Lecture Notes for

Introduction to Topology
MA3F1

David Mond

November 13th 2013

Contents

Chapter 1

Introduction

These are revised and corrected lecture notes from the course taught in the autumn of 2013.

As always, please let me know of typos and other errors. I am very grateful to all the people who pointed out errors in earlier drafts.

1.1 Conventions

Topologists use a lot of diagrams showing spaces and maps. For example

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "h"'] \arrow[dr, "p"'] & Y \arrow[d, "q"] \\ W \arrow[r, "g"'] & Z \end{tikzcd}$$

has four spaces and five maps. The diagram is *commutative* if all compositions agree – in this case if $q \circ f = p = g \circ h$.

Sometimes if $X \subset Y$ then a hooked arrow $X \hookrightarrow Y$ is used to denote the inclusion map.

In some arguments in this module and lecture notes, where we are concerned about the existence of a certain map, a dashed arrow is used to indicate the map in question, instead of the standard solid arrow, in order to emphasise our interest in this particular map. For example, in Example 2.0.6 4 below, we revise quotient topologies, and give the (transparently easy) proof of the following proposition.

Proposition 1.1.1. *Let $\sim$ be an equivalence relation on the space X, and let Q be the set of equivalence classes with its quotient topology. If $f : X \to Y$*

is a continuous map, then there is a continuous map $\bar{f} : Q \to Y$ *making the following diagram commute, if and only if* $f(x_1) = f(x_2)$ *every time* $x_1 \sim x_2$.

$$\begin{tikzcd} X \arrow[d, "q"'] \arrow[dr, "f"] & \\ Q \arrow[r, dashed, "\bar{f}"'] & Y \end{tikzcd}$$

The phrase *passing to the quotient* is often used here. The proposition can be stated as "the continuous map $f : X \to Y$ passes to the quotient to define a continuous map $\bar{f} : Q \to Y$ if and only if for all $x_1, x_2, \in X$, $x_1 \sim x_2 \implies f(x_1) = f(x_2)$."

Pictures and Diagrams

I have drawn lots of pictures, because I believe they help. But even more helpful than my pictures, would be pictures you yourself draw. A picture or diagram is helpful because it engages the visual cortex, which is capable of focusing simultaneously on more information than other parts of the brain. However, it can suffer overload. The mind can grind to a halt when confronting a picture, especially where each element has some conceptual complexity. If this happens with any of the pictures in these lecture notes, the best thing is to draw your own, adding details as they are mentioned in the mathematical development the picture is intended to illustrate. Building a picture up step by step is easier and more helpful than trying to understand one that is already completed.

Chapter 2

Topology versus Metric Spaces

This section is mostly propaganda and a little revision.

Let X be a metric space with metric d. We denote by $B(x_0, \varepsilon)$ the open ball centred at x_0 with radius ε:

$$B(x_0, \varepsilon) := \{x \in X : d(x, x_0) < \varepsilon\}.$$

Definition 2.0.2. Let X be a metric space with metric d. A set $U \subset X$ is *open* if for all $x \in U$ there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subset U$, and is *closed* if its complement is open.

Lemma 2.0.3. *The collection $\mathscr{T}$ of open sets in a metric space (X, d) has the following properties:*

1. $X \in \mathscr{T}, \emptyset \in \mathscr{T}$
2. *the union of any collection of members of $\mathscr{T}$ is in $\mathscr{T}$ ("$\mathscr{T}$ is closed under arbitrary unions").*
3. *The intersection of a finite collection of members of $\mathscr{T}$ is in $\mathscr{T}$ ("$\mathscr{T}$ is closed under finite intersections").*

Definition 2.0.4.

1. If X is a set (not necessarily a metric space) then any collection $\mathscr{T}$ of subsets of X with properties 1- 3 of 2.0.3 is called a *topology* on X. The sets belonging to $\mathscr{T}$ are usually called the *open subsets* of X (with respect to $\mathscr{T}$). The set X together with a topology $\mathscr{T}$ is called a *topological space*.

2. Let $(X_1, \mathscr{T}_1)$ and $(X_2, \mathscr{T}_2)$ be topological spaces. A map $f : X_1 \to X_2$ is said to be continuous with respect to $\mathscr{T}_1$ and $\mathscr{T}_2$ if for every $U \in \mathscr{T}_2$, $f^{-1}(U) \in \mathscr{T}_1$. In other words, f is continuous if the preimage of every open set is open. The map f is a *homeomorphism* if it is continuous and has a continuous inverse.

The topology on a metric space (X, d) defined by 2.0.2 is called the *metric topology.*

Lemma 2.0.5. *Let $f : (X_1, d_1) \to (X_2, d_2)$ be a map of metric spaces. Then f is continuous as a map of metric spaces, if and only if it is continuous with respect to the metric topologies on X_1 and X_2.*

Proof. Revision exercise. $\square$

Why study topology?

Many topological spaces do not have natural metrics.

Example 2.0.6. 1. On $C^0[0,1]$, there are several well-known metrics: for example, $d_0(f,g) = \sup\{|f(x) - g(x)| : x \in [0,1]\}$ and $d_H(f,g) = \int_0^1 |f(x)-g(x)|^2 dx$. But what about $C^0(\mathbb{R})$? Neither metric on $C^0[0,1]$ extends to a metric on $C^0(\mathbb{R})$, since both rely on the compactness of $[0,1]$ for their finiteness.

2. Since $C^1[0,1] \subset C^0[0,1]$, the metrics on $C^0[0,1]$ restrict to metrics on $C^1[0,1]$. But these metrics are a little deficient: for example the two functions with graphs

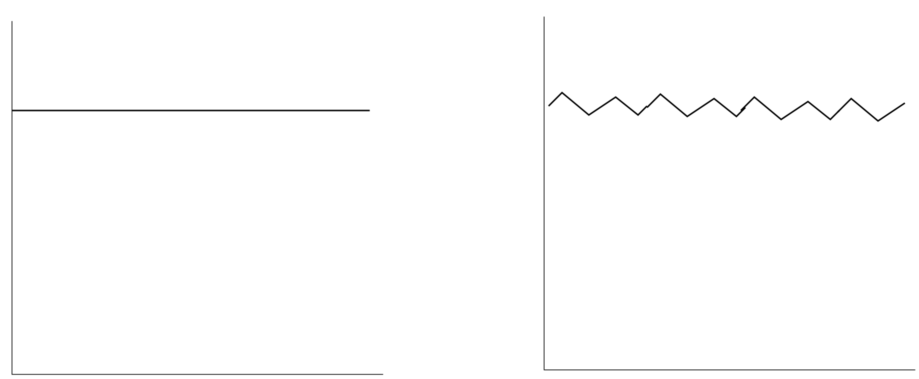


Figure 1

are close in the metrics on $C^0[0,1]$ even though one jiggles up and down very fast while the other is constant. It is possible to mimic the two metrics on $C^0[0,1]$ to define metrics on $C^1[0,1]$ which remedy this. We set

$$d_1(f,g) = \sup\{|f(x) - g(x)| + |f'(x) - g'(x)| : x \in [0,1]\},$$

$$d_H^1(f,g) = \int_0^1 (f(x) - g(x))^2 + (f'(x) - g'(x))^2 dx,$$

for example. The same can be done for $C^k[0,1]$ for finite k. But what if $k = \infty$?

In this course we will *not* be interested in spaces of functions (as in Example 2.0.6 1 and 2). Instead, we will be more interested in "geometric objects" (though I will not attempt a definition of this term here!).

3. Let S be the set of all straight lines in the plane, and let $S_0 \subset S$ be the set of lines passing through 0. There is a "natural" metric on S_0: $d(\ell_1, \ell_2) =$ angle between ℓ_1 and ℓ_2. It has the property that if we apply an isometry of $\mathbb{R}^2$ which leaves 0 fixed, such as a rotation about 0 or a reflection in a line through 0, then the "distance" we have defined does not change:

$$d(f(\ell_1), f(\ell_2)) = d(\ell_1, \ell_2)$$

for any two lines $\ell_1, \ell_2 \in S_0$ and any isometry f of $\mathbb{R}^2$ fixing 0.

On the other hand, one can prove that there is no metric on S with a similar property, that for all lines ℓ_1 and ℓ_2 and isometries f of $\mathbb{R}^2$, $d(f(\ell_1), f(\ell_2)) = d(\ell_1, \ell_2)$. Note that there are more isometries now: in addition to those that fix 0, there are translations, reflections in lines not passing through 0, etc.

This is rather serious: with any metric, S will take on a "shape" which does not allow all of the symmetries which we would like it to have. In mathematics, "abstraction" is the process by which one throws away all of the aspects of a problem not deemed essential. Although premature abstraction can make mathematics incomprehensible, by depriving us of the details of motivating examples, abstraction is ultimately a process of simplification. To endow S with a metric with spurious bumps and lumps would go in the opposite direction.

Nevertheless, even without a metric on S, it is natural (and correct) to suspect that S has some topological properties. For example, S seems to be path-connected: one can deform any line in the plane to any other in a "continuous" way. Later we will show that there is a way of giving S a reasonable topology which makes this precise, and for which each of the symmetries described above is a homeomorphism.

4. If X is a set, and $\sim$ is an equivalence relation on S, then let Q be the set of equivalence classes. It is often referred to as the quotient of X by the equivalence relation, and often denoted $X/\sim$. For $x \in X$, let $[x]$ be its equivalence class, and define a map $q : X \to Q$ (the "quotient map") by $q(x) = [x]$. If X is a topological space, there is a natural way of giving Q a topology: we declare a set $U \subseteq Q$ open if $q^{-1}(U)$ is open. It is evident that this makes the map q continuous.

   Proposition 2.0.7. *Let $\sim$ be an equivalence relation on the space X, and let Q be the set of equivalence classes, with the quotient topology. If $f : X \to Y$ is a continuous map, then there is a continuous map $\bar{f}$ making the following diagram commute, if and only if $f(x_1) = f(x_2)$ every time $x_1 \sim x_2$.*

$$\begin{tikzcd} X \arrow[d, "q"'] \arrow[dr, "f"] & \\ Q \arrow[r, dashed, "\bar{f}"'] & Y \end{tikzcd}$$

   Proof. It is obvious that a necessary and sufficient condition for the existence of a map $\bar{f} : Q \to Y$ such that $\bar{f} \circ q = f$, is that for all $x_1, x_2 \in X$, $x_1 \sim x_2 \implies f(x_1) = f(x_2)$. It remains to show that $\bar{f}$ is continuous. Let U be open in Y. We have to show that $\bar{f}^{-1}(U)$ is open in Q. But $\bar{f}^{-1}(U)$ is open if and only if its preimage in X is open – i.e. if and only if $q^{-1}(\bar{f}^{-1}(U))$ is open in X. However $q^{-1}(\bar{f}^{-1}(U)) = f^{-1}(U)$, and this is open by the continuity of f. $\square$

   The quotient topology is extremely useful. For example, it allows us to give a mathematical definition of "gluing objects together", as we now describe.

5. If X and Y are topological spaces, $A \subset X$, and $f : A \to Y$ is a map, we define a space $X \cup_f Y$, as follows: as a set it is the quotient of

the disjoint union $X \coprod Y$ by the equivalence relation generated by the relation

$$x \sim y \text{ if } x \in A \text{ and } y = f(x), \tag{2.0.1}$$

and its topology is the quotient topology we get from $X \coprod Y$. Note that the relation specified by (2.0.1) is not itself an equivalence relation, since we do not explicitly require that $x \sim x$ and $y \sim y$ for all $x \in X$ and $y \in Y$, nor do we ensure transitivity (if $f(x_1) = y = f(x_2)$ then we must require $x_1 \sim x_2$, otherwise our relation will not be transitive). But if we extend (2.0.1) by these requirements, we do get an equivalence relation, which we refer to as the equivalence relation generated by (2.0.1).

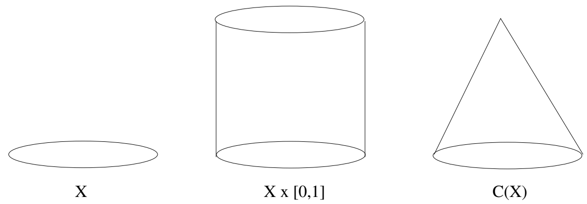
6. We can also glue a space to itself: if, as before, $A \subset X$ and now $f : A \to X$ is a map, we define a new space by subjecting X to the equivalence relation defined by

$$x \sim x' \quad \text{if } x \in A \text{ and } x' = f(x).$$

As before, in this description of the equivalence relation we do not specify "$x \sim x$ for all $x \in X$" or "$x \sim x'$ if $x' \in A$ and $x = f(x')$". The first of these holds for *all* equivalence relations, and so must hold here (since we are defining an equivalence relation). The second is a consequence of the statement that $x \sim x'$ if $x \in A$ and $x' = f(x)$, by the symmetry property of equivalence relations.

If $X = [0, 2] \times [0, 1], A = \{0\} \times [0, 1]$, and $f : A \to X$ is defined by $f(0, y) = (2, 1 - y)$, the space we get is the Möbius strip. Although it is possible to give the Möbius strip a metric, it is not needed, or natural. For everything that interests us about the Möbius strip, a topology is enough, as we will see. Indeed, when we make a Möbius strip out of paper following the above gluing data, we get something that is not rigid, that bends as we push it. Since it is not rigid, its identity as a metric subspace of $\mathbb{R}^3$ is superfluous and irrelevant.

7. Let X be any topological space. The *cone on* X is the topological space defined as follows: we form the product $X \times [0, 1]$ and then identify all of the points of $X \times \{1\}$ with one another – we squish $X \times \{1\}$ to a single point.



That is,

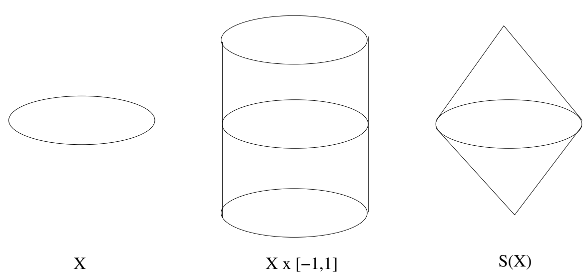
$$C(X) = \frac{X \times [0,1]}{(x,1) \sim (x',1) \text{ for all } x, x' \in X}.$$

Important remark The drawing of $C(X)$ here shows X as a subset of $\mathbb{R}^2$, and shows $C(X)$ as the union of the line segments in $\mathbb{R}^2 \times \mathbb{R}$ joining points of $X \times \{0\}$ to a point of the form $(x_0, 1) \in \mathbb{R}^2 \times \mathbb{R}$. How reasonable is this representation? We will prove the following statement:

Proposition 2.0.8. *Let $X \subset \mathbb{R}^n$ and let $p_0 \in \mathbb{R}^n$ be any point. Then $C(X)$ is homeomorphic to the union of line segments joining points of $X \times \{0\} \subset \mathbb{R}^n \times \mathbb{R}$ to the point $(p_0, 1) \in \mathbb{R}^n \times \mathbb{R}$.*

8. Let X be any topological space. The *suspension* of X, $S(X)$, is the space obtained by forming the product $X \times [-1, 1]$ and then squishing all of $X \times \{-1\}$ to a single point, and all of $X \times \{1\}$ to a single point. That is

$$S(X) = \frac{X \times [-1,1]}{(x,-1) \sim (x',-1) \text{ for all } x, x' \in X; (x,1) \sim (x',1) \text{ for all } x, x' \in X}$$

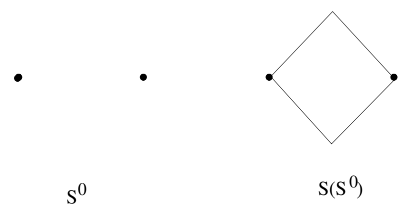


We will prove

Proposition 2.0.9. *Let $X \subset \mathbb{R}^n$ and let $p_0 \in \mathbb{R}^n$. Then $S(X)$ is homeomorphic to the union of the line segments joining points of $X \times \{0\} \subset \mathbb{R}^n \times \{0\}$ to the points $(p_0, 1)$ and $(p_0, -1)$ in $\mathbb{R}^n \times \mathbb{R}$.*

9. The sphere S^n is the set of points

$$\{x \in \mathbb{R}^{n+1} : ||x|| = 1\}.$$



In particular, $S^0 = \{-1, 1\} \subset \mathbb{R}$. It does not look very round. By the last proposition, $S(S^0)$ is homeomorphic to the (boundary of) a square, and therefore to the circle S^1. In fact, for *all* $n \in \mathbb{N}$, $S(S^n)$ is homeomorphic to S^{n+1}. This also follows from the proposition. I leave it as an exercise.

2.1 Subspaces

If X is a topological space and $Y \subset X$, then the *subspace topology* on Y is the topology in which the open sets are sets $Y \cap A$, where A is open in X. Endowed with this topology, Y is a *subspace* of X. When $Y \subset X$ and we speak of a subset of Y being open, or closed, we *always* mean with respect to the subspace topology. Proof of the following statements is an easy exercise.

Proposition 2.1.1. *Suppose that X is a topological space and Y is a subspace of X.*

1. *Suppose that $W \subset Y$. Then W is closed in Y if and only if there exists V closed in X such that $W = Y \cap V$.*

2. *If A is open in Y and Y is open in X then A is open in X.*

3. *If W is closed in Y and Y is closed in X then W is closed in X.*

4. *If X is Hausdorff then so is Y.* $\square$

2.2 Homeomorphism

A map $\varphi : X \to Y$ is a homeomorphism if it is continuous and has a continuous inverse. The fact that it has an inverse means that it is a bijection, but it is important to note that *not every continuous bijection is a homeomorphism.*

Example 2.2.1. *Let S^1 denote the unit circle $\{(x,y) \in \mathbb{R}^2 : x^2+y^2=1\}$, and consider the map $f : [0,1) \to S^1$ defined by $f(t) = (\cos 2\pi t, \sin 2\pi t)$. Then f is a continuous bijection, but clearly not a homeomorphism, since S^1 is compact while $[0,1)$ is not.*

Exercise Since f is not a homeomorphism, its inverse cannot be continuous. This amounts to the fact that there are open sets $U \subset [0,1)$ whose image $f(U)$ in S^1 is not open. Find them.

Nevertheless, there is a very useful result which assures us that with an extra hypothesis, a continuous bijection is a homeomorphism.

Proposition 2.2.2. *Let $f : X \to Y$ be a continuous bijection, with X compact and Y Hausdorff. Then f is a homeomorphism.*

Proof. $U \subset X$ open $\implies$ $X \setminus U$ closed $\implies$ $X \setminus U$ compact (since X is compact) $\implies$ $f(X \setminus U)$ compact $\implies$ $f(X \setminus U)$ closed in Y (since Y is Hausdorff – a compact subset of a Hausdorff space is closed) $\implies$ $f(U)$ is open (as f is a bijection, $f(U) = Y \setminus f(X \setminus U)$). □

We will refer to this as the "compact-to-Hausdorr lemma".

We now use it to prove Propositions 2.0.8 and 2.0.9.

Proof. of 2.0.8: Denote by Y the union of line segments joining the points $(x,0)$ of $X \times \{0\}$ to the point $(p,1)$:

$$Y = \{(1-t)(x,0) + t(p,1) : x \in X,\ t \in [0,1]\}.$$

- Step 1: define a suitable map $f : X \times [0,1] \to Y$, by

  $$f(x,t) = (1-t)(x,0) + t(p,1) = ((1-t)x + tp, t).$$

  This is obviously continuous (why?) and has the following properties:
  (i) All points $(x,1)$ (exactly the points that are identified to one another by the equivalence relation $\sim$) are mapped to $(1,p)$.
  (ii) f is 1-to-1 except for this: we have $f(x,t) = f(x',t')$ only if $(x,t) = (x',t')$, unless $t = t' = 1$.
  (iii) f is surjective.

- Step 2: f passes to the quotient to define a continuous map $\bar{f}$: $C(X) \to Y$,

$$\begin{tikzcd} X \times [0,1] \arrow[d, "q"'] \arrow[dr, "f"] & \\ C(X) = \frac{X\times[0,1]}{\sim} \arrow[r, dashed, "\bar{f}"'] & Y \end{tikzcd}$$

  by (i) of Step 1 and Proposition 2.0.7. This map is injective and surjective, by (ii) and (iii) of Step 1. So it is a continuous bijection.

- Step 3: $C(X)$ is compact, as it is the image under the continuous map q of the compact space $X \times [0,1]$. Y is Hausdorff, as it is contained in $\mathbb{R}^{n+1}$. So by Proposition 2.2.2, $\bar{f}$ is a homeomorphism.

□

Exercise (i) Prove 2.0.9. (ii) In fact 2.0.8 and 2.0.9 hold even without the hypothesis that X be compact. To prove this for 2.0.8, find an explicit inverse to the map $\bar{f}$ constructed in Step 2. The same approach works for 2.0.9.

If two topological spaces are homeomorphic then in everything that concerns their topology alone, they are the same. At around the time when the notions of topological space and homeomorphism were first introduced, Pieano and others surprised mathematicians with their construction of surjective continuous maps ("Peano curves") from $[0,1]$ to $[0,1]^2$. These maps suggested that perhaps $\mathbb{R}$ and $\mathbb{R}^2$ might be homeomorphic, in which case $\mathbb{R}^n$ and $\mathbb{R}^m$ would be homeomorphic for all m and n. Fortunateley this turned out not to be the case; Brouwer proved in 19.. that if the open sets $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^n$ are homeomorphic then $m = n$. The proof is more difficult than one might imagine, and we will not be able to give it in this module. It can easily be proved using the methods of algebraic topology. We will be able to show that $\mathcal{R}$ and $\mathbb{R}^2$ are not homeomorphic, and that $\mathbb{R}^2$ and $\mathbb{R}^3$ are not homeomorphic – see exercises .. and ...

The question naturally arises, to find ways of deciding whether two given spaces are homeomorphic. Some very major mathematics has developed by restricting this problem to particular classes of topological spaces.

Definition 2.2.3. A Hausdorff topological space X is an *n-dimensional manifold* if each point has a neighbourhood $U \subset X$ which is homeomorphic to an open set in $\mathbb{R}^n$.

An obvious example is $\mathbb{R}^n$ itself. Slightly less obvious is the n-sphere S^n. The two open sets $U_1 = S^n \setminus \{N\}$ and $U_2 = S^n \setminus \{S\}$ cover S^n, and each is homeomorphic to $\mathbb{R}^n$, via stereographic projection. The picture shows stereographic projection for $n = 2$, from $U_1 \to \mathbb{R}^2 \times \{0\} = \mathbb{R}^2$. The map ϕ takes the point x to the point where the straight line from the north pole N to x meets the plane $\{z = 0\}$.

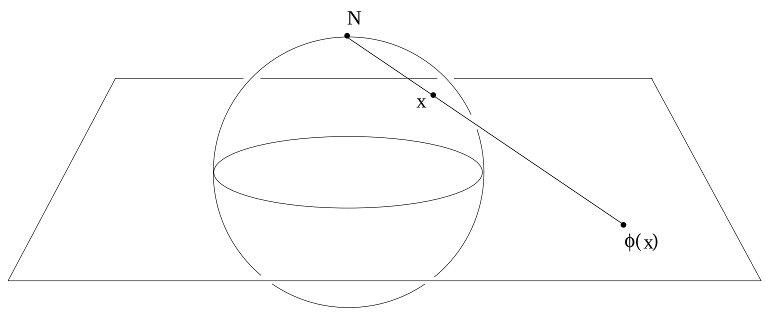


Figure 2

The 2-sheeted cone $X := \{(x, y, z) \in \mathbb{R}^3 : x^2+y^2-z^2 = 0\}$ is *not* a manifold. One can see this very easily: every connected neighbourhood U of the point $(0, 0, 0)$ in X is disconnected by the removal of the single point $(0, 0, 0)$. If $\varphi : U \to V$ were a homeomorphism from U to an open set V in $\mathbb{R}^2$, then V would be disconnected by the removal of the single point $\varphi(0, 0, 0)$. But this is impossible: no open set in $\mathbb{R}^2$ can be disconnected by the removal of a single point.

Unlike the classification of all topological spaces, the classification of n-dimensional manifolds, at least for low values of n, is feasible. Very early on (about 1910) it was proved by Brouwer that if an m-dimensional manifold M_1 is homeomorphic to an n-dimensional manifold M_2 then $m = n$.

A 2-dimensional manifold is called a surface. Compact surfaces have been completely classified, and we will review the clasification and sketch its proof later in the course. The classification is surprisingly simple. Compact surfaces are divided into two classes, orientable and non-orientable. The following diagram shows the first four members of a list of all compact orientable surfaces, up to homeomorphism.

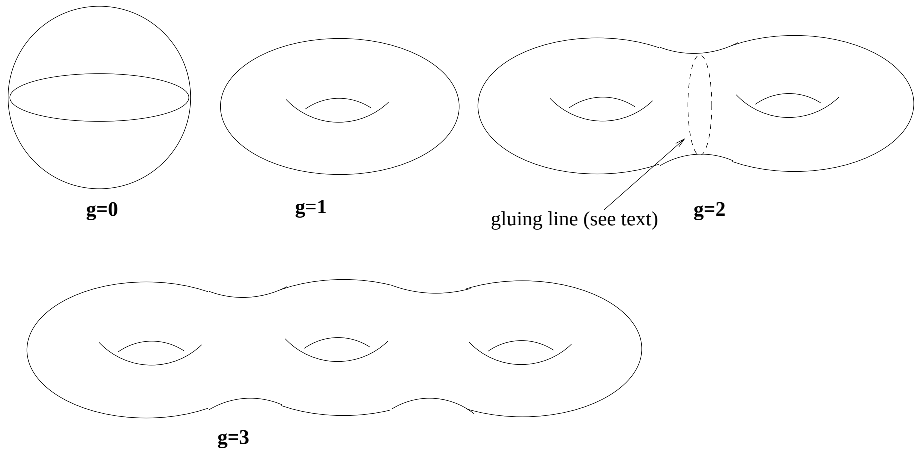


Figure 3

The integer g is called the *genus* of the surface. Up to homeomorphism, there is a unique compact orientable surface of genus g, for each $g \in \mathbb{N}$. The surface of genus g can be constructed from the surface of genus $g-1$ as follows: remove a small disc from surface of genus $g-1$, and a disc of the same size from the surface of genus 1 (the torus). Now glue the two surfaces together along the edges of the holes. The dotted line on the surface of genus 2 shows where this gluing has been done. By repeatedly applying this procedure one obtains a whole succession of compact orientable surfaces, one for each natural number. And these are all there are, up to homeomorphism!

Recent breakthroughs on the classification of 3-dimensional manifolds have centred on the proof by the Russian mathematician Perelman of the celebrated Poincaré Conjecture:

Theorem 2.2.4. *Every compact simply connected 3-dimensional manifold is homeomorphic to S^3.*

Poincaré conjectured that this was true in the late 19th century. Wondering whether the universe was homeomorphic to a 3-sphere (this was before General Relativity) he sought ways of establishing that this was indeed the case.

To understand the meaning of the term "simply connected", we need a couple of simple definitions.

Definition 2.2.5. *1. Let X and Y be topological spaces. Two continuous maps $f : X \to Y$ and $g : X \to Y$ are homotopic if there is a* *continuous map* $F : X \times [0,1] \to Y$ *such that* $F(x,0) = f(x)$ *and* $F(x,1) = g(x)$.

2. *A topological space* X *is simply connected if if it is path connected and every continuous map* $f : S^1 \to X$ *is homotopic to a constant map. That is, if every continuous map* $S^1 \to X$ *can be continuously deformed in* X *to a constant map.*

The 2-sphere is simply connected – you can't tie a string to a slippery ball. The (hollow) torus is not simply connected: neither of the two closed curves shown in the diagram can be contracted *on the torus* to a point.

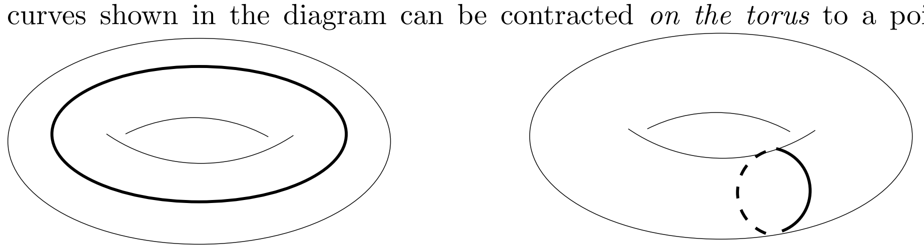


Figure 4

2.3 Overview of the fundamental group

Among all of the surfaces in the list shown in Figure 3, only the sphere is simply connected. This is the 2-dimensional version of Poincaré's conjecture/Perelman's theorem. To measure the extent to which a space is not simply connected, we will associate a group, the so-called Fundamental Group, to each space X. This will be one of the major themes of the module.

NB The following discussion omits a technical detail which will be properly dealt with when the topic is covered in the module (as opposed to the present introductory outline).

The fundamental group of the space X is denoted by $\pi_1(X)$, and we will see that it has the following properties:

1. A continuous map $f : X \to Y$ induces a homomorphism $f_* : \pi_1(X) \to \pi_1(Y)$.

2. Given continuous maps $f : X \to Y$ and $g : Y \to Z$, we have

$$(g \circ f)_* = g_* \circ f_* \tag{2.3.1}$$

3. If $i_X : X \to X$ is the identity map, then

$$(i_X)_* \text{ is the identity homomorphism.} \tag{2.3.2}$$

These have the following consequence:

Proposition 2.3.1. *If $f : X \to Y$ is a homeomorphism then $f_* : \pi_1(X) \to \pi_1(Y)$ is an isomorphism.*

Proof. Let $g : Y \to X$ be the inverse of f. Then $g \circ f = i_X$ and $f \circ g = i_Y$. By (2.3.1), we therefore have

$$g_* \circ f_* = (i_X)_* \quad \text{and} \quad f_* \circ g_* = (i_Y)_*.$$

By (2.3.2), this means that f_* and g_* are mutually inverse homomorphisms, and thus are isomorphisms. $\square$

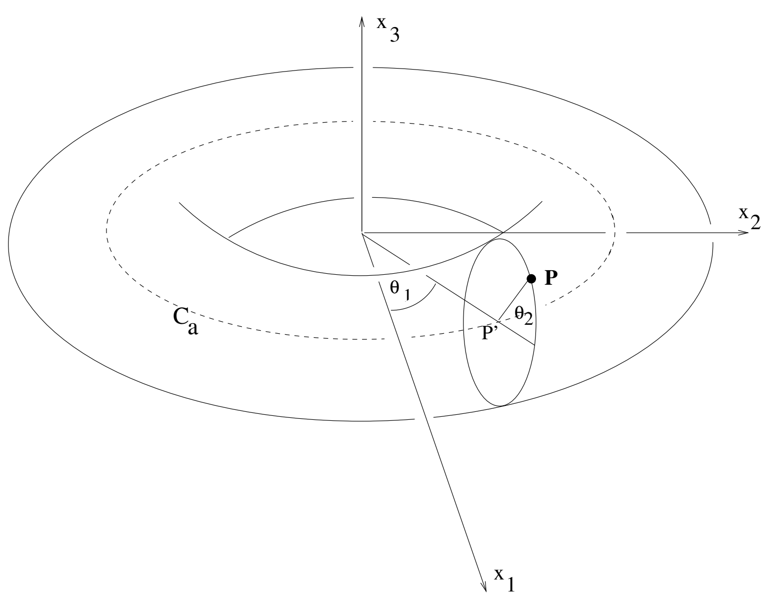
It follows that as a means of proving that two spaces are *not* homeomorphic, one can try to prove that their fundamental groups are not isomorphic.

The fundamental group is just the first of many groups that can be associated to a topological space X, all with the properties described in (2.3.1) and (2.3.2). The module Algebraic Topology studies one important family of these groups, the *homology groups*.

Chapter 3

Examples and Constructions

1. Choose real numbers $0 < b < a$. The *torus* $T = T_{ab} \subset \mathbb{R}^3$ is shown in the following diagram. It is the set of points a distance b from the circle C_a (shown with a dashed line) in the x_1x_2 plane with centre 0 and radius a.



Each point $P \in T$ is determined by the two angles θ_1, θ_2 shown; in terms of these angles its co-ordinates in the ambient $\mathbb{R}^3$ are

$$((a + b\cos\theta_2)\cos\theta_1, (a + b\cos\theta_2)\sin\theta_1, b\sin\theta_2). \tag{3.0.1}$$

This formula specifies a bi-periodic map Φ from the $\theta_1\theta_2$-plane to the torus.

In fact the torus is the image of the square $[0, 2\pi] \times [0, 2\pi]$ under Φ, or indeed of any square $[c, c + 2\pi] \times [d, d + 2\pi]$.

Since any point on the torus is uniquely specified by the two angles θ_1, θ_2, it is easy to guess at a map $S^1 \times S^1 \to T$: the point on $S^1 \times S^1$ with angles θ_1, θ_2 is mapped to the point on the torus specified by these angles. The point specified on $S^1 \times S^1$ has cartesian coordinates $(x_1, x_2, x_3, x_4) = (\cos\theta_1, \sin\theta_1, \cos\theta_2, \sin\theta_2)$, and replacing the trigonometric functions in (3.0.1) by these coordinates, we get the expression

$$((a + bx_1)x_3, (a + bx_1)x_4, bx_4) \tag{3.0.2}$$

which determines a continuous map $\mathbb{R}^4 \to T$. Its restriction $\phi : S^1 \times S^1 \to T$ is bijective, and since $S^1 \times S^1$ is compact and T is Hausdorff (as a subset of the metric space $\mathbb{R}^3$), ϕ is a homeomorphism by the compact-to-Hausdorff lemma 2.2.2.

2. Consider the map $\exp : \mathbb{R} \to S^1$, $f(t) = e^{2\pi i t} = \cos 2\pi t + i \sin 2\pi t$. Note that exp is both continuous, and a homomorphism of groups: $f(t_1 + t_2) = f(t_1)f(t_2)$. Like any map, exp determines an equivalence relation on its domain: $t_1 \sim t_2$ if $\exp(t_1) = \exp(t_2)$. Evidently $\exp(t_1) = \exp(t_2)$ if and only if $t_1 - t_2 \in \mathbb{Z}$, and so the equivalence classes are cosets of the subgroup $\mathbb{Z}$, and the quotient space is correctly described as $\mathbb{R} / \mathbb{Z}$. Indeed, the kernel of exp is $\mathbb{Z}$. The first isomorphism theorem of group theory says that the map $\overline{\exp} : \mathbb{R} / \ker \exp = \mathbb{R} / \mathbb{Z} \to S^1$ induced by exp is an isomorphism of groups. I claim that it is also a homeomorphism.

$$\begin{tikzcd} \mathbb{R} \arrow[d, "q"'] \arrow[dr, "\exp"] & \\ \mathbb{R} / \mathbb{Z} \arrow[r, "\overline{\exp}"'] & S^1 \end{tikzcd} \tag{3.0.3}$$

The argument is completely standard: $\bar{f}$ is continuous by the passing-to-the-quotient lemma (Proposition 2.0.7); $\mathbb{R} / \mathbb{Z}$ is compact, since it is the image of the compact space $[0, 1]$ under the continuous map $q : \mathbb{R} \to \mathbb{R} / \mathbb{Z}$; S^1 is Hausdorff since it is a metric space; f is surjective, and therefore by construction $\bar{f}$ is bijective. Once again, the conclusion follows by the compact-to-Hausdorff lemma (Proposition 2.2.2).

Thus the quotient $\mathbb{R} / \mathbb{Z}$ is the same as S^1 both as a group and as a topological space.

3. If $f_1 : X_1 \to Y_1$ and $f_2 : X_2 \to Y_2$ are homeomorphisms then so is the

map

$$f_1 \times f_2 : X_1 \times X_2 \to Y_1 \times Y_2$$

defined by $(f_1 \times f_2)(x_1, x_2) = (f_1(x_1), f_2(x_2))$. If f_1 and f_2 are isomorphisms of groups, then so also is $f_1 \times f_2$.

For this reason, the product map

$$\exp \times \exp : \mathbb{R} \times \mathbb{R} \to S^1 \times S^1 \tag{3.0.4}$$

induces an isomorphism of groups and a homeomorphism of spaces

$$\frac{\mathbb{R} \times \mathbb{R}}{\mathbb{Z} \times \mathbb{Z}} \to S^1 \times S^1. \tag{3.0.5}$$

By composing this with the homeomorphism $\phi : S^1 \times S^1 \to T$ of Example 1 above, we obtain a homeomorphism

$$\frac{\mathbb{R} \times \mathbb{R}}{\mathbb{Z} \times \mathbb{Z}} \to T \tag{3.0.6}$$

4. If we restrict the map $\exp \times \exp$ of (3.0.4) to the unit square $[0,1] \times [0,1]$, we still get a surjective map to $S^1 \times S^1$. It follows, by the now standard arguments of 2.0.7 and the compact-to-Hausdorff lemma, that the image of $[0,1] \times [0,1]$ in $\mathbb{R} \times \mathbb{R}/(\mathbb{Z} \times \mathbb{Z})$ is mapped homeomorphically to $S^1 \times S^1$. This image is $([0,1] \times [0,1])/\sim$ where

$$(0, y) \sim (1, y) \text{ for all } y \in [0,1], \quad (x, 0) \sim (x, 1) \text{ for all } x \in [0,1].$$

By composing with the homeomorphism $S^1 \times S^1 \to T$ of Example 1 above, we obtain a homeomorphism

$$\frac{[0,1] \times [0,1]}{\sim} \to T. \tag{3.0.7}$$

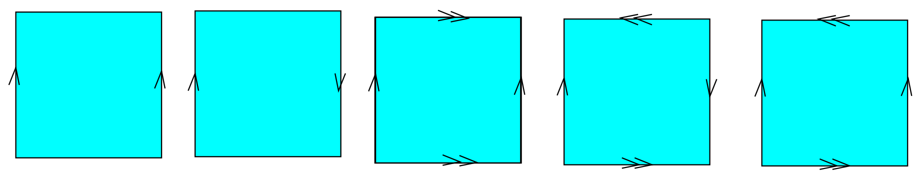
The homeomorphism (3.0.7) can be understood in very down-to-earth terms: glue two opposite sides of a square together to obtain a cylinder, and then glue the two ends of the cylinder together to obtain a torus. If we glue the edges of the square together in different ways, we obtain other spaces - which we will study later.

5. Notation: we indicate the equivalence relation on the square $[0,1] \times [0,1]$ generated by $(0,t) \sim (1,t)$ by drawing arrows on the left and right

edges of the square, pointing in the same direction, as in the first square below. The quotient is the cylinder $S^1 \times [0,1]$. The equivalence relation

$$(0,t) \sim (1, 1-t)$$

is indicated by drawing arrows pointing in opposite directions, as in the second square below. The quotient is the Möbius strip. The quotient of the third square is the torus T. Note that we use a double arrow on the top and bottom edges - we do not want to identify them with the left and right edges. The quotient of the fourth square is the projective plane $\mathbb{R}\mathbb{P}^2$.
Exercise: Prove this. And what is the quotient of the fifth square?



6. If one takes a paper rectangle and glues one pair of opposite edges after giving it a whole twist (rather than the half twist used to make the Möbius strip), the space obtained is homeomorphic to $[0,1] \times S^1$, since the gluing relation is the same. It doesn't look like a cylinder because it is not embedded in $\mathbb{R}^3$ in the usual way.

6. Take a Möbius strip and cut it along the central circle; it does not fall into two pieces, but instead one obtains a strip with a whole twist, homeomorphic to $S^1 \times [0,1]$. The new strip, with its whole twist, can be carefully wrapped onto itself so that it becomes a Möbius strip with double thickness. Try this! Note that it seems impossible to perform this double wrapping with the cyclinder embedded in $\mathbb{R}^3$ in the usual way.

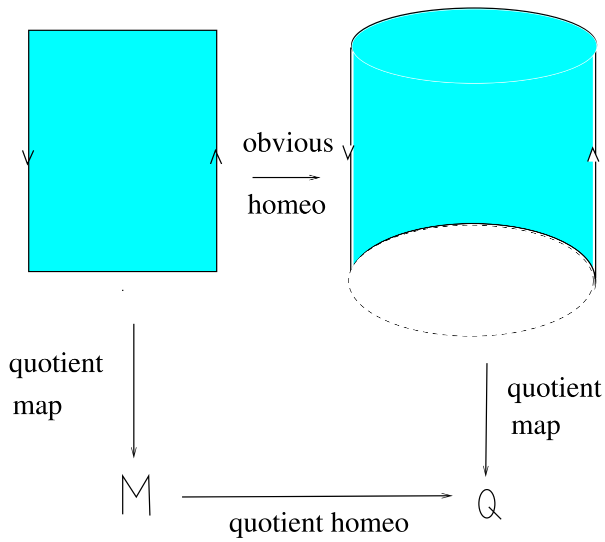
7. This "double wrapping" of $S^1 \times [0,1]$ onto the Möbius strip is, in effect, a 2-to-one map. This map can easily be described mathematically. On $S^1 \times [0,1]$ define an equivalence relation $\sim$ by

$$(x,y,t) \sim (-x,-y,1-t).$$

This becomes slightly more elegant if we replace $[0,1]$ by $[-1,1]$ and define the equivalence relation by

$$(x,y,t) \sim (-x,-y,-t).$$

I claim that the quotient space Q is homeomorphic to the Möbius strip. Since each equivalence class of the relation $\sim$ consists of two points, the map $S^1 \times [-1,1] \to Q$ is 2-to-1, so in this way we get a 2-to-1 map of $S^1 \times [-1,1]$ to the Möbius strip.



To see that Q is homeomorphic to the Möbius strip, consider the restriction of q to $S^1_+ \times [-1,1]$, where S^1_+ is the semicircle $\{(x,y) \in S^1 : y \geq 0\}$. This restriction maps the back half of the cylinder onto Q. Now S^1_+ is homeomorphic to $[0,1]$, so we have a composite surjection

$$h : [0,1] \times [-1,1] \simeq S^1_+ \times [-1,1] \xrightarrow{q} Q.$$

The only points that are identified by h are points of the form $(0,y)$ and $(1,-y)$. Thus, Q is the quotient of the square $[0,1] \times [-1,1]$ by the same equivalence relation used to define the Möbius strip.

Chapter 4

The fundamental group

Let X be a topological space and let $x_0 \in X$. A *path* in X is a continuous map $p : [0,1] \to X$. A path p is a *loop* if $p(0) = p(1)$, and is a *loop based at* x_0 if $p(0) = p(1) = x_0$. If p_0 and p_1 are paths in X with $p_0(1) = p_1(0)$, one obtains a new path $p_0 * p_1$ by first performing p_0 and then performing p_1, both at double speed:

$$(p_0 * p_1)(t) = \begin{cases} p_0(2t) & \text{if } t \in [0, 1/2] \\ p_1(2t-1) & \text{if } t \in [1/2, 1] \end{cases}$$

Notice that this is well defined at $t = 1/2$ because both formulas on the right hand side give the same value for $t = 1/2$, by the assumption that $p_0(1) = p_1(0)$.

If p_0 and p_1 are both loops in X based at x_0 then $p_0 * p_1$ is always defined, and is again a loop based at x_0.

Clearly in general $p_0 * p_1 \neq p_1 * p_0$. This "product" does not allow us to define a group yet: there is no neutral element, and also associativity fails, since while $p*(q*r)$ performs p at double speed in the interval $[0, 1/2]$, then q at quadruple speed in $[1/2, 3/4]$ and r at quadruple speed in $[3/4, 1]$, instead $(p * q) * r$ performs p at quadruple speed in $[0, 1/4]$, then q at quadruple speed in $[1/4, 1/2]$, and then r at double speed in $[1/2, 1]$.

To remedy these problems we introduce an equivalence relation in the set of loops based at x_0. In fact it is useful to widen the scope slightly. Suppose p_0 and p_1 are two paths joining the same pair of points x_0 and x_1: $p_0(0) = p_1(0) = x_0$, $p_0(1) = p_1(1) = x_1$. We say p_0 and p_1 are *homotopic* if there is a continuous map $P : [0,1] \times [0,1] \to X$ such that

1. $P(t, 0) = p_0(t), P(t, 1) = p_1(t)$for all $t \in [0, 1]$,

   and *end-point preserving homotopic* if also

2. $P(0,u)$ and $P(1,u)$ are independent of u, so that for all u the path p_u defined by $p_u(t) = P(t,u)$ is a path from x_0 to x_1.

Now suppose that p_0 and p_1 are loops based at x_0. Then Condition 2 means that for each $u \in [0,1]$, the path $p_u : [0,1] \to X$ is a loop based at x_0; Condition 1, together with the continuity of P, says that the family of loops $\{p_u : u \in [0,1]\}$ is a continuous deformation from p_0 to p_1.

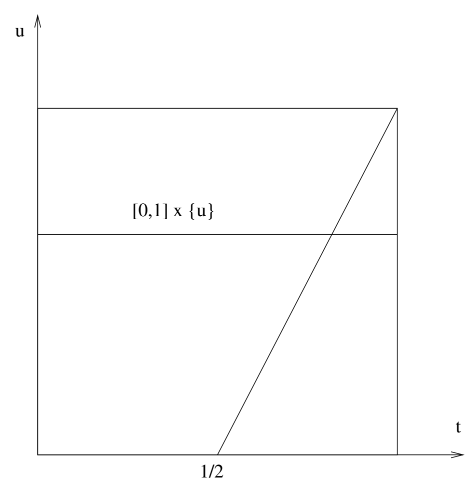
We will refer to such a map as an end-point-preserving homotopy (e.p.p.h.) from p_0 to p_1.

Define a constant path e by $e(t) = x_0$ for all t. Clearly e is a loop, though not a very interesting one.

Example 4.0.2. The loops $p * e$ and $e * p$ are both end-point preserving homotopic to p. To see this for $p * e$, we define an end-point-preserving homotopy between $p * e$ and p by

$$P(t,u) = \begin{cases} p(2t/1+u) & \text{if } t \in [0,(1+u)/2] \\ e(t) & \text{if } t \in [(1+u)/2,1] \end{cases}$$

The square shows the domain of P; the interval $[0,1] \times \{u\}$ can be thought of as the domain of p_u.



The map P is well-defined, for the two formulas used to define it coincide when both apply (i.e. when $t = (1+u)/2$): both give the value x_0 at such points. To see it is continuous, we observe that it is clearly continuous on the subspaces

$$X_1 = \{(t,u) \in [0,1] \times [0,1] : 0 \leq t \leq (1+u)/2\}$$

and

$$X_2 = \{(t, u) \in [0, 1] \times [0, 1] : (1 + u)/2 \leq t \leq 1\},$$

and invoke the following Lemma:

Lemma 4.0.3. *Suppose that the topological space X is the union of closed subspaces X_i, $i = 1, \ldots, n$, and that there exist continuous maps $f_i : X_i \to Y$ with the property that if $x \in X_i \cap X_j$ then $f_i(x) = f_j(x)$. Then the map $f : X \to Y$ defined by*

$$f(x) = f_i(x) \text{ if } x \in X_i$$

is continuous.

Proof. A map is continuous if and only if the preimage of every open set is open. This holds if and only if the preimage of every closed set is closed. Let $V \subset Y$ be closed. Then $f^{-1}(V) = \bigcup_{i=1}^{n} f_i^{-1}(V)$. Each $f_i^{-1}(V)$ is closed in X_i, as $f_i : X_i \to Y$ is continuous. Because X_i is closed in X, it follows that $f_i^{-1}(V)$ is closed in X. Hence $f^{-1}(V)$, as the union of a finite number of closed subsets of X, is closed in X.

□

I leave to the reader the proof that $e*p$ is end-point-preserving homotopic to p.

We will construct many homotopies $[0, 1] \times [0, 1] \to X$ by dividing up the domain into a number of regions, defining F on each region by a simple formula from which it is apparent that F is continuous on the region, and then checking (or leaving to the reader to check) that the different formulae agree on the overlaps between the regions. Continuity will then follow from Lemma 4.0.3.

Lemma 4.0.4. *The relation of end-point-preserving homotopy is an equivalence relation.*

Proof. Reflexivity is obvious; symmetry is shown by observing that if P is an e.p.p.h. from p_0 to p_1 then the map $\tilde{P}$ defined by $\tilde{P}(t, u) = P(t, 1 - u)$ is an e.p.p.h. from p_1 to p_0. For transitivity, suppose that P is an e.p.p.h from p_0 to p_1 and Q is an e.p.p.h. from p_1 to p_2. Define $PQ : [0, 1] \times [0, 1] \to X$ by

$$PQ(t, u) = \begin{cases} P(t, 2u) & \text{if } 0 \leq u \leq 1/2 \\ Q(t, 2u - 1) & \text{if } 1/2 \leq u \leq 1. \end{cases}$$

It is easy to see that PQ is an e.p.p.h. from p_0 to p_2. □

For each loop p based at x_0, denote by $[p]$ its equivalence class. We denote by $\pi_1(X, x_0)$ the set of equivalence classes of loops in X based at x_0.

Proposition 4.0.5. *The operation of composition of loops respects the equivalence relation. That is, if $p_0 \sim p_1$ and $q_0 \sim q_1$ then $p_0 * q_0 \sim p_1 * q_1$.*

Proof. Suppose that P and Q are epph's from p_0 to p_1 and q_0 to q_1. Define an epph PQ from $p_0 * q_0$ to $p_1 * q_1$ by

$$PQ(t,u) = \begin{cases} P(2t, u) & \text{if } t \in [0, 1/2] \\ Q(2t-1, u) & \text{if } t \in [1/2, 1] \end{cases}$$

□

The idea of the proof is that P is a family of loops p_u beginning with p_0 and ending with p_1, and Q is a family of loops beginning with q_0 and ending with q_1, so therefore $p_u * q_u$ is a family of loops starting with $p_0 * q_0$ and ending with $p_1 * q_1$.

Lemma 4.0.5 allows us to define a binary operation, which we will continue to denote by $*$, on the set of equivalence classes of loops based at x_0. We define

$$[p_1] * [p_2] = [p_1 * p_2]. \tag{4.0.1}$$

In case (4.0.1) appears to work without the need for Lemma 4.0.5, we can write it slightly differently:

> to define the product of two equivalence classes A and B, choose a loop p in A and a loop q in B, and define AB to be the equivalence class of $p * q$.

This version makes it clearer that for the definition to work, the equivalence class of $p*q$ must not depend on the choice of $p \in A$ and $q \in B$, i.e. that any other choice would give the same result. That is what Lemma 4.0.5 says.

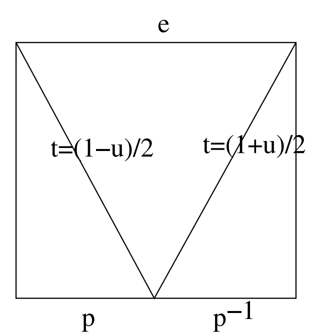
Theorem 4.0.6. *$\pi_1(X, x_0)$, with the binary operation $*$, is a group.*

Proof. Only three steps are needed.

1. *Neutral element* By Example 4.0.2, $[e] * [p] = [p] * [e] = [p]$, so $[e]$ is a neutral element. That it is the only neutral element follows from the existence of inverses, which we prove next.

2. *Inverses* The inverse of $[p]$ is the equivalence class of the loop p^{-1} defined by $p^{-1}(t) = p(1-t)$. To see that $[p] * [p^{-1}] = [e] = [p^{-1}] * [p]$, we have to construct suitable epph's. For the first equality, define

$$P(t,u) = \begin{cases} p(2t) & \text{if } 0 \le t \le \frac{1-u}{2} \\ p(1-u) & \text{if } \frac{1-u}{2} \le t \le \frac{1+u}{2} \\ p^{-1}(2t-1) & \text{if } \frac{1+u}{2} \le t \le 1 \end{cases}$$



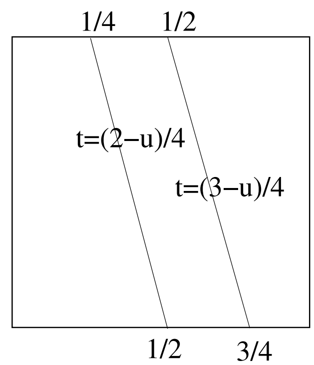
You should check that the three definitions agree on the overlaps. I leave as an exercise the proof that $p^{-1} * p$ is epph to e.

3. *Associativity* We have to show that $p * (q * r)$ is epph to $(p * q) * r$. Consider the following square. The lower edge is divided into three subintervals, corresponding to the definition of $p * (q * r)$,

$$p * (q * r)(t) = \begin{cases} p(2t) & \text{if } t \in [0, 1/2] \\ q(4t-2) & \text{if } t \in [1/2, 3/4] \\ r(4t-3) & \text{if } t \in [3/4, 1] \end{cases}$$

and the upper edge is divided into three subintervals, this time corresponding to the definition of $(p * q) * r$. Define

$$P(t,u) = \begin{cases} p(\frac{4t}{2-u}) & \text{if } 0 \le t \le (2-u)/4 \\ q(4t-2+u) & \text{if } (2-u)/4 \le t \le (3-u)/4 \\ r(\frac{4t-3+u}{1+u}) & \text{if } (3-u)/4 \le t \le 1 \end{cases}$$



The diagram shows the three regions involved in the definition of the epph P.

$\square$

Now that we know that $\pi_1(X, x_0)$ is a group, we should calculate it for some examples.

Example 4.0.7. 1. For any $x_0 \in \mathbb{R}^n$, $\pi_1(\mathbb{R}^n, x_0)$ is the trivial group $\{1\}$. For every loop p based at x_0 is epph to the constant loop e: the epph

$$P(t, u) = x_0 + (1 - u)(p(t) - x_0)$$

shrinks p to e by pulling each point $p(t)$ back towards x_0 along the ray joining it to x_0.

2. A topological space X is *contractible* (to a point $x_0 \in X$) if there exists a homotopy $F : X \times [0, 1] \to X$ such that for all $x \in X$, $F(x, 0) = x$ and $F(x, 1) = x_0$.

Proposition 4.0.8. *If X is contractible to x_0 then $\pi_1(X, x_0) = \{1\}$*

Proof. We prove this for the special case where $F(x_0, u) = x_0$ for all $u \in [0, 1]$. In this case, for any $p \in L(X, x_0)$, we define an epph from p to the constant loop e simply by

$$P(t, u) = F(p(t), u).$$

This shows that $[p] = [e]$ so that $\pi_1(X, x_0) = \{[p] : p \in L(X, x_0)\} = \{[e]\}$.

If $F(x_0, u)$ is not necessarily constant, then the path

$$p_u(t) = F(p(t), u)$$

is a loop, but based at $F(x_0, u)$ instead of at x_0. Of course, p_1 *is* a loop based at x_0, namely the constant loop e.

For $u \in [0, 1]$ let γ_u be the path $\gamma_u(t) = F(x_0, tu)$. Observe that $\gamma_u(0) = x_0$ and $\gamma_u(1) = F(x_0, u)$, and that in particular $\gamma_0(t) = x_0$ for all t (so $\gamma_0 = e$).

Consider the path $q_u = (\gamma_u * p_u) * g_u^{-1}$. This first travels from x_0 to $F(x_0, u)$, then follows the loop p_u, then travels back to x_0. So it is a loop based at x_0. Moreover

$$q_0 = (e * p) * e^{-1} = (e * p) * e$$

and

$$q_1 = (\gamma_1 * p_1) * \gamma_1^{-1} = (\gamma_1 * e) * \gamma_1^{-1}$$

Thus the map

$$Q(t, u) = q_u(t)$$

is an epph from $(e * p) * e$ to $(\gamma_1 * e) * \gamma_1^{-1}$. Since p is epph to $(e * p) * e$ and $(\gamma_1 * e) * \gamma_1^{-1}$ is epph to $\gamma_1 * \gamma_1^{-1}$ and therefore to e, it follows by the transitivity of the relation epph that p is epph to e. □

In fact if X is contractible to x_0 and $x_1 \in X$ then X is contractible to x_1. For contractibility implies path conectedness (**Exercise**), and by following a map F contracting X to x_0 with a path from x_0 to x_1, we get a contraction to x_1. In more detail, if p is a path in X from x_0 to x_1 and $F : X \times [0, 1] \to X$ is a contraction to x_0, then the map $G : X \times [0, 1] \to X$ defined by

$$G(x, t) = \begin{cases} F(x, 2u) & \text{if } u \in [0, 1/2] \\ p(2u - 1) & \text{if } u \in [1/2, 1] \end{cases}$$

contracts X to x_1.

3. $\pi_1(S^2, x_0) = \{1\}$ even though S^2 is not contractible. This is pretty obvious visually, but proving it requires a bit of effort. We use the fact that for any $y \in S^2$, $S^2 \smallsetminus \{y\}$ is homeomorphic to $\mathbb{R}^2$ (Exercise I.3). Let p be any loop based at x_0. If p is not surjective, let y be a point of S^2 not in the image of p. Let $\varphi : S^2 \smallsetminus \{y\} \to \mathbb{R}^2$ be a

homeomorphism. Let $P : [0,1] \times [0,1] \to \mathbb{R}^2$ be an epph from $\varphi \circ p$ to the contant loop based at $\varphi(x_0)$. Then

$$\tilde{P}(t,u) = \varphi^{-1}(P(\varphi \circ p(t), u))$$

is an epph from p to the constant loop based at x_0. However we still have to deal with the possibility that p may be surjective – surjective continuous maps from $[0,1]$ to S^2 do exist. For this case, we show first that p is epph to a loop which is not surjective, and then reason as before. The proof that p is epph to a loop which is not surjective is left as a guided exercise in Exercises II Section C.

The alert reader will have noticed that so far we have only shown examples of spaces for which $\pi_1(X, x_0) = \{1\}$. We will now set out to remedy this.

4.1 Calculation of $\pi_1(S^1, x_0)$

Notation

The standard circle S^1 can be thought of as the set of complex numbers with modulus 1. As such it contains the complex number 1.

Intuition

Any loop $p : [0,1] \to S^1$ is epph to a loop which does not change direction (i.e. is either clockwise all the time, or anticlockwise all the time). We will call such a loop *monotone*. This can be seen intuitively: imagine a loop made of some elastic material, wound round a smooth circular wire, along which the loop can slide. If the loop changes direction at any point then it can be deformed until it is monotone. This can be accomplished by drawing the loop tight: as the tension increases, the wrinkles are smoothed out. In fact the same physical argument suggests that any loop is epph to a constant speed monotone loop: if the elastic loop is free to slide around the circle, the tension will even itself out. The constant speed monotone loops are all of the form

$$p_n(t) = \exp(nt);$$

(when n is positive, p_n winds n times anticlockwise round the circle. If n is negative, it winds n times clockwise). Since, as we have sketched, each loop

p is epph to one of these, there is a bijection $\pi_1(S^1, 1) \to \mathbb{Z}$. Furthermore, $[p_n] * [p_m] = [p_{n+m}]$, so the bijection is an isomorphism.

Our proof more or less follows these lines. To begin with, recall the map $\exp : \mathbb{R} \to S^1$ defined by $\exp(t) = e^{2\pi i t}$.

Proposition 4.1.1. *Let $X \subset \mathbb{R}^n$ be a compact set with the property that $0 \in X$ and for all $x \in X$, the line segment $[0, x]$ is entirely contained in X. Suppose that $f : X \to S^1$ is continuous, and $f(0) = 1$. Then there exists a continuous map $\tilde{f} : X \to \mathbb{R}$ such that $\exp \circ \tilde{f} = f$ and $\tilde{f}(0) = 0$.*

Proof. As X is compact, f is uniformly continuous and so there exists $\delta > 0$ such that if $|x - x'| < \delta$ then $|f(x) - f(x')| < 2$, and in particular $f(x) \neq -f(x')$ (recall that we are on the unit circle). Compactness also implies that X is bounded so there exists $N \in \mathbb{N}$ such that $X \subset B(0, \delta N)$, and thus for all $x \in X$, $\|\frac{x}{N}| < \delta$ and hence also $\|jx/N - (j-1)x/N\| < \delta$ for $j \in \mathbb{N}$. It follows that

$$f\left(j\frac{x}{N}\right) \neq -f\left((j+1)\frac{x}{N}\right).$$

For $j = 0, \ldots, N-1$ define

$$h_j(x) = f\left((j+1)\frac{x}{N}\right)\left(f\left(j\frac{x}{N}\right)\right)^{-1}.$$

We have $h_j(0) = 1$ for all j. By the above, $h_j(x) \neq -1$ for any j and any $x \in X$. The map $(-1/2, 1/2) \xrightarrow{\exp} S^1 \smallsetminus \{-1\}$ is a homeomorphism with inverse which we will call ℓ. Note that $\ell(1) = 0$. Define $\varphi_j : X \to \mathbb{R}$ by

$$\varphi_j = \ell \circ h_j.$$

We have $\varphi_j(0) = \ell(1) = 0$ for all j.

Define

$$\tilde{f}(x) = \varphi_0(x) + \cdots + \varphi_{N-1}(x).$$

Then

$$\begin{aligned}
\exp(\tilde{f}(x)) &= \exp\big(\varphi_0(x) + \cdots + \varphi_{N-1}(x)\big) \\
&= \exp(\varphi_0(x)) \times \cdots \times \exp(\varphi_{N-1}(x)) \\
&= \exp(\ell(h_0(x)) \times \cdots \times \exp(\ell(h_{N-1}(x)) \\
&= h_0(x) \times \cdots \times h_{N-1}(x) \\
&= \frac{f(x/N)}{f(0)} \times \frac{f(2x/N)}{f(x/N)} \times \cdots \times \frac{f(Nx/N)}{f((N-1)x/N)} \\
&= f(x) \qquad (4.1.1)
\end{aligned}$$

$\square$

The map $\tilde{f}$ is called a *lift* of f; the reason is explained by the following diagram:

$$\begin{tikzcd} & \mathbb{R} \arrow[d, "\exp"] \\ X \arrow[ur, "\tilde{f}"] \arrow[r, "f"'] & S^1 \end{tikzcd}$$

Example 4.1.2. Let $p_n : [0,1] \to S^1$ be the loop defined by $p_n(t) = \exp(nt)$. Then as lift we can take $\tilde{p}_n(t) = nt$. If we took $\tilde{p}(t) = nt + m$ for any fixed $m \in \mathbb{Z}$, then we would still have $\exp \circ \tilde{p}_n = p_n$, but the condition that $\tilde{p}_n(0) = 0$ of course only holds if $m = 0$.

Corollary 4.1.3. *With X as in Proposition 4.1.1, suppose that $f : X \to S^1$ is continuous, and suppose that $f(0) = \exp(a)$. Then there exists a unique lift $\tilde{f} : X \to \mathbb{R}$ such that $\exp \circ \tilde{f} = f$ and $\tilde{f}(0) = a$.*

Proof. Let $f_0(x) = f(x)/f(0)$, and let $\tilde{f}_0$ be a lift of f_0 as constructed in 4.1.1. Recall that $\tilde{f}_0(0) = 0$. Define $\tilde{f}(x) = \tilde{f}_0(x) + a$. Then

$$\exp(\tilde{f}(x)) = \exp(\tilde{f}_0(x)) \exp(a) = f_0(x) \times f(0) = f(x).$$

If $\tilde{g}$ is another lift of f also satisfying $\tilde{g}(0) = a$, then since $\exp(\tilde{f}(x)) = \exp(\tilde{g}(x))$ for all x, $\tilde{f}(x) - \tilde{g}(x) \in \mathbb{Z}$ for all x. Since $\tilde{f}(0) = \tilde{g}(0)$, $\tilde{f}$ and $\tilde{g}$ must coincide everywhere. $\square$

Note that of course, if $\tilde{f} : X \to \mathbb{R}$ is a lift of $f : X \to S^1$, then necessarily $\tilde{f}(0) \in \exp^{-1}(f(0))$. The corollary says that there *is* a lift starting at any point in $\exp^{-1}(f(0))$, and only one: once we fix the starting point, the lift is unique.

Definition 4.1.4. Let p be a loop in S^1, and let $\tilde{p}$ be a lift of p, as in the corollary. Define the *degree* of p to be $\tilde{p}(1) - \tilde{p}(0)$.

If $\tilde{p}$ and $\tilde{\tilde{p}}$ are two lifts of the same $p : X \to S^1$, then since $\exp(\tilde{p}(x)) = \exp(\tilde{\tilde{p}}(x))$ for all $x \in X$, it follows that $\tilde{p}(x) - \tilde{\tilde{p}}(x) \in \mathbb{Z}$ for all x. Since X is connected, $\tilde{p} - \tilde{\tilde{p}}$ must therefore be constant. So when $X = [0,1]$ and p is a loop in S^1 based at x_0, then $\tilde{p}(1) - \tilde{p}(0) = \tilde{\tilde{p}}(1) - \tilde{\tilde{p}}(0)$ and $\deg(p)$ is well defined.

Now we show that, as claimed in the intuitive introduction to this section, "wrinkles can be smoothed and tension evened":

Proposition 4.1.5. *Any loop p in S^1 based at 1 is epph to a constant speed monotone loop – i.e. to a loop $p_n(t) = \exp(nt)$, for some $n \in \mathbb{Z}$.*

Proof. Let $\tilde{p}$ be a lift of p with $\tilde{p}(0) = 0$. Then $\tilde{p}(1) = n$ for some $n \in \mathbb{Z}$. Let $g_n(t) = nt$, so that $\exp \circ g_n = p_n$. We will construct an epph from $\tilde{p}$ to g_n; by composing it with exp, we will get an epph from p to p_n.

Define $R(t, u) = (1-u)\tilde{p}(t)+ug_n(t)$. Then R is continuous, $R(t, 0) = \tilde{p}(t)$ and $R(t, 1) = g_n(t)$. Also since $\tilde{p}(0) = g_n(0)$ and $\tilde{p}(1) = g_n(1)$, R itself is an epph. It follows that $\exp \circ R$ is an epph from $p = \exp \circ \tilde{p}$ to $\exp \circ g_n = p_n$. □

Proposition 4.1.6. *Let p_0 and p_1 be loops in S^1 based at x_0. If they are epph, then $deg(p_0) = deg(p_1)$.*

Proof. Let $P : [0, 1] \times [0, 1] \to S^1$ be an epph from p_0 to p_1. Applying Corollary 4.1.3 to P, we obtain a lift $\tilde{P} : [0, 1] \times [0, 1] \to \mathbb{R}$, such that $\exp \circ \tilde{P} = P$. It follows that $\tilde{p}_0$ and $\tilde{p}_1$ (defined, of course, by $\tilde{p}_0(t) = \tilde{P}(t, 0)$ and $\tilde{p}_1(t) = \tilde{P}(t, 1)$) are lifts of p_0 and p_1. Thus

$$\deg(p_0) = \tilde{P}(1, 0) - \tilde{P}(0, 0) \text{ and } \deg(p_1) = \tilde{P}(1, 1) - \tilde{P}(0, 1).$$

But for all u, $P(0, u) = x_0 = P(1, u)$ (P is an epph) and it follows that $\tilde{P}(0, u)$ and $\tilde{P}(1, u)$ both lie in the discrete set $\exp^{-1}(x_0)$. Therefore $\tilde{P}(0, u)$ and $\tilde{P}(1, u)$ are constant, and in particular

$$\tilde{P}(1, 0) - \tilde{P}(0, 0) = \tilde{P}(1, 1) - \tilde{P}(0, 1).$$

□

Theorem 4.1.7. *$deg : \pi_1(S^1, x_0) \to \mathbb{Z}$ is an isomorphism.*

Proof. This theorem summarises quite a lot of information.

1. Proposition 4.1.6 shows that deg is well defined on epph classes, so gives a well defined map $\pi_1(S^1, 1) \to \mathbb{Z}$.

2. The fact that $\deg(p_n) = n$ shows that deg is surjective.

3. Proposition 4.1.5 shows that deg is injective. For if $\deg f_0 = \deg f_1 = n$ then both f_0 and f_1 are epph to p_n, and hence to one another, so that $[f_0] = [f_1]$ in $\pi_1(S^1, 1)$.

4. To see that $\deg(p * q) = \deg(p) + \deg(q)$, suppose that $\tilde{p}$ is a lift of p with $\tilde{p}(0) = 0$. We choose a lift $\tilde{q}$ of q starting not at 0, but at $\tilde{p}(1)$, as we may, by Corollary 4.1.3. The path $r := \tilde{p}*\tilde{q}$ is a lift of $p*q$. Finally, $\deg(p * q) = r(1) - r(0) = \tilde{q}(1) - \tilde{q}(0) + \tilde{p}(1) - \tilde{p}(0) = \deg(p) + \deg(q)$.

□

4.2 How much does $\pi_1(X, x_0)$ depend on the choice of x_0?

Read this subsection later if you'd rather see some more concrete examples. We show that provided X is path-connected, the group $\pi_1(X, x_0)$ is independent of the choice of x_0, at least up to isomorphism. We will need the following generalisation of composition of loops.

Definition 4.2.1. Let a and b be paths in X, not necessarily loops. If $a(1) = b(0)$ then $a * b$ is the path defined by

$$a * b(t) = \begin{cases} a(2t) & \text{if } t \in [0, 1/2] \\ b(2t-1) & \text{if } t \in [1/2, 1]. \end{cases}$$

Note that we need $a(1) = b(0)$ in order for the two halves of the definition to agree on the overlap of their domains (when $t = 1/2$).

The argument used to prove associativity of $*$ in Theorem 4.0.6 shows:

Proposition 4.2.2. *Suppose a, b and c are paths in X with $a(1) = b(0)$ and $b(1) = c(0)$. Then the two paths $(a * b) * c$ and $a * (b * c)$ are epph to one another.* □

Suppose that there is a path a in X from x_0 to x_1. We use it to construct a map $\pi_1(X, x_1) \to \pi_1(X, x_0)$.

If p is a loop in X based at x_1 then $(a * p) * a^{-1}$ is a loop in X based at x_0. Here a^{-1} is, as usual, the path a traversed backwards: $a^{-1}(t) = a(1-t)$. If p_0 and p_1 are loops in X based at x_1, then it is easy to show that

$$[p_0] = [p_1] \text{ in } \pi_1(X, x_1) \implies [(a * p_0) * a^{-1}] = [(a * p_1) * a^{-1}] \text{ in } \pi_1(X, x_0).$$

So by this procedure we have defined a map

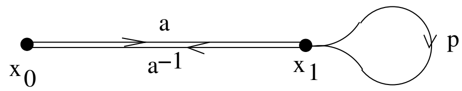
$$\pi_1(X, x_1) \to \pi_1(X, x_0),$$

which we will call c_a. We retain the name of the path a as a subindex, because different paths from x_0 to x_1 may give rise to different isomorphisms (see Exercises 2 13).

Note that because the $*$ operation between paths is associative, we could just as well have defined c_a by

$$[p] \mapsto [a * (p * a^{-1})];$$

we get the same homomorphism $\pi_1(X, x_1) \to \pi_1(X, x_0)$ with this version of the definition.



Proposition 4.2.3. $c_a : \pi_1(X, x_1) \to \pi_1(X, x_0)$ *is an isomorphism.*

Proof. Step 1: The map $c_{a^{-1}}$ is the inverse of c_a, so c_a is a bijection. To see this, observe that

$$\begin{aligned} c_{a^{-1}}(c_a([p])) = c_{a^{-1}}([(a * p) * a^{-1}]) &= [\big(a^{-1} * ((a * p) * a^{-1})\big) * a] \\ &= [\big((a^{-1} * a) * p\big) * (a^{-1} * a)] \end{aligned}$$

(here, and in what follows, we use the asociativity of $*$, Proposition 4.2). Now $a^{-1} * a$ is epph to the constant loop based at x_1, which we denote by e_{x_1}. Thus

$$[\big((a^{-1} * a) * p\big) * (a^{-1} * a)] = [(e_{x_1} * p) * e_{x_1}]$$

and this is equal to $[p]$ itself.
Step 2: $c_a([p] * [q]) = c_a([p]) * c_a([q])$. For

$$\begin{aligned} c_a([p]) * c_a([q]) &= [(a * p) * a^{-1}] * [(a * q) * a^{-1}] \\ &= [\big((a * p) * a^{-1}\big) * \big((a * q) * a^{-1}\big)] \\ &= [\big((a * p) * (a^{-1} * a)\big) * (q * a^{-1})] \end{aligned}$$

Since $a * a^{-1}$ is epph to the constant loop e_{x_0}, we have

$$\begin{aligned} [\big((a * p) * (a^{-1} * a)\big) * (q * a^{-1})] &= [\big((a * p) * e_{x_0}\big) * (q * a^{-1})] \\ &= [(a * p) * (q * a^{-1})] \\ &= [(a^*(p * q)) * a^{-1}] \\ &= c_a([p] * [q]). \end{aligned}$$

□

If you are irritated by the fussing with parentheses, you can adopt the following

Convention In any iterated $*$ product inside square brackets (and therefore denoting a epph class) which does not involve compound inverses, parentheses can be dispensed with.

This is justified by Proposition 4.2. By "compound inverses" we mean expressions of the form $(p * q)^{-1}$. Here the brackets are important: if we remove them we have to change the order of the terms (Exercises II 2(ii)).

4.3 The degree of a map $S^1 \to S^1$ and the Fundamental Theorem of Algebra

Let $f : S^1 \to S^1$ be continuous. By composing f with $\exp : [0,1] \to S^1$ we get a loop in S^1. Let $f : S^1 \to S^1$ be a continuous map. Define the *closed degree* [1] of f to be the degree of the loop $f \circ \exp : [0,1] \to S^1$.

Example 4.3.1. The closed degree of $P_n(z) = z^n$ is n.

Proposition 4.3.2. *The closed degree of a continuous map $S^1 \to S^1$ is invariant under homotopy.*

Proof. Suppose that $F : S^1 \times [0,1] \to S^1$ is a homotopy between f_0 and f_1, and define $f_u : S^1 \to S^1$ by $f_u(z) = F(z,u)$. The map $G : [0,1] \times [0,1] \to S^1$ defined by $G(t,u) = F(\exp(t),u)$ is a homotopy, but not an epph, since although the end points $G(0,u)$ and $G(1,u)$ are equal, they may vary as u varies. To fix this we replace F by a new homotopy $\hat{F}$ which ensures that for each u, $f_u(1) = 1$:

$$\hat{F}(z,u) = F(1,u)^{-1} F(z,u).$$

The effect of multiplying by $F(1,u)^{-1}$ is to rotate the whole image round so that 1 is always mapped to 1.

By construction, $\hat{G}(t,u) = \hat{F}(\exp(t),u)$ is an epph. Hence

$$\deg(\hat{g}_0) = \deg(\hat{g}_1), \quad \text{where } \hat{g}_u(t) = \hat{G}(t,u)$$

and so

$$\mathrm{cldeg}(\hat{f}_0) = \mathrm{cldeg}(\hat{f}_1), \quad \text{where } \hat{f}_u(t) = \hat{F}(t,u).$$

All we need now is to show that $\mathrm{cldeg}(\hat{f}_0) = \mathrm{cldeg}(f_0)$ and $\mathrm{cldeg}(\hat{f}_1) = \mathrm{cldeg}(f_1)$. But $\hat{g}_u$ and g_u differ only by a rotation $g_u(1)^{-1}$, so this is obvious. $\square$

Proposition 4.3.3. *If the closed degree of $f : S^1 \to S^1$ is n then f is homotopic to P_n.*

Proof. This is essentially Proposition 4.1.5. There's just one small and irritating difficulty, which is that when we work with loops we use end-point-preserving homotopy (epph), whereas with maps $S^1 \to S^1$ we simply use homotopy (there are no end points to refer to).

[1] The standard name for this is simply the degree of f. We use "closed degree" to avoid confusion with the degree of a loop.

Suppose first that $f : S^1 \to S^1$ has closed degree n and that $f(1) = 1$. Then $f \circ \exp$ has degree n, so $f \circ \exp$ and p_n are epph, by Prop 4.1.5. Let

$$F : [0, 1] \times [0, 1] \to S^1$$

be an epph from $f \circ \exp$ to p_n. Then $F(0, u) = F(1, u) = 1$ for all u, so F passes to the quotient to define a homotopy

$$G : S^1 \times [0, 1] \to S^1$$

from f to the map P_n defined by $P_n(z) = z^n$ (note that $P_n \exp = p_n$).

The general case (where we do not insist that $f(1) = 1$) is dealt with by noting that any $f : S^1 \to S^1$ is homotopic to a map g for which $g(1) = 1$: suppose that $f(1) = \exp(c)$ for some $c \in [0, 1]$. Then define a homotopy H from f to a map g for which $g(1) = 1$, by $H(z, u) = exp(-cu)f(z)$. The previous argument applies to g.

Thus f is homotopic to g, and g is homotopic to P_n. So f is homotopic to P_n. □

The circle S^1 is the boundary of the closed unit disc $\overline{B(0, 1)}$. We say that $f : S^1 \to Y$ extends to a continuous map $\overline{B(0, 1)} \to Y$ if there exists a continuous map $F : \overline{B(0, 1)} \to Y$ such that the restriction of F to S^1 is f.

Corollary 4.3.4. *If the continuous map $f : S^1 \to S^1$ extends to a continuous map $F : \overline{B(0, 1)} \to S^1$ then f is homotopic to a constant map and $cldeg(f) = 0$.*

Proof. Define a homotopy G by

$$G(z, u) = F(uz).$$

Clearly $G(z, 1) = F(z) = f(z)$ for $z \in S^1$, and $G(z, 0) = F(0)$ □

Theorem 4.3.5. *Let $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + \alpha_1 z + a_0$ be a complex polynomial of degree $n > 0$. Then p has a root in $\mathbb{C}$.*

Proof. We may divide through by a_n and so assume that $a_n = 1$. Let $p_u(z) = z^n + u(a_{n-1} z^{n-1} + \cdots + a_0)$. So p_1 is our polynomial p, and $p_0(z) = z^n$.

We now choose $R \in \mathbb{R}$ so large that when $u \in [0, 1] \subset \mathbb{R}$, p_u can have no root on the circle of radius R with centre 0. This is easy enough: if

$$R > \max\{1, |a_{n-1}| + |a_{n-2}| + \cdots + |a_0|\}$$

then when $|z| \geq R$,

$$z^n > |u \sum_k |a_k||z^k| \geq |u \sum_k a_k z^k|$$

and so $p_u(z) \neq 0$.

Now consider the map $f_u : S^1 \to S^1$ defined by

$$f_u(z) = \frac{p_u(Rz)}{\|p_u(Rz)\|}.$$

By our definition of R, the numerator and denominator are never 0 for $u \in [0,1]$, so f_u is well defined and continuous. Now $f_0(z) = z^n$, so $\mathrm{cldeg}(f_0) = n$. Because f_1 is homotopic to f_0 (by the family f_u), we have $\mathrm{cldeg}(f_1) = n$ also. This proves that p must have a root in the disc $B(0,R)$, for otherwise the map f_1 coud be extended to the unit disc $B(0,1)$ by the formula

$$F(z) = \frac{p(Rz)}{\|p(Rz)\|}.$$

□

4.4 Winding numbers

Suppose that $p : [0,1] \to \mathbb{R}^2$ is a loop and that z_0 is a point not in the image of p. We will define the *winding number* of p with respect to z_0, $W_p(z_0)$, which measures the number of times p winds round z_0 in the anticlockwise direction. It is defined as follows: since $p(t) \neq z_0$ for all $t \in [0,1]$, it follows that for each t,

$$p_{z_0}(t) := \frac{p(t) - z_0}{\|p(t) - z_0\|}$$

is a well defined loop in S^1. We define

$$W_p(z_0) = \deg(p_{z_0}).$$

There are several points of interest in this definition. It seems intuitively clear that if z_0 and z_1 are in the same connected component of the complement of the loop p, then p must wind round z_0 and z_1 the same number of times. The technology we have makes it possible to prove this. Since the image of the loop p is compact, it is closed in $\mathbb{R}^2$, so its complement is open, and therefore so is each connected compponent of the complement. Therefore each connected component of the complement is path connected.

Let γ be a path in the complement of the image of p, with $\gamma(0) = z_0$ and $\gamma(1) = z_1$. The map

$$P(t,u) = p_{\gamma(u)}(t) = \frac{p(t) - \gamma(u)}{\|p(t) - \gamma(u)\|}$$

is a homotopy of p_{z_1} to p_{z_2}. It does not fix end-points, unfortunately, but this can be fixed by using a trick like the one used in the proof of Proposition 4.3.2, replacing $P(t,u)$ by $Q(t,u) = P(1,u)^{-1}P(t,u)$. This is an epph, so $\deg(q_0) = \deg(q_1)$. As q_u is just p_u, rotated by multiplying by the constant $p_u(1)^{-1}$, $\deg(p_u) = \deg(q_u)$ for all u.

Thus the winding number of the loop p defines an integer-valued function on the set of connected components of the complement of the image of p.

Example 4.4.1. One might imagine that if $U \subset \mathbb{C}$ is an open set and if p, q are loops in U based at x_0, such that

$$W_p(y) = W_q(y)$$

for all $y \notin U$, then $[p] = [q]$ in $\pi_1(U, x_0)$. But it is not so! **Exercise** Find a counterexample. Once you have one, it will be clear how to construct many. Hint: Example 4.6.1(4) below is relevant here.

A second interesting point is that if the loop p is a differentiable map then the winding number can be computed by an integral formula from complex analysis:

$$W_p(z) = \frac{1}{2\pi i}\int_p \frac{dz}{z - z_0}.$$

I leave this as an exercise in Exercises 3.

4.5 Induced homomorphisms

If $f : X \to Y$ is continuous and $f(x_0) = y_0$ then each loop in X based at x_0 can be composed with f to give a loop in Y based at y_0. Epph's also can be composed with f, so if $p_0 \sim p_1$ then $f \circ p_0 \sim f \circ p_1$.Thus we get a well-defined map $f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ sending $[p]$ to $f_*([p]) := [f \circ p]$. All three parts of the following proposition are easy to prove.

Proposition 4.5.1. *(i)* $f \circ (p * q) = (f \circ p) * (f \circ q)$*, so* f_* *is a homomorphism.*

(ii) If $f : X \to Y$ and $g : Y \to Z$, with $f(x_0) = y_0$ and $g(y_0) = z_0$ then

$$g_* \circ f_* = (g \circ f) * .$$

(iii) If f is the identity map on X then f_* is the identity map on $\pi_1(X, x_0)$.

Corollary 4.5.2. *If $f : X \to Y$ is a homeomorphism sending x_0 to y_0 then $f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ is an isomorphism.*

Proof. By the proposition, $(f^{-1})_*$ is the inverse of f_*. □

4.6 Categories and Functors

Properties (i), (ii) and (iii) of 4.5.1 together amount to the fact that the assignment

$$(X, x_0) \to \pi_1(X, x_0)$$

is a *functor* from the category of topological spaces with base-points to the category of groups. We will not make use of this fact, but given the importance of the notion of functor, it is worth a brief discussion (but will not figure in the exam!).

A *category* consists of two things: a collection of objects, like, say, topological spaces, and a collection of "morphisms" between the objects, like, say, continuous maps. Our discussions so far have mostly concerned two categories: **Top**, whose objects are all topological spaces and whose morphisms are all continuous maps, and **Groups**, whose objects are all groups and whose morphisms are all group homomorphisms.

Other examples of categories are

1. **Ab**, with abelian groups as objects and homomorphisms (of abelian groups) as morphisms. Note that **Ab** is a subcategory of **Groups** in an obvious way.

2. **Rings**, with rings as objects and ring homomorphisms as morphisms.

3. $\mathbf{Vec}_{\mathbb{R}}$, with vector spaces over $\mathbb{R}$ as objects and linear maps as morphisms.

These are meant to obey certain rules; for example, the composition of two morphisms is meant to be a morphism. It certainly is in all of the examples mentioned above.

A *functor* F from category $\mathbf{A}$ to category $\mathbf{B}$ is a rule which assigns, to each object A of $\mathbf{A}$, an object $F(A)$ of $\mathbf{B}$, and to each morphism $f : A_1 \to A_2$

of **A** a morphism $F(f) : F(A_1) \to F(A_2)$ of **B**. This assignment should have the following two properties:

- F1: $F(g \circ f) = F(g) \circ F(f)$
- F2: $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$.

(which we have already met in Proposition 4.5.1). In fact there is a variant of this definition: sometimes it turns out that a morphism $X \to Y$ does not give rise to a morphism $F(X) \to F(Y)$, but instead to a morphism $F(Y) \to F(X)$. This is what happens, for example, with the functor C^0 which takes a topological space X and associates to it the ring $C^0(X) := \{f : X \to \mathbb{R} : f \text{ is continuous}\}$, with ring operations defined pointwise:

$$\begin{aligned} (f+g)(x) &= f(x) + g(x) \\ (fg)(x) &= f(x)g(x) \end{aligned}$$

If $\varphi : X \to Y$ is a continuous map then composition with φ defines a ring homomorphism $C^0(f) : C^0(Y) \to C^0(X)$, taking $f : Y \to \mathbb{R}$ to $f \circ \varphi : X \to \mathbb{R}$. This ring homomorphism is usually denoted φ^* rather than $C^0(\varphi)$. Instead of property F1 listed above, we have

- CF1: $(\psi \circ \varphi)^* = \varphi^* \circ \psi^*$

when $\varphi : X \to Y$ and $\psi : Y \to Z$ are continuous maps. (check this for yourself).

Functors which satisfy F1 and F2 are called *covariant functors*, while those which satisfy CF1 and F2 are *contravariant functors*

Example 4.6.1. 1. The example we have been most concerned with, π_1, is not, strictly speaking, a functor from **Top** to **Groups**. Instead, we have to concoct the category **PTop** of *topological spaces with base-point* (or "pointed topological spaces")

- The objects of **PTop** are pairs (X, x_0) where X is a topological space and x_0 is a point in X.
- The morphisms of **PTop** are continuous maps of pairs $(X, x_0) \to (Y, y_0)$, i.e. continuous maps $X \to Y$ taking x_0 to y_0.

Now π_1 really is a covariant functor from **PTop** to **Groups**. This is precisely the content of Proposition 4.5.1.

2. Suppose that V is a vector space over $\mathbb{R}$. Then so is $V^* := \mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R})$, the set of linear maps from V to $\mathbb{R}$. But now if $T : V \to W$ is a linear map, then its transpose, T^*, is a linear map $W^* \to V^*$. One can go on to check that indeed the process of taking the dual is a contravariant functor from $\mathbf{Vec}_{\mathbb{R}}$ to itself.

3. Algebraic topology studies a whole collection of covariant functors **Top** $\to$ **Ab**, the *homology functors* H_n, and a whole series of contravariant functors **Top** $\to$ **Ab**, the *cohomology functors* H^n.

4. Let G be a group. Its *commutator subgroup* $[G, G]$ is the subgroup generated by all elements $x^{-1}y^{-1}xy$ for $x, y \in G$. The product $x^{-1}y^{-1}xy$ is often written $[x, y]$ and is called a *commutator*. It is equal to the neutral element of G if and only if x and y commute – i.e. if and only if $xy = yx$. It is proved in group theory that $[G, G]$ is a normal subgroup, and that $G/[G, G]$ is an abelian group (**Exercise: prove these assertions.**) It is called the abelianisation of G, and we will denote it for now by $A(G)$.

Exercise (i) Show that any homomorphism $\varphi : G_1 \to G_2$ passes to the quotient to define a homomorphism $G_1/[G_1, G_1] \;\to\; G_2/[G_2, G_2]$.

(ii) Show that if we denote the quotient homomorphism in (i) by $A(\varphi)$, then $A(\varphi_1 \circ \varphi_2) = A(\varphi_1) \circ A(\varphi_2)$. Since it is obvious that A satisfies F2, i.e. that $A(\mathrm{id}_G) = \mathrm{id}_{A(G)}$, this proves that A is a covariant functor from **Groups** to **Ab**.

It is a theorem due to Poincaré that the abelianisation of $\pi_1(X, x_0)$ is in a natural way isomorphic to $H_1(X)$, the first homology group of the topological space X.

4.7 Homotopy Invariance

Definition 4.7.1. (i) Two continuous maps $f_0, f_1 : X \to Y$ are *homotopic* to one another if there is a continuous map $F : X \times [0, 1] \to Y$ such that $F(x, 0) = f_0(x),\ F(x, 1) = f_1(x)$.

Homotopy is an equivalence relation: reflexivity and symmetry are fairly obvious, and if F is a homotopy from f_0 to f_1 and G is a homotopy from f_1 to f_2 then by performing first F and then G, both at double speed with respect to the parameter $u \in [0, 1]$, one gets a homotopy from f_0 to f_2.

What is the effect of homotopy on the map $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$?

If f_0 and f_1 both map x_0 to y_0, and if $F(x_0, u) = y_0$ for all u, then for any loop p in X based at x_0, the map $P : [0,1] \times [0,1] \to Y$, defined by $P(t, u) = F(p(t), u)$, is an epph from $f_0 \circ p$ to $f_1 \circ p$, and it follows that $f_{0,*} : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ and $f_{1,*} : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ are equal.

But what if $f_0(x_0) \neq f_1(x_0)$?

Write $f_0(x_0) = y_0, f_1(x_0) = y_1$, and suppose $y_0 \neq y_1$. Clearly the path $u \mapsto F(x_0, u)$ runs from y_0 to y_1. Let us denote it by a. Recall from Proposition 4.2.3 that a determines an isomorphism $c_a : \pi_1(Y, y_1) \to \pi_1(Y, y_0)$, defined by

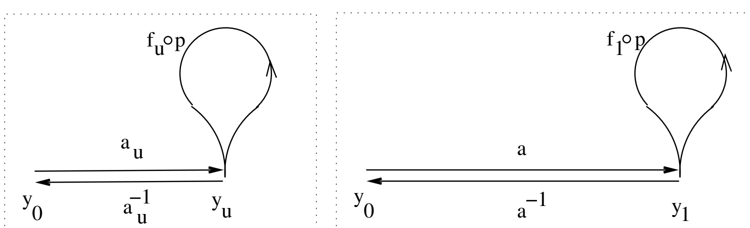
$$c_a([p]) = [(a * p) * a^{-1}].$$

Proposition 4.7.2. *If $F : X \times [0,1] \to Y$ is a homotopy between f_0 and f_1 then $f_{0*} : \pi_1(X, x_0) \to \pi_1(Y, f_0(x_0))$ and $f_{1*} : \pi_1(X, x_0) \to \pi_1(Y, f_1(x_0))$ are related by*

$$f_{0*} = c_a \circ f_{1*},$$

where a is the path in Y, from $f_0(x_0)$ to $f_1(x_0)$, defined by $a(t) = f_t(x_0)$.

Proof. If p is a loop in X based at x_0, then $f_u \circ p$ is a loop in Y based at $f_u(x_0) =: y_u$. For each $u \in [0,1]$, let $a_u : [0,1] \to Y$ be a path from y_0 to y_u, chosen so that the family of paths a_u varies continuously with u (e.g. $a_u(t) = a(ut)$).



Then the map

$$P(t, u) = \left(\left(a_u * (f_u \circ p)\right) * a_u^{-1}\right)(t)$$

is an epph from $f_0 \circ p$ to $\left(a * (f_1 \circ p)\right) * a^{-1}$. The epph class of $\left(a * (f_1 \circ p)\right) * a^{-1}$ is, by definition, $c_a(f_{1*}([p]))$. So $f_{0*}([p]) = c_a(f_{1*}([p]))$. $\square$

Definition 4.7.3. (ii) The map $f : X \to Y$ is a *homotopy equivalence* if there exists a continuous map $g : Y \to X$ such that $g \circ f$ is homotopic to the identity map on X, and $f \circ g$ is homotopic to the identity map on Y.

In this case we say that X and Y are homotopy-equivalent, or "have the same homotopy type". We call g a "homotopy inverse" to f, and vice versa.

Example 4.7.0. 1. The circle S^1 is homotopy equivalent to $\mathbb{R}^2 \setminus \{0\}$. Take $f : S^1 \to \mathbb{R}^2 \setminus \{0\}$ to be inclusion, and define $g : \mathbb{R}^2 \setminus \{0\} \to S^1$ by $g(x) = x/\|x\|$. Then $g \circ f = 1_X$ (so $g \circ f$ is homotopic to the identity map on X!), and we need only show that $f \circ g$ is homotopic to the identity on $\mathbb{R}^2 \setminus \{0\}$. We have $f \circ g(x) = g(x)$ since f is inclusion. Define $F(x,t) = (1-t)x + tg(x)$. Then $F(x,0) = x$ and $F(x,1) = g(x)$. As t varies from 0 to 1, $F(x,t)$ moves from x to $x/\|x\|$ along the line-segment joining them.

In a case like this, where $f : X \to Y$ is the inclusion of a subset and $g : Y \to X$ is a left inverse of f one calls g a *retraction* of Y to X, and X a *retract* of Y. If $F : Y \times [0,1] \to Y$ is a homotopy such that $F(x,0) = x$ and $F(x,1) = g(x)$, *and such that if* $x \in X$ *then* $F(x,t) = x$ *for all* t, one says that F is a *deformation retraction*, and X is a *deformation-retract* of Y.

2. **Exercise** Check that if X is a deformation-retract of Y then the inclusion of X in Y is a homotopy-equivalence.

3. S^1 is homotopy-equivalent to any annulus $\{x \in \mathbb{R}^2 : a \leq \|x\| \leq b\}$, where $0 < a \leq b$.

4. An argument like the one for the case $n = 2$ shows that S^{n-1} is a deformation-retract of $\mathbb{R}^n \setminus \{0\}$.

5. Let $p_1 \neq p_2$ be points in $\mathbb{R}^2$. Then $\mathbb{R}^2 \setminus \{p_1, p_2\}$ is homotopy-equivalent to a figure 8. One gains intuition for this by making a drawing, and working out how to slide each of the points of $\mathbb{R}^2 \setminus \{p_1, p_2\}$ along some path to the figure 8, shown in the drawing below as the space X made up of two tangent circles. Note that a deformation-retraction can be realised in stages:

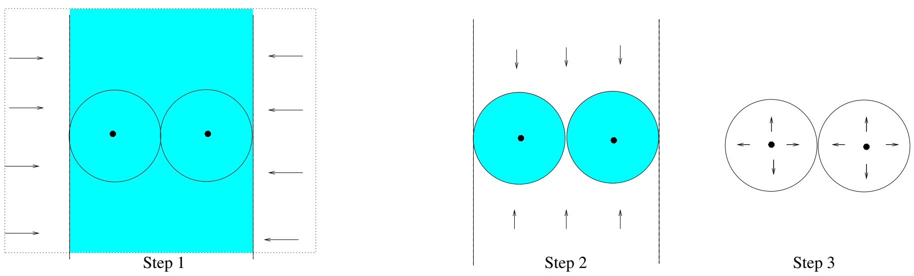
Lemma 4.7.1. *Suppose* $X \subset Y \subset Z$ *and that* $F : Z \times [0,1] \to Z$ *is a deformation-retraction of* Z *to* Y*, and that* $G : Y \times [0,1] \to Y$ *is a deformation-retraction of* Y *to* X*. Then the map* $H : Z \times [0,1] \to Z$ *defined by*

$$H(z,t) = \begin{cases} F(z,2t) & \text{if } t \in [0,1/2] \\ G(F(z,1), 2t-1) & \text{if } t \in [1/2,1] \end{cases}$$

is a deformation-retraction of Z *to* X*.*

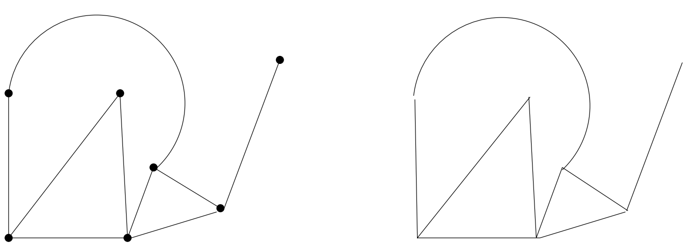
Proof. Exercise. □

To construct a deformation-retraction of $\mathbb{R}^2 \setminus \{p_1, p_2\}$ to the figure 8 shown, I suggest first constructing a horizontal deformation-retraction to the vertical strip shown,



then a vertical deformation retraction to the union of the two discs, and finally a radial deformation retraction of the discs (minus their centres) to their boundary.

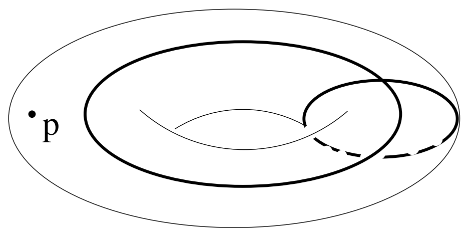
The term *graph* is used for any space homeomorphic to the union of a finite number of line segments with some of their endpoints identified. The end points of the line segments are called *nodes*, and it is conventional to depict them as heavy blobs, as in the left-hand diagram below. As a topological space, the same graph is more accurately represented in the right-hand picture. The circle and the figure 8 are both graphs.



6. **Exercise**(i) What simple graph is $\mathbb{R}^2 \setminus \{p_1, \ldots, p_n\}$ homotopy-equivalent to?
(ii) $\mathbb{R}^3 \setminus \{p\}$ is homotopy-equivalent to a 2-sphere. What can be said about $\mathbb{R}^3 \setminus \{p_1, p_2\}$?

7. Let T be the torus (see Chapter 3 and $p \in T$. Show that $T \setminus \{p\}$ has as

deformation retract a union of two circles with one point in common, as shown in the picture.



It is probably easier to construct a deformation retraction by using the model of the torus as the quotient of a square, (Chapter 3 Example 5), and taking, as the two circles, the image in the torus of the boundary of the square.

8. What simple graph does $\mathbb{R}\mathbb{P}^2 \setminus \{p\}$ deformation-retract to? Again, use the description of $\mathbb{R}\mathbb{P}^2$ as a quotient of a square, as in Chapter 3 Example 5.

In this subject it is more fun, and mathematically valuable, to develop arguments visually, without writing out all of the details. One can often see quite easily how to construct a suitable deformation-retraction, when writing out the details would be long and tedious. Of course it is important that one should, in principle, be able to write out an argument.

Corollary 4.7.2. *Suppose that $f : X \to Y$ is a homotopy equivalence. Then $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is an isomorphism.*

Proof. Write $x_1 = g(f(x_0))$. Let F be a homotopy from id_X to $g \circ f$, and let a be the path in X from x_0 to x_1, defined by $a(u) = F(x_0, u)$. By Proposition 4.7.2,

$$(\mathrm{id}_X)* = c_a \circ (g \circ f)_* = c_a \circ g_* \circ f_*. \tag{4.7.1}$$

Now $(\mathrm{id}_X)_*$ is the identity map on $\pi_1(X, x_0)$, and c_a is an isomorphism. It follows from (4.7.1) that f_* is injective and g_* surjective. A similar argument, interchanging the roles of f and g and using the fact that $f \circ g$ is homotopic to id_Y, shows that g_* is injective and f_* is surjective. Thus both f_* and g_* are bijective homomorphisms, i.e. isomorphisms. □

Proposition 4.7.3. *Homotopy equivalence is an equivalence relation on the set of all topological spaces.*

Proof. Exercise □

4.8 The fundamental group of a product

Let X and Y be spaces and let $\mathrm{pr}_X : X \times Y \to X$ and $\mathrm{pr}_Y : X \times Y \to Y$ denote the standard projections.

Theorem 4.8.1. *If X and Y are path-connected spaces, and $x_0 \in X$ and $y_0 \in Y$, the homomorphism $pr_{X*} \times pr_{Y*} : \pi_1(X \times Y, (x_0, y_0)) \to \pi_1(X, x_0) \times \pi_1(Y, y_0)$ is an isomorphism.*

Proof. If p and q are loops in X and Y based at x_0 and y_0 then the map $t \mapsto (p(t), q(t))$ is a loop in $X \times Y$ based at (x_0, y_0), and satisfies $p = \mathrm{pr}_X \circ r$ and $q = \mathrm{pr}_Y \circ r$. Thus $\mathrm{pr}_{X*} \times \mathrm{pr}_{Y*}$ is surjective.

To see that it is 1-to-1, suppose r is a loop in $X \times Y$ based at (x_0, y_0), and $\mathrm{pr}_X \circ r$ and $\mathrm{pr}_Y \circ r$ are epph to constant loops, via epphs's F and G. Then the map $H : [0,1] \times [0,1] \to X \times Y$ defined by

$$H(t, u) = (F(t, u), G(t, u))$$

is an epph from r to the constant loop based at (x_0, y_0). Thus $\ker(\mathrm{pr}_{X*} \times \mathrm{pr}_{Y*})$ is trivial, so $\mathrm{pr}_{X*} \times \mathrm{pr}_{Y*}$ is 1-to-1. $\square$

4.9 The fundamental group of a union

Since the torus is the product of two circles, by Theorem 4.8.1 its fundamental group is isomorphic to $\mathbb{Z} \times \mathbb{Z}$. On the other hand, if we take away a single point, getting a space homotopy-equivalent to the union of two circles with a single point in common (see Example 4.7.0 no. 7), we get a much more complicated fundamental group.

Let X_1 and X_2 denote the two circles, and let x_0 be the point in common. Pick a generator $[p_i]$ of $\pi_1(X_i, x_0)$ $i = 1, 2$, and use the same name for the corresponding loop in $X_1 \cup X_2$. I claim that

$$\text{in } \pi_1(X_1 \cup X_2, x_0), \quad [p_1] * [p_2] \neq [p_2] * [p_1]. \tag{4.9.1}$$

We will prove this later. It is a consequence of van Kampen's Theorem, (due equally to Seifert), which shows how, under reasonably favourable circumstances, one can determine the fundamental group of a union of spaces, $X = \bigcup_{\alpha \in A} X_\alpha$, from the fundamental groups of the X_α.

From the theorem we will derive the following description of $\pi_1(X, x_0)$, where X is the union of two circles with a single point in common. The elements of $\pi_1(X, x_0)$ are all products

$$[p_{i_1}]^{j_1} * [p_{i_2}]^{j_2} * \cdots * [p_{i_n}]^{j_n} \tag{4.9.2}$$

each i_j is either 1 or 2, each index j_k is an integer, and $n \in \mathbb{N}$. The expression (4.9.2) is *reduced* if none of the j_k is equal to 0 and $i_{k+1} \neq i_k$ for each k. The latter requirement is simply to avoid expressions like

$$[p_1]^2 * [p_1]^{-3} * [p_2] * [p_2]^4$$

which can be condensed to

$$[p_1]^{-1} * [p_2]^5.$$

The theorem says that in $\pi_1(X, x_0)$, the two reduced expressions

$$[p_{i_1}]^{j_1} * [p_{i_2}]^{j_2} * \cdots * [p_{i_m}]^{j_n} \text{ and } [p_{i'_1}]^{j'_1} * [p_{i'_2}]^{j'_2} * \cdots * [p_{i'_n}]^{j'_n}$$

are equal if and only if $m = n$ and $i_k = i'_k$ and $j_k = j'_k$ for all k. In particular, $[p_1] * [p_2] \neq [p_2] * [p_1]$.

Elements of the form (4.9.2) are multiplied simply by juxtaposing them and then condensing the resulting expression so that it becomes reduced. Thus for example

$$\left([p_1] * [p_2]^7 * [p_1]^2\right)\left([p_1]^{-2} * [p_2]^{-3} * [p_1]\right) = [p_1] * [p_2]^4 * [p_1].$$

To state van Kampen's Theorem it we need first the notion of the *free product* of groups. Let $G_\alpha, \alpha \in A$ be a collection of groups, indexed by the set A. The free product

$$*_{\alpha \in A} G_\alpha$$

is the set of finite words [2] $g_1 g_2 \cdots g_m$, such that

1. each g_i belongs to one of the G_α,
2. no g_i is the neutral element of its group, and
3. g_i and g_{i+1} never belong to the same G_α.

A word with these three properties is said to be *reduced*. The binary operation in $*_{\alpha \in A} G_\alpha$ is defined by juxtaposing words and then condensing the resulting word until it becomes reduced, by means of two procedures:

1. if the last element of the first word lies in the same G_α as the first element in the second word (so that the juxtaposed word violates condition 3), simply multiply these two elements in G_α.

[2] "Word" is the standard term here, but you could also say or "ordered lists", or "ordered m-tuples".

2. if this product gives the neutral element in G_α, omit it. If omitting it then results in another violation of rule 4, repeat procedure 1.

For example, let G and H be groups and denote their elements by $g_1, g_2, \dots$ and $h_1, h_2, \dots$, respectively. Consider the words $w_1 = g_1h_1g_2$ and $w_2 = g_3h_2$ in $G * H$. The juxtaposition of w_1 and w_2 (in that order) is $g_1h_1g_2g_3h_2$; this is not reduced, as the third and fourth elements both lie in G; applying procedure 1, it becomes $g_1h_1(g_2g_3)h_2$, which is reduced, unless $g_3 = g_2^{-1}$; in that case g_2g_3 is the neutral element of G, and using procedure 2 we further condense it to $g_1h_1h_2$, and then by procedure 1 to $g_1(h_1h_2)$. This is a reduced word provided $h_1 \neq h_2^{-1}$.

The neutral element in $*_{\alpha\in A}G_\alpha$ is the empty word, with no letters.

The inverse of $g_1g_2 \cdots g_m$ is $g_m^{-1}g_{m-1}^{-1} \cdots g_2^{-1}g_1^{-1}$.

Lemma 4.9.1. *$*_{a\in A}G_\alpha$, with this operation, is a group.*

For each $\alpha_i \in A$, there is an injective homomorphism $G_\alpha \to\ *_{\alpha\in A}\, G_\alpha$, sending $g \in G_\alpha$ to the one-letter word g in $*_{\alpha\in A}G_\alpha$. Thus we can regard each G_{α_i} as a subgroup of $*_{\alpha\in A}G_\alpha$. We do so in what follows.

The free product has an important universal property:

Lemma 4.9.2. *Let $G_\alpha, \alpha \in A$ be a collection of groups, let H be a group, and for each $\alpha \in A$ let $\varphi_\alpha : G_\alpha \to H$ be a group homomorphism. Then there is a unique homomorphism $\varphi : G \to H$ extending the φ_α.*

Proof. Given the word $g_1 \cdots g_m$, suppose $g_i \in G_{\alpha_i}$ for $i = 1 \dots m$. Define $\varphi(g_1 \cdots g_m) = \varphi_{\alpha_1}(g_1)\varphi_{\alpha_2}(g_2) \cdots \varphi_{\alpha_m}(g_m)$. $\square$

In fact the free product of the G_α is characterised by this property: any other group with the same property is isomorphic to it (see Exercises III).

Suppose that the space X is a union of subsets X_α, $\alpha \in A$, with each X_α containing the point x_0. Then for each α, there is a homomorphism $i_{\alpha *} : \pi_1(X_\alpha, x_0) \to \pi_1(X, x_0)$ induced by the inclusion $i_\alpha : X_\alpha \hookrightarrow X$. From the universal property we therefore obtain a homomorphism

$$\varphi : *_{\alpha\in A}\pi_1(X_\alpha, x_0) \to \pi_1(X, x_0) \tag{4.9.3}$$

extending all of the $i_{\alpha *}$.

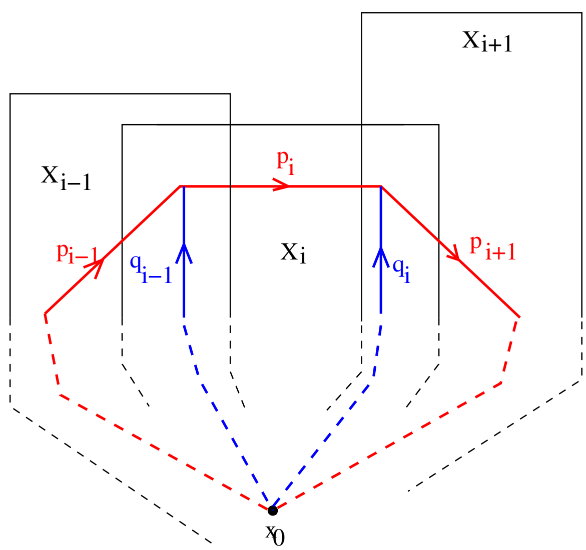
Van Kampen's theorem has two parts, which we state and prove separately.

Theorem 4.9.3. *Let X be the union of path connected open sets A_α, each one containing the point x_0, and with $X_\alpha \cap X_\beta$ path connected for all $\alpha\,\beta \in A$. Then the homomorphism φ of (4.9.3) is surjective.*

Proof. Let $p \in L(X, x_0)$. Each point $t \in [0, 1]$ is contained in an interval (t', t'') such that $p((t', t''))$ is entirely contained in some X_α. A finite number of these intervals cover $[0, 1]$. It follows that there is a sequence $t_0 = 0 < t_1 < \cdots < t_n = 1$ such that for $i = 1, \ldots, n$, $p([t_{i-1}, t_i])$ is entirely contained in one of the X_a. Re-label these X_α so that $p([t_{i-1}, t_i]) \subset X_i$. Let q_i be a path in $X_i \cap X_{i+1}$ from x_0 to $p(t_i)$, and let $p_i : [0, 1] \to X$ be the path $p_i(t) = p\big((1 - t)t_{i-1} + tt_i\big)$; p_i simply traces out the part of the path p from $p(t_{i-1})$ to $p(t_i)$. Then for each i, the path

$$(q_{i-1} * p_i) * q_i^{-1}$$

is a loop in X_i.



Finally, $[p]$ is the image under φ of the word

$$[p_1 * q_1^{-1}][q_1 * p_2 * q_2^{-1}]\cdots[q_{n-2} * p_{n-1} * q_{n-1}^{-1}][q_{n-1} * p_n]$$

in $*_{\alpha\in A} G_\alpha$. $\square$

To describe the kernel of φ, notice that any $p \in L(X\alpha \cap X_\beta, x_0)$ gives rise to elements $[p^{(\alpha)}]$ and $[p^{(\beta)}]$ of $\pi_1(X_\alpha, x_0)$ and $\pi_1(X_\beta, x_0)$. These two count as *different* elements in the free product $*_\alpha \pi_1(X_\alpha, x_0)$, even though both are sent by φ to the *same* element, the class of p in $\pi_1(X, x_0)$. Thus φ is not injective.

Note that $[p^{(\alpha)}] \in \pi_1(X_\alpha, x_0)$ is an equivalence class of loops in X_α, and $[p^{(\beta)}] \in \pi_1(X_\beta, x_0)$ is an equivalence class of loops in X_β. They are distinct equivalence classes, even though both are represented by the same loop p.

Because $\varphi([p^{(\alpha)}] = \varphi([p^{(\beta)}])$, the two-letter words $[p^{(\beta)}]^{-1}[p^{(\alpha)}]$ and $[p^{(\beta)}][p^{(\alpha)}]^{-1}$, and their inverses, lie in $\ker\varphi$. The second part of van Kampen's Theorem says that provided each triple intersection $X_\alpha \cap X_\beta \cap X_\gamma$ is path connected, these elements generate the kernel of φ. More precisely:

Theorem 4.9.4. *Let X and the X_α be as in 4.9.3, and suppose in addition that each triple intersection $X_\alpha \cap X_\beta \cap X_\gamma$ is path-connected. Then the kernel of φ is the normal subgroup N of $*_\alpha \pi_1(X_\alpha, x_0)$ generated by all the elements $[p_\beta]^{-1}[p_\alpha]$ just described (i.e. is the smallest normal subgroup containing them all).*

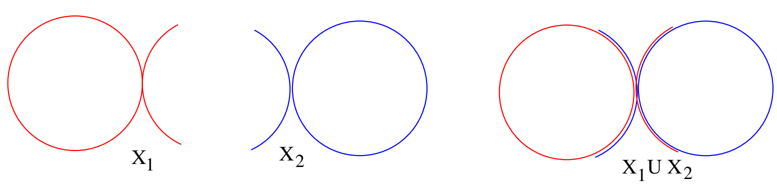
Before proving this, note the following consequence:

Corollary 4.9.5. *If, in 4.9.3, there are just two X_α, say X_1 and X_2, and their intersection is simply connected, then $\varphi : \pi_1(X_1, x_0) * \pi_1(X_2, x_0) \to \pi_1(X, x_0)$ is an isomorphism.*

The union of two circles with a single point in common is called the *wedge sum* of the two circles, and denoted $S^1 \vee S^1$.

Corollary 4.9.6. $\pi_1(S^1 \vee S^1) \simeq \mathbb{Z} * \mathbb{Z}$.

Proof. (of 4.9.6): As base point we take the point of intersection, x_0. We cannot apply van Kampen immediately, as neither S^1 is open in $S^1 \vee S^1$. This is easily remedied: $S^1 \vee S^1$ is the union of the two open subspaces X_1 and X_2 shown in the following diagram, each of which has its respective circle as a deformation retract.



□

Proof. (of 4.9.4) (closely based on the proof in *Algebraic Topology* by Allen Hatcher): Given $[p] \in \pi_1(X, x_0)$, call the preimages of $[p]$ in $*_{\alpha\in A}\pi_1(X_\alpha, x_0)$

factorisations of $[p]$. Given any factorisation $[p_1][p_2]\cdots[p_m]$ of $[p]$, we can transform it by moves of two types:

M1: Combine two adjacent terms $[p_i]$ and $[p_{i+1}]$ into one, $[p_i * p_{i+1}]$, if $[p_i]$ and $[p_{i+1}]$ can be represented by loops in the same $L(X_\alpha, x_0)$.

M2: If $[p]$ is the class of a loop in $X_\alpha \cap X_\beta$, transform a word of the form $w[p^{(\alpha)}]w'$ into the word $w[p^{(\beta)}]w'$.

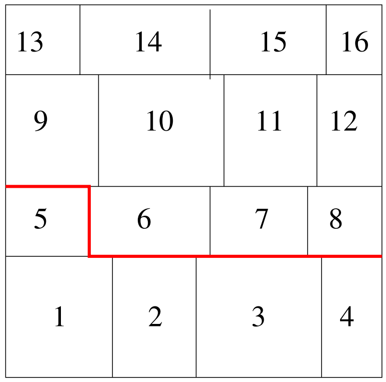
We also allow the reverse of these moves. A move of type M2 or its inverse may result in a non-reduced word, and we may then need one or more moves of type M1 to condense it.

Clearly neither move changes the image of the word in $\pi_1(X, x_0)$. A move of type M2 does not change the class of the word in $*_\alpha \pi_1(X_\alpha, x_0)/N$, for the starting word $w[p^{(\alpha)}]w'$ is transformed into the word $w[p^{(\beta)}]w'$ by left multiplying it by the element $w[p^{(\beta)}][p^{(\alpha)}]^{-1}w^{-1}$ of N. [3]

If we can show that any two factorisations are related by a sequence of these operations, then we will have proved the theorem.

So let $[p_1][p_2]\cdots[p_k]$ and $[p'_1]\cdots[p'_\ell]$ be two factorisations. Then the two loops $p_1 * p_2 * \cdots * p_k$ and $p'_1 * \cdots * p'_\ell$ are epph in X. Let $F : [0,1] \times [0,1] \to X$ be an epph from $p_1 * p_2 * \cdots * p_k$ to $p'_1 * \cdots * p'_\ell$. There exist partitions $0 = t_0 < t_1 < \cdots < t_m = 1$ and $0 = u_0 < u_1 < \cdots < u_n = 1$ such that each rectangle $R_{ij} := [t_{i-1}, t_i] \times [s_{j-1}, s_j]$ is mapped by F into a single X_α, which we fix and re-label X_{ij}. We assume that the partition $0 = t_0 < t_1 < \cdots < t_m = 1$ includes the partitions by which the interval is divided in forming the products $p_1 * p_2 * \cdots * p_k$ and $p'_1 * p'_2 * \cdots * p'_\ell$. Since F maps each R_{ij} into the open set X_{ij}, we can shift the vertical edges of the R_{ij} slightly so that no point of $[0,1] \times [0,1]$ lies in more than three of the R_{ij}, with $F(R_{ij})$ still contained in X_{ij}. We may assume that $n > 2$, so that we can carry out this perturbation only on the vertical edges of the rectangles R_{ij} with $2 \le j \le n-1$.

[3] Note that we need the *normal* subgroup generated by the $[p^{(\beta)}][p^{(\alpha)}]^{-1}$, for this to work: in general the element we need to left-multiply by, $w[p^{(\beta)}][p^{(\alpha)}]^{-1}w^{-1}$, will not lie in the subgroup generated by the $[p^{(\beta)}][p^{(\alpha)}]^{-1}$; it will of course lie in the normal subgroup.



Now relabel the R_{ij} as shown in the figure, and re-label the X_{ij} containing them in the same way. If γ is a path in $[0,1] \times [0,1]$ from the left edge to the right edge, then the restriction F_γ lies in $L(X, x_0)$. Let γ_r be the path separating the first r rectangles $R_1, \ldots, R_r$ from the remainder. The figure shows γ_5 in red. Thus, γ_0 is the lower edge $[0,1] \times \{0\}$ and γ_{mn} is the upper edge $[0,1] \times \{1\}$.

For each corner v of one of the R_k with $F(v) \neq x_0$, let q_v be a path in X from x_0 to $F(v)$. We can choose q_v to lie in the intersection of the two or three open sets X_i containing v, since we have assumed all double and triple intersections are path-connected. If we insert into $F|\gamma_v$ the appropriate paths $q_v^{-1} * q_v$ at successive corners, as in the proof of surjectivity of φ, then we obtain a factorisation of $[F|\gamma_v]$ by regarding the loop corresponding to a horizontal or vertical edge joining adjacent corners as lying in the X_i corresponding to one of the rectangles R_i containing the segment. For each segment, a different choices of the R_i containing it changes the factorisation of $[F|\gamma_r]$ to an equivalent factorisation (using M2). The factorisations associated to successive paths γ_r and γ_{r+1} are also equivalent, since pushing γ_r across R_{r+1} to γ_{r+1} changes $F|\gamma_r$ to $F|\gamma_{r+1}$ by a homotopy within X_{r+1} (whose active ingredient is $F|R_{r+1}$).

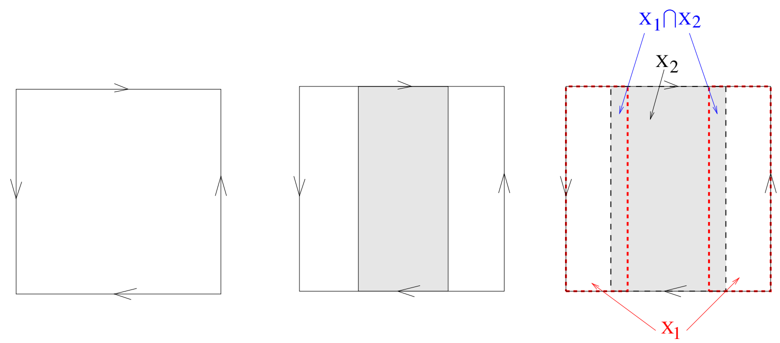
Since all of the factorisations we obtain in this way are equivalent to one another, it follows that the original factorisations $[p_1][p_2]\cdots[p_k]$ and $[p_1'][p_2']\cdots[p_\ell']$ are equivalent to one another, proving the theorem.

$\square$

4.10 Applications of van Kampen's Theorem

1. $\mathbb{RP}^2$ may be obtained by gluing a disc and a Möbius band. To see this, recall that $\mathbb{RP}^2$ can be obtained as the quotient of a square by

identifying opposite edges as shown in the left hand diagram below.



The shaded strip in the middle figure becomes a Möbius band under the identification of upper and lower edges. We denote this by X_2. The remainder of the square becomes a (topological) disc after the right and left edges, and upper and lower edges, are glued. If it is difficult to see this from the diagram, you should cut out a piece of paper and make the required gluings with sellotape. Call this disc X_1. To make X_1 and X_2 open in $\mathbb{R}\,\mathbb{P}^2$, we fatten each one, as shown in the right hand diagram. The intersection $X_1 \cap X_2$ is the union of the two narrow vertical strips , with the top of the left hand strip glued to the bottom of the right hand strip, and vice versa. Gluing top left to bottom right forms a strip. Then gluing bottom left to top right forms either a Möbius strip or a cylinder. (**Exercise**: Which?) In either case, $X_1 \cap X_2$ deformation-retracts to a circle. The Möbius strip X_2 also deformation-retracts to a circle. Thus,

$$\pi_1(X_1, x_0) = \{1\}, \quad \pi_1(X_2, x_0) \simeq \mathbb{Z}, \quad \pi_1(X_1 \cap X_2, x_0) \simeq \mathbb{Z}\,.$$

Picking a basepoint in $X_1 \cap X_2$, and not bothering to mention it in our formulae, by 4.9.4 we have

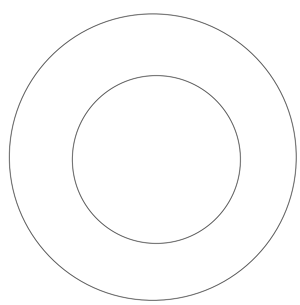
$$\pi_1(\mathbb{R}\,\mathbb{P}^2) \simeq \frac{\{1\} * \pi_1(X_2)}{N} \simeq \frac{\{1\} * \mathbb{Z}}{N}.$$

By definition of free product, for any group G, $\{1\} * G$ can be naturally identified with G. What happens to N under this identification? Recall that N is the (normal) subgroup generated by the two letter words of the form $[p^{(1)}][p^{(2)}]^{-1}, [p^{(1)}]^{-1}[p^{(2)}]$ and their inverses, where p is a loop in $X_1 \cap X_2$, and $[p^{(1)}]$ and $[p^{(2)}]$ are the classes it gives rise to in $\pi_1(X_1)$ and $\pi_1(X_2)$ respectively. In this case $G \simeq \mathbb{Z}$ is abelian, so every subgroup is normal. This makes things easier! Since $\pi_1(X_1) = \{1\}$, the two letter word $[p^{(1)}]^{-1}[p^{(2)}]$ reduces to $[p^{(2)}]$; so N is identified to the subgroup of $\mathbb{Z} = \pi_1(X_2)$ consisting of the image of $\pi_1(X_1 \cap X_2)$. This is the subgroup $2\,\mathbb{Z}$, since $X_1 \cap X_2$ is a thickening of the boundary circle of the Möbius strip X_2. So

$$\pi_1(\mathbb{R}\,\mathbb{P}^2) \simeq \frac{\mathbb{Z}}{2\,\mathbb{Z}}.$$

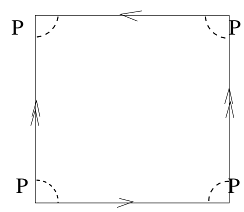
It is hard to imagine a loop p generating $\pi_1(\mathbb{R}\,\mathbb{P}^2)$: p itself cannot be contracted to a point on $\mathbb{R}\,\mathbb{P}^2$, but $p * p$ can be! Can you see from the diagram how this works? Can you see a way of contracting $p * p$ to a point? Equivalently, find a map $D^2 \to \mathbb{R}\,\mathbb{P}^2$ extending the map from S^1 to the loop p.

2. Obtaining $\mathbb{R}\,\mathbb{P}^2$ as quotient of a square is a little unnatural. More natural, because more homogeneous, is its description as quotient of S^2 by the antipodal identification. The whole of $\mathbb{R}\,\mathbb{P}^2$ is in fact obtained from one half of the sphere, say the northern hemisphere, with the only pairs of antipodal points now lying on the equator. So we have a third description of $\mathbb{R}\,\mathbb{P}^2$ as quotient of a disc by the antipodal identification on the boundary circle only. The diagram below shows a disc D containing a smaller concentric disc D'.



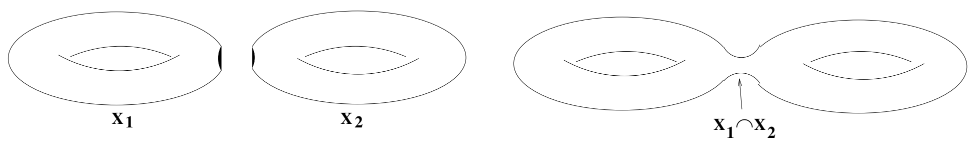
Under the quotient map $D \twoheadrightarrow D/\sim = \mathbb{R}\,\mathbb{P}^2$, D' is mapped homeomorphically onto its image, and thus gives a disc in $\mathbb{R}\,\mathbb{P}^2$. What is the image of $D \smallsetminus D'$?. Use this to give another calculation of $\pi_1(\mathbb{R}\,\mathbb{P}^2)$.

3. The surface obtained by identifying opposite edges of a square as shown here



is known as the *Klein bottle*. It is a 2-dimensional manifold, since each point clearly has a neighbourhood homeomorphic to a disc in $\mathbb{R}^2$; a suitable neighbourhood of the point P is shown in the diagram. You can use van Kampen to find its fundamental group, in the same way as for $\mathbb{R}\,\mathbb{P}^2$, though the result is a little more complicated.

4. The genus 2 surface can be obtained by gluing together two holed tori X_1 and X_2, along the edges of the holes, as shown below.



Each holed torus deformation-retracts to $S^1 \vee S^1$, so has fundamental group $\mathbb{Z} * \mathbb{Z}$. After thickening X_1 and X_2 to make them open in their union, the intersection $X_1 \cap X_2$ is a cylinder, so $\pi_1(X_1 \cap X_2) \simeq \pi_1(S^1) \simeq \mathbb{Z}$. Thus
$$\pi_1(S) \simeq \frac{\mathbb{Z} * \mathbb{Z} * \mathbb{Z} * \mathbb{Z}}{N}$$
where N is a certain normal subgroup. In fact (Exercise) $N \simeq \mathbb{Z}$, though the way that it is expressed in $\mathbb{Z} * \mathbb{Z} * \mathbb{Z} * \mathbb{Z}$ needs to be made explicit. How to do this?

5. How to calculate the fundamental group of the complement of a standard, planar circle in $\mathbb{R}^3$? It can be done by an easy application of van Kampen. You have to find path-connected open sets whose union is $\mathbb{R}^3 \setminus S^1$ and whose intersection also is path-connected. Begin by using your geometrical imagination, or making a drawing, rather than straight away trying to write down a precise description of two suitable open sets.

6. In general, if K is a circle embedded in $\mathbb{R}^3$ (i.e. a knot), then $\pi_1(\mathbb{R}^3 \setminus K)$ depends on the embedding. It is a powerful algebraic invariant of the embedding. Standard techniques are available for its computation;

they are beyond the scope of this course. In every case, its abelianisation is isomorphic to $\mathbb{Z}$.

Chapter 5

Covering Spaces

Definition 5.0.1. A continuous map $f : X \to Y$ is a *covering map* if for every point $y_0 \in Y$ there is an open neighbourhood V of y_0 in Y such that $f^{-1}(V)$ is a disjoint union of open sets U_α, such that for each U_α, the restriction $f| : U_\alpha \to V$ is a homeomorphism. In this case we also say that X is a *covering space* of Y. The maps $(f|_{U_\alpha})^{-1} : V \to U_\alpha$ are *local inverses* of f.

We adopt the (slightly non-standard) term "well-covered" to describe an open set such as the set V in the definition of covering space, whose preimage consists of a disjoint union of open sets U_α each of which is mapped homeomorphically to V by f.

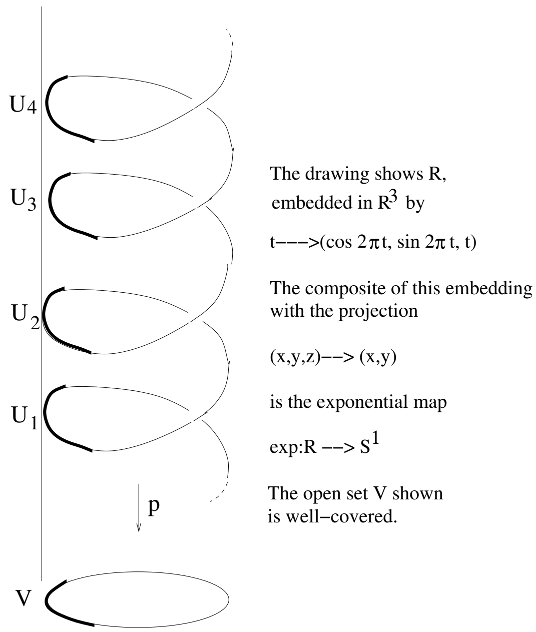
Example 5.0.2. Examples are plentiful.

1. The exponential map $\exp : \mathbb{R} \to S^1$ is a covering map. In the proof of Proposition 4.1.1, that paths in S^1 lift to paths in $\mathbb{R}$, we used the local inverse $\ell : S^1 \setminus \{-1\} \to (-1/2, 1/2)$ to exp. In fact, for every $b \in S^1$, and every $a \in \mathbb{R}$ such that $\exp(a) = b$, we can use ℓ to define a local inverse $\ell_a : S^1 \setminus \{-b\} \to (a - 1/2, a + 1/2)$ to exp, simply by

$$\ell_a(z) = \ell(b^{-1}z) + a.$$

So taking, as V, the open set $S^1 \setminus \{-b\}$, and picking (arbitrarily) one $a \in \exp^{-1}(b)$, we have

$$\exp^{-1}(V) = \bigcup_{n \in \mathbb{Z}} (a - 1/2 + n, a + 1/2 + n).$$



2. the map $\mathbb{C} \to \mathbb{C} \smallsetminus \{0\}$ sending z to e^z is a covering map. Its local inverses are called “branches of the natural logarithm function”.

3. The map $S^1 \to S^1$ sending z to z^n is a covering map. Its local inverses are called n’th roots. **Exercise** Make a drawing for $n = 3$, showing a well-covered open set in the target S^1 and its preimages in the source S^1.

4. In Example 3 Numbers 6 and 7, we described a 2-to-1 map from the cylinder to the Möbius strip. This is a covering map; this becomes especially clear when we wrap the cylinder (embedded in $\mathbb{R}^3$ as a strip with a whole twist) twice round the Möbius strip. Look out for the classroom demonstration.

5. Many quotient maps are covering maps. The map $S^2 \to \mathbb{R}\mathbb{P}^2$ sending each point x to its equivalence class is a covering map; so is $\exp \times \exp : \mathbb{R}^2 \to S^1 \times S^1$; by composing with the homeomorphism $\phi : S^1 \times S^1 \to T$ of Chapter 3 Example 1, we get a covering map $\mathbb{R}^2 \to T$.

6. A map $f : X \to Y$ is a *local homeomorphism* if for each point $x \in X$ there are open sets $U \subset X$, containing x, and $V \subset Y$, containing $f(x)$, such that $f| : U \to V$ is a homeomorphism. A local homeomorphism

is not necessarily a covering map: if $y \in Y$ has infinitely many preimages $x_i \in X$, then although for each x_i there is a neighbourhood U_i of x_i in X and a neighbourhood V_i or y in Y such that $f| : U_i \to V_i$ is a homeomorphism, a problem may arise trying to find *one* neighbourhood V of y which works for all of the x_i. But if f is finite-to-one, such a V can always be found, by intersecting the V_i corresponding to the x_i (**Exercise**). Thus if X is compact and Hausdorff, and $f : X \to Y$ is a local homeomorphism, then it is a covering map (**Exercise**).

7. (For those who study Manifolds) If X and Y are compact n-dimensional smooth manifolds and $f : X \to Y$ is a smooth map whose derivative $d_x f : T_x X \to T_{f(x)} Y$ is an isomorphism for all $x \in X$, then f is a covering map; this follows from the previous example by the inverse function theorem.

8. *Integer maps on the torus* Recall that the quotient of $\mathbb{R}^2$ by the equivalence relation

$$(x_1, y_1) \sim (x_2, y_2) \quad \text{if} \quad (x_1, y_1) - (x_2, y_2) \in \mathbb{Z} \times \mathbb{Z}$$

is $S^1 \times S^1$.

Let A be a 2×2 matrix with integer entries. It defines a linear map $\mathbb{R}^2 \to \mathbb{R}^2$, with the property that if $(x_1, y_1) \sim (x_2, y_2)$ then $A(x_1, y_1) \sim A(x_2, y_2)$. Hence, A passes to the quotient to define a continuous map $\bar{A} : S^1 \times S^1 \to S^1 \times S^1$. In view of the homeomorphism $S^1 \times S^1 \simeq T$, A gives rise to a continuous map $T \to T$, which we will also call $\bar{A}$.

Proposition 5.0.3. *If* $\det A \neq 0$ *then* $\bar{A}$ *is a covering map.*

Proof. See Exercises 4. □

9. *Restriction of a covering map to a subspace:* Suppose that $p : \tilde{Y} \to Y$ is a covering map and $Y_0 \subset Y$. Let $\tilde{Y}_0 := p^{-1}(Y_0)$, and let $p_0 : \tilde{Y}_0 \to Y_0$ be the restriction of $p : \tilde{Y} \to Y$. Then $p_0 : \tilde{Y}_0 \to Y_0$ is a covering map; indeed, if $V \subset Y$ is well-covered by $p : \tilde{Y} \to Y$ then $V \cap Y_0$ is well-covered by $p_0 : \tilde{Y}_0 \to Y_0$ (**Exercise**).

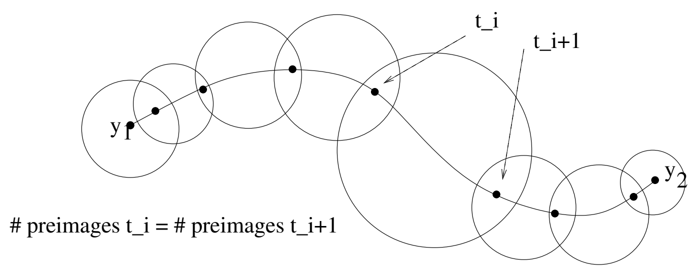
10. *Pull-back of a covering map:* Suppose that $p : \tilde{Y} \to Y$ is a covering map and $f : X \to Y$ is a continuous map. Define a new space $X \times_Y \tilde{Y}$ by

$$X \times_Y \tilde{Y} = \{(x, \tilde{y}) \in X \times \tilde{Y} : f(x) = p(\tilde{y})\}.$$

Let $\mathrm{pr}_X : X \times_Y \tilde{Y} \to X$ be the restriction of the usual projection $X \times \tilde{Y} \to X$. Then $\mathrm{pr} : X \times_Y \tilde{Y} \to X$ is a covering map; indeed, any set $W \subset X$ mapped by f into a well covered open set $V \subset Y$ is well covered by $\mathrm{pr} : X \times_Y \tilde{Y} \to X$ (**Exercise**). The new covering map $\mathrm{pr} : X \times_Y \tilde{Y} \to X$ is called the *pull-back of* $p : \tilde{Y} \to Y$ *by* f, and sometimes we write $f^*(\tilde{Y})$ and $f^*(p)$ in place of $X \times_Y \tilde{Y}$ and pr.

5.0.1 The degree of a covering map

Let $p : \tilde{Y} \to Y$ be a covering map. Let $V \subset Y$ be a well-covered set, and let $y \in V$. For each of the open sets U_α making up $p^{-1}(V)$, as in the definition of covering map, $U_\alpha \cap p^{-1}(y)$ consists of a single point. It follows that $p^{-1}(y)$ has as many points as there are open sets U_α, and thus for any two points $y_1, y_2 \in V$, there is a bijection $p^{-1}(y_1) \to p^{-1}(y_2)$. If Y is path connected, any two points y_1 and y_2 can be joined by a path, which, being compact, can be covered by a finite number of well-covered connected open sets. From this it follows (**Exercise**)



that there is a bijection $p^{-1}(y_1) \to p^{-1}(y_2)$. The cardinality of $p^{-1}(y)$ is called the *degree* of the covering map.

5.0.2 Path lifting and Homotopy Lifting

Lemma 5.0.4. *Let $p : \tilde{Y} \to Y$ be a covering map, and let $f : X \to Y$ be a continuous map, with X connected. Suppose that $\tilde{f}_1 : X \to \tilde{Y}$ and $\tilde{f}_2 : X \to \tilde{Y}$ are lifts of f. If $\tilde{f}_1(x) = \tilde{f}_2(x)$ for some point $x \in X$, then $\tilde{f}_1$ and $\tilde{f}_2$ agree everywhere.*

Proof. Let A be the set of points in X where $\tilde{f}_1$ and $\tilde{f}_2$ agree. By hypothesis, A is not empty. We will show it is both open and closed. This forces $A = X$, by the connectedness of X.

Suppose $x \in X$. Let V be a well-covered neighbourhood of $f(x)$ in Y, with $p^{-1}(V) = \coprod_{\alpha \in A} U_\alpha$. Suppose $\tilde{f}_1(x) \in U_{\alpha_1}$, and $\tilde{f}_2(x) \in U_{\alpha_2}$. By continuity of $\tilde{f}_1$ and $\tilde{f}_2$, there are open neighbourhoods N_1 and N_2 of x in X such that $\tilde{f}_1(N_1) \subset U_{\alpha_1}$ and $\tilde{f}_2(N_2) \subset U_{\alpha_2}$. The fact that $\tilde{f}_1$ and $\tilde{f}_2$ are lifts of f then implies that on N_1 and N_2 respectively, $\tilde{f}_1 = (p|_{U_{\alpha_1}})^{-1} \circ f$ and $\tilde