

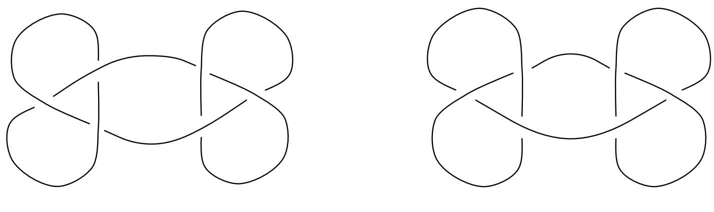
(b) Show that the wreath of a connected sum of two (regular diagrams of) knots is the sum of their wreaths, namely

$$w(K\#K') = w(K) + w(K')\,.$$

(c) Conclude that the Jones polynomial of a connected sum of two (regular diagrams of) knots is the product of their Jones polynomials, namely

$$V_{K\#K'}(t) = V_K(t)V_{K'}(t)\,.$$

(d) Making use of (c), compute the Jones polynomials of granny's knot and of the square knot, which are depicted in Figure 6.23.



Granny's knot Square knot

Figure 6.23 Connected sums of trefoil knots

6.4.17 Exercise.

(a) Show that if two knots K and K' have the same chromatic number n, then el chromatic number of their connected sum is also n.

(b) More generally, show that if $n(K)$ denotes the chromatic number of a knot K, then one has

$$n(K\#K') = \min\{n(K), n(K')\}\,,$$

where min represents the minimum of both numbers.

6.4.18 Exercise. Making use of the previous exercises, compute the chromatic numbers n and n' of granny's knot K and of the square knot K', and verify that these numbers do not coincide with $m = |V_K(-1)|$ nor $m' = |V_{K'}(-1)|$. Check that, however, $n|m$ and $n'|m'$.

6.4.19 Note. The book [1] contains a table of polynomials of all prime knots with no more than nine crossings.

6.5 The knot group

Given two equivalent knots K_1 and K_2, notice that there is a homeomorphism $\varphi : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$, which preserves the orientation and is such that $\varphi(K_1) = \varphi(K_2)$. Hence, by restriction, we obtain a homeomorphism $\mathbb{R}^3 - K_1 \approx \mathbb{R}^3 - K_2$. Thus we have proved the following.

6.5.1 **Proposition**. *Equivalent knots have homeomorphic complements.* □

Once more we are facing the homeomorphism problem. The previous proposition implies that if two knots have nonhomeomorphic complements, then they are not equivalent knots. Therefore, each invariant which allows us to distinguish topological spaces, up to homeomorphism, will in principle be useful to distinguish between knots.

6.5.2 Note. Notice that a knot and its mirror image have homeomorphic complements, since some reflection of $\mathbb{R}^3$ is a homeomorphism that maps any knot onto its mirror image. As we mentioned in the Introduction, a few years ago Gordon and Luecke [12] proved the converse of Proposition 6.5.1. More precisely, they showed that *two knots, or one of them and the mirror image of the other, are equivalent if and only if their complements are homeomorphic.* This way, the equivalence problem of knots is equivalent to the homeomorphism problem in the case of certain 3-dimensional submanifolds of $\mathbb{R}^3$.

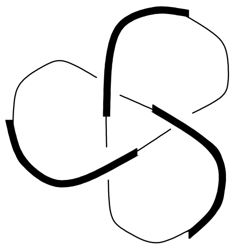
A good invariant to distinguish between two knots is the fundamental group of their complements.

6.5.3 Definition. Let $K \subset \mathbb{R}^3$ be a knot. The fundamental group $\pi_1(\mathbb{R}^3 - K)$ is called *knot group* of K and is denoted by $\pi(K)$.

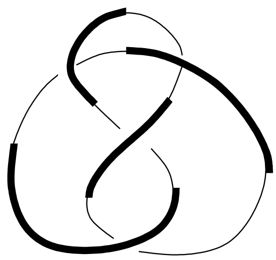
Hence, the following holds immediately.

6.5.4 **Proposition**. *The knot group $\pi(K)$ is an invariant of the equivalence class of K, that is, if K and K' are equivalent knots, then their knot groups $\pi(K)$ and $\pi(K')$ are isomorphic.* □

In what follows, we shall look for a reasonable way of presenting the knot group in terms of generators and relations, depending on the diagram of the knot in question. To do this, take a knot K inside the upper halfspace of $\mathbb{R}3$ $(x_3 > 0)$, so that its vertical projection into the plane $x_3 = 0$ is regular. Now break apart the projection into portions corresponding to "overcrossings' or to "undercrossings" in such a way that they alternate, as shown in Figure 6.24 for the case of the trefoil and the figure eight knots.



K is the trefoil knot



K is the figure eight knot

Figure 6.24 Overcrossings and undercrossings of a knot K

We now replace each undercrossing by its projection into the plane $x_3 = 0$ and we join it with the adjacent overcrossings by vertical line segments in such a way that we obtain a new knot, denoted again by K, which is obviously equivalent to the original knot, as shown in Figure 6.25.

We can now proceed to compute the knot group of K by decomposing $\mathbb{R}^3 - K$ in several pieces with recognizable fundamental groups and applying the Seifert–van Kampen theorem in each step.

First, we compute $\pi_1(\mathbb{R}^3_+ - K)$, where R^3_+ denotes the closed upper halfspace, $x_3 \geq 0$. Take a base point x_0 with a large enough height (x_3) so that it is far above the knot. Furthermore, give the knot an orientation. For each overcrossing take a loop based at x_0 which goes around the the arc in clockwise sense, as shown in Figure 6.26.

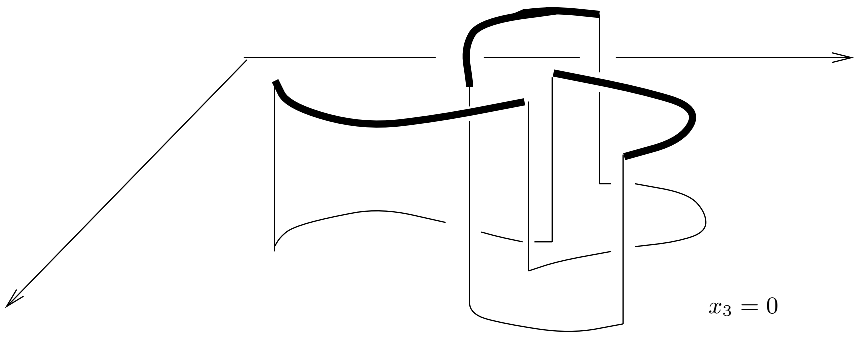


Figure 6.25 Modified knot K

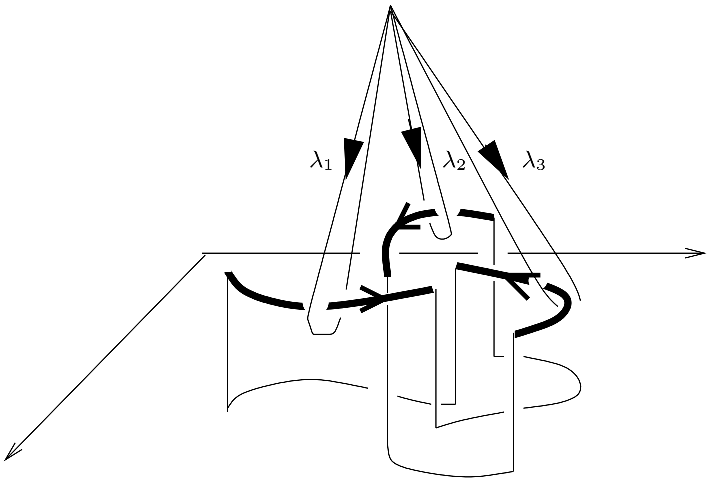


Figure 6.26 Generators of the knot group

Let us call these loops $\lambda_1, \lambda_2, \ldots, \lambda_k$, and let

$$\alpha_1, \alpha_2, \ldots, \alpha_k \in \pi_1(\mathbb{R}^3_+ - K)$$

be the elements determined by them. We obtain the following result.

6.5.5 **Lemma**. *The group $\pi_1(\mathbb{R}^3_+ - K)$ is freely generated by the elements $\alpha_1, \alpha_2, \ldots, \alpha_k$.*

Proof: Denote by $\widehat{K}$ the subspace of K built up by the overcrossings and the vertical segments of K, whose lower ends lie in the plane $x_3 = 0$. Therefore the inclusion $\mathbb{R}^3_+ - K \hookrightarrow \mathbb{R}^3_+ - \widehat{K}$ is a homotopy equivalence and the fundamental group of both spaces is the same.

The union K_i of each overcrossing with the two vertical line segments at its endpoints have a closed neighborhood B_i in $\mathbb{R}^3_+$ which is homeomorphic to a 3-

ball, and has the form of, say, a horseshoe $H_i \times I$, as shown in Figure 6.27(a). Each complement $B_i - K_i$ is homeomorphic to a cylinder with the axis removed $(\mathbb{B}^2 - 0) \times I$, as shown in Figure 6.27(b). Hence it is homotopy equivalent to $\mathbb{B}^2 - 0$, therefore also to $\mathbb{S}^1$. Thus its fundamental group is isomorphic to $\mathbb{Z}$ generated by the class α_i' of a loop that goes around K once; more precisely around the overcrossing of K_i.

If X is the resulting space obtained by removing in $\mathbb{R}^3_+$ the interiors of each B_i and the interior of the disc in which B_i intersects the plane $x_3 = 0$, then X is homeomorphic to $\mathbb{R}^3_+$ and in particular, it is simply connected. Thus we have that $\mathbb{R}^3 - \widehat{K} = X \cup (B_1 - K_1) \cup (B_2 - K_2) \cup \cdots \cup (B_k - K_k)$. On the other hand, the intersection of X with each portion $B_i - K_i$ is homeomorphic to a disc. Therefore, it is simply connected as well. If we assume, inductively, that the fundamental group of $\mathbb{R}^3 - \widehat{K} = X \cup (B_1 - K_1) \cup (B_2 - K_2) \cup \cdots \cup (B_i - K_i)$ is free generated by $\alpha_1, \alpha_2, \ldots, \alpha_i$, then the Seifert–van Kampen theorem implies that the fundamental group of $\mathbb{R}^3 - \widehat{K} = X \cup (B_1 - K_1) \cup (B_2 - K_2) \cup \cdots \cup (B_{i+1} - K_{i+1})$ is obtained adding a new generator around K_i, which of course can be taken as α_{i+1}. This proves our lemma. $\square$

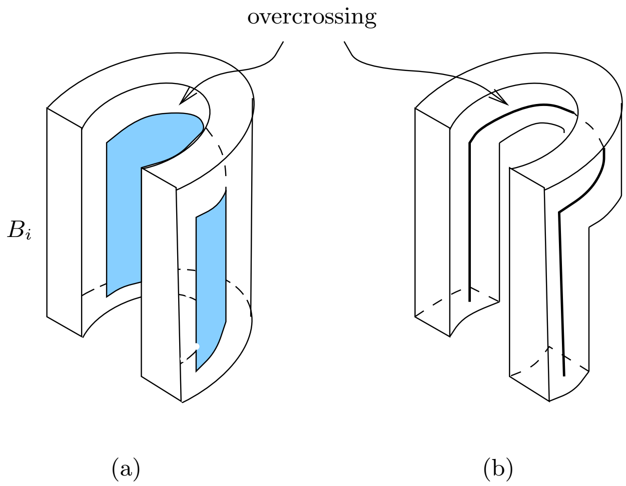


Figure 6.27 The horseshoes

Indeed, due to the fact that we wish to apply the Seifert–van Kampen theorem, we need two open sets. One of them, which we still shall denote $\mathbb{R}^3_+ - \widehat{K}$, is obtained by taking the union with $(\mathbb{R}^2 - \widehat{K}) \times (-\varepsilon, 0]$. Namely, we let it grow a little bit downwards, without altering its homotopy type.

We still have to add the open set $\mathbb{R}^3_-\widehat{K}$, which represents the points of negative height in $\mathbb{R}^3$, but removing from them, not only undercrossings of $\widehat{K}$, but the products of the corresponding portions of arc with the interval $(-\varepsilon, 0)$. Homotopically, this space is equivalent to the original one, and the intersection of both enlarged spaces is also homotopically equivalent to the original intersection of height $x_3 = 0$. Thus we may consider the original spaces for the reasoning.

Let us take a look at the undercrossing of the knot between the ith and the $(i+1)$th overcrossings and assume that the jth overcrossing passes over it, as shown in Figure 6.28. Take the loops λ_i and λ_{i+1} and shift them in order to bring them closer to the crossing and then take loops λ_j and λ'_j which represent α_j, but are such that the pass on different sudes of the undercrossing, as shown in Figure 6.28. Let us now enlarge the projection of the undercrossing in $\mathbb{R}^3_-$, to obtain a ball of dimension 3, which we call D_i, and consider the result of taking the union of $\mathbb{R}^3_+ - \widehat{K}$ with the set-difference $D_i - \widehat{K}$. In order to base all loops at x_0, we adjoin to D_i a line segment from x_0 to A and then, vertically, another one to B un the upper cap of D_i. Thus $D_i - \widehat{K}$ is simply connected, as it is clear, and $(D_i - \widehat{K}) \cap (\mathbb{R}^3_+ - \widehat{K})$ consists of a disc with a (smooth) arc removed from its interior. Hence it has the homotopy type of a circle and thus its fundamental group is infinite cyclic (see Exercise 6.5.11), generated, say, by a loop μ_i based at x_0 which goes once clockwise around the undercrossing in the plane $x_3 = 0$. Then μ_i represents an element of the fundamental group, which we denote by β_i.

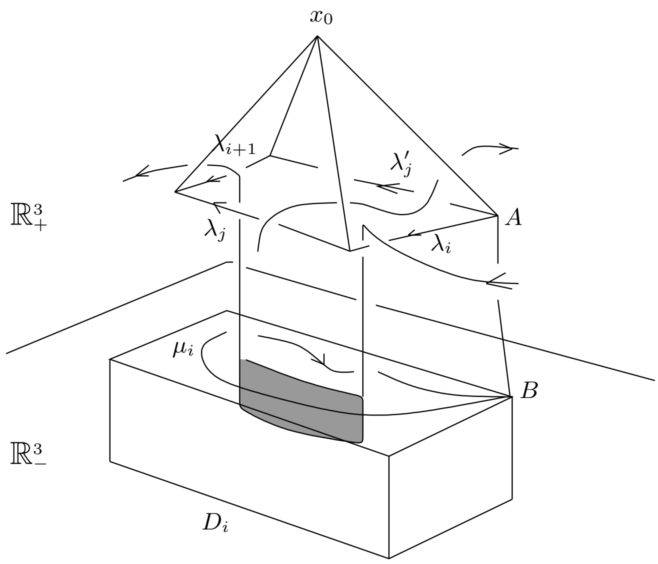


Figure 6.28 Applying the Seifert–van Kampen theorem

According to the Seifert–van Kampen theorem, if we wish to compute the fundamental group of $(\mathbb{R}^3_+ - \widehat{K}) \cup (D_i - \widehat{K})$, we must include the relation $\iota_*(\beta_i) = 1$ in $\pi_1(\mathbb{R}^3_+ - \widetilde{K})$, where ι is the inclusion

$$(\mathbb{R}^3_+ - \widehat{K}) \cup (D_i - \widehat{K}) \hookrightarrow \mathbb{R}^3_+ - \widetilde{K}\,.$$

But $\iota_*(\beta_i)$ is represented by the loop μ_i considered as a loop inn $\mathbb{R}^3_+ - \widetilde{K}$. Sliding μ_i vertically upwards, we obtain a loop which is homotopic to $\lambda_i \lambda_j \overline{\lambda_{i+1}} \overline{\lambda'_j}$, as one can also see in Figure 6.28. Hence the effect of adding $D_i - \widehat{K}$ is equivalent to imposing the relation $\alpha_i\alpha_j\alpha_{i+1}^{-1}\alpha_j^{-1} = 1$ in the group. Notice that inverting the orientation of the jth overcrossing, the relation changes by $\alpha_{i+1}\alpha_j\alpha_i^{-1}\alpha_j^{-1} = 1$.

The other possibility would be that the "undercrossing" would have been included only to put apart two consecutive overcrossings. In this case, the loop μ_i is clearly homotopic to $\lambda_i\overline{\lambda_{i+1}}$. Consequently, the relation needed in this case would have been $\alpha_i\alpha_{i+1}^{-1} = 1$. We denote any of both relations by r_i.

Since the total number of undercrossings is k, the first $k-1$ undercrossings determine relations $r_1, r_2, \ldots, r_{k-1}$ and they yield a description of the fundamental group $\pi_1(Y)$, where

$$Y = (\mathbb{R}^3_+ - \widehat{K}) \cup (D_1 - \widehat{K}) \cup \cdots \cup (D_{k-1} - \widehat{K})$$

is $\langle \alpha_1, \ldots, \alpha_k \mid r_1, \ldots, r_{k-1}\rangle$.

We shall see now that this is already the final description, we are looking for. This is so since the relation coming from the last undercrossing is a consequence of the previous ones. Let Z be the closure of $\mathbb{R}^3 - Y$. To finish the construction of $\mathbb{R}^3 - \widehat{K}$, we have to take the union of $Z - \widehat{K}$ and Y. But $Z - \widehat{K}$ is simply connected and $Y \cap (Z - \widehat{K})$ has an infinite cyclic fundamental group generated by a loop around the last undercrossing. Now we may choose this loop as a large circle in the plane $x_3 = 0$ which runs around all of the projection of our knot. If we slide this circle vertically upwards until it is completely over the knot, and then we contract it, we notice that it represents the trivial element of $\pi_1(Y)$. By the Seifert–van Kampen theorem, we obtain the main result of this Section.

6.5.6 **Theorem**. *The knot group of a knot K is generated by the elements $\alpha_1, \alpha_2, \ldots, \alpha_k$ subject to the relations $r_1, r_2, \ldots, r_{k-1}$.* □

6.5.7 Examples.

(a) To compute the knot group of the trivial knot, we put apart the circle into two semicircles, under the convention that one is the overcrossing. The recipe given above gives us a generator and no relation. Hence *the knot group of the trivial knot is infinite cyclic, namely it is isomorphic to $\mathbb{Z}$.*[‡]

(b) To compute the knot group of the left-handed trefoil knot, one takes overcrossings and undercrossings according to Figure 6.24. Hence we have three generators $\alpha_1, \alpha_2, \alpha_3$ subject to the relations $x_1x_2x_1^{-1}x_3^{-1}$ and $x_2x_3x_2^{-1}x_1^{-1}$.

[‡] As we saw in Chapter 4 (Example 4.4.14), the complement of a trivial knot is homotopically the same as the complement of a solid torus. Thus the knot group of the trivial knot is the fundamental group of a solid torus. Since this has the homotopy type of a circle, the knot group is isomorphic to $\mathbb{Z}$.

If we eliminate α_3 and write $a = \alpha_1$ and $b = \alpha_2$, then we have that *the knot group of the trefoil knot is*

$$G = \langle a, b \mid abab^{-1}a^{-1}b^{-1} \rangle .$$

If we take the permutation group in three letters, Σ_3, there is a homomorphism $G \longrightarrow \Sigma_3$, such that $a \mapsto (12)$, $b \mapsto (23)$, because $(12)(23)(12) = (13) = (23)(12)(23)$. Since (12) and (23) generate Σ_3, the homomorphism is an epimorphism. This proves that G is not abelian and therefore, it is not $\mathbb{Z}$. This way we see again that the trefoil knot is not equivalent to the trivial knot.

(c) If we consider now the *square knot*, shown in Figure 6.23, we can take over- and undercrossings as shown in Figure 6.29 and then we can label them from 1 to 6. The letters a, b, c represent the generators of the knot group corresponding to three of the overcrossings, as shown in the figure, and using the relations given by the undercrossings $1, 2, 4, 5$, we express the other generators in terms of these three.

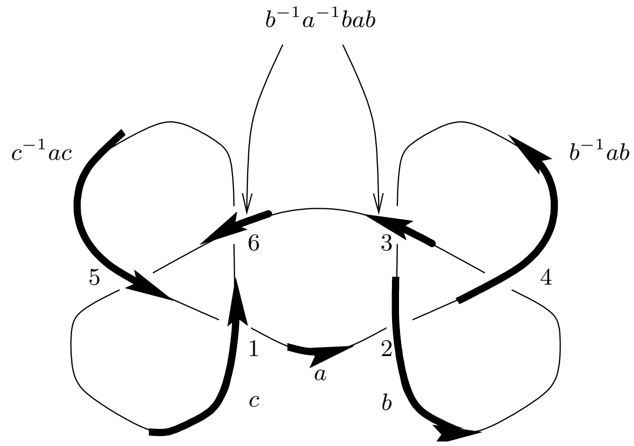


Figure 6.29 Computation of the knot group of the square knot

The relations corresponding to the undercrossings 3 and 6 are

$$(b^{-1}ab)(b^{-1}a^{-1}bab)b^{-1}(b^{-1}a^{-1}bab)^{-1} ,$$

which transforms into $abab^{-1}a^{-1}b^{-1}$, and

$$(c^{-1}ac)(b^{-1}a^{-1}bab)(c^{-1}ac)^{-1}c^{-1} ,$$

which transforms into $acac^{-1}a^{-1}c^{-1}$, by replacing bab by aba. Thus *the knot group of the square knot is*

$$\langle a, b, c \mid abab^{-1}a^{-1}b^{-1}, acac^{-1}a^{-1}c^{-1} \rangle .$$

Nonequivalent knots may have the same knot group. For instance, the left-hand side of the square knot is like a right-handed trefoil knot, while the right-hand side is its mirror image. If we change the left-hand side by another right-handed trefoil knot, then we obtain *granny's knot*, which, as we know, is a different knot (see 6.4.16). However, it is an *exercise* for the reader to verify that the knot group of the granny's knot is isomorphic to the knot group of the square knot. (Compare this with the result of Exercise 6.4.16.)

The problem of deciding if two groups which are given in terms of generators and relations, are isomorphic or not is, in general, impossible to solve. In the best case, it is a difficult task. Therefore, the problem of classifying knots, since it is somehow equivalent to the problem of classifying groups, has no solution. Already in the case of the classification of surfaces we had to face the problem by abelianizing the fundamental groups. Unfortunately, in the case of knot groups the abelianization trivializes the problem, as shown in the following.

6.5.8 **Proposition**. *The abelianization of any knot group is an infinite cyclic group.* $\square$

Therefore, in the case of knots, this procedure brings nothing.

6.5.9 EXERCISE. Compute the knot groups of the stevedor's and the true lovers' knots.

6.5.10 EXERCISE. Let K be a tame knot and let T be a tubular neighborhood of it, which is obtained by taking all points in $\mathbb{R}^3$, whose distance to the knot is less than or equal to some $\varepsilon > 0$, where ε is small enough, that the tube has no self-itersections. Show that $\mathbb{R}^3 - T^\circ$ is a strong deformation retract of $\mathbb{R}^3 - K$. Conclude that the inclusion induces an isomorphism $\pi_1(\mathbb{R}^3 - T^\circ) \cong \pi_1(\mathbb{R}^3 - K)$.

6.5.11 EXERCISE. Let $A \subset \overset{\circ}{\mathbb{B}}{}^2$ be a closed line segment. Show that $\mathbb{B}^2 - A \approx \mathbb{B} - 0$. Generalize the result to the case when A is a "tame" arc, that is, when A is obtained from a line segment through an ambient homeomorphism (of $\mathbb{R}^3$ or of a neighborhood of the segment in $\mathbb{R}^3$). Conclude that for any arc $A \subset \overset{\circ}{\mathbb{B}}{}^2$, the group $\pi_1(\mathbb{B}^2 - A)$ is infinite cyclic.

References

[1] C.C. Adams, *The Knot Book*, W.H. Freeman and Company, New York, 1994.

[2] M. Aguilar, S. Gitler, C. Prieto, *Algebraic Topology from a Homotopical Viewpoint*, Universitext, Springer-Verlag, New York, Berlin, Heidelberg, 2002.

[3] L.V. Ahlfors, *Complex Analysis*, Second Edition, McGraw-Hill Book Co., New York, 1966.

[4] M.A. Armstrong, *Basic Topology*, UTM, Springer-Verlag, New York, Berlin, Heidelberg, 1983).

[5] N. Bourbaki, *Éléments de Mathématique*, première partie, livre III, *Topologie Générale*, Hermann, Paris, 1961.

[6] G.E. Bredon, *Introduction to Compact Transformation Groups*, Academic Press, New York, 1972.

[7] T. tom Dieck, *Topologie*, de Gruyter, Berlin, New York 1991

[8] A. Dold, *Teoría de Punto Fijo*, Vols. I, II and III, translated and adapted by C. Prieto, Monografías del Instituto de Matemáticas, UNAM, 1984.

[9] M.H. Freedman, The topology of 4-dimensional manifolds, *J. Differential Geom.* 17 (1982), 357–453.

[10] S. Gitler, C. Prieto, Eds., *Tendencias Interdisciplinarias en las Matemáticas*, Aportaciones Mat.–Comunicaciones, SMM, México, 2000.

[11] R.E. Gompf, A.I. Stipsicz, 4-*Manifolds and Kirby Calculus*, Graduate Studies in Math. Vol. 20, Amer. Mat. Soc., Providence, 1999.

[12] C.McA. Gordon, J. Luecke, Knots are determined by their complements, *J. Amer. Math. Soc.* 2 (1989), 371–415. [A short account on this work can be consulted in *Bull. Amer. Math. Soc.* (N.S.) 20 (1989), 83–87.]

[13] J. Hass, J. C. Lagarias, The number of Reidemeister moves needed for unknotting, *J. Amer. Math. Soc.* 14 (2001) 399–428

[14] J. Hempel, 3-*Manifolds*, Annals of Mathematics Studies, Princeton University Press & University of Tokyo, Princeton, 1976.

[15] I.M. James, Ed., *History of Topology*, North-Holland, Amsterdam, 1999.

[16] V.F.R. Jones, A polynomial invariant for knots and links via Von Neumann algebras, *Bull. Amer. Math. Soc.* 12 (1985), 103–111.

[17] L. H. Kauffman, New invariants in the theory of knots, *Amer. Math. Monthly* 95, 3 (1988) 195–242.

[18] L. H. Kauffman, *Knots and Physics*, Series on Knots and Everything 1, World Scientific, 1991.

[19] W.M. Massey, *A Basic Course in Algebraic Topology*, Springer-Verlag, 1991.

[20] C. Prieto, Nudos, enlaces y realidad, *Universidad de México*, Núm. 578-579 (marzo-abril 1999), 19–24.

[21] C. Prieto, *Elements of Point-Set Topology*, 2009 (to appear).

[22] W. Rudin, *Real and Complex Analysis*, McGraw-Hill Book Co., New York, 1974.

[23] R. Stöcker, H. Zieschang, *Algebraische Topologie*, B.G. Teubner, Stuttgart, 1988.

Index

Symbols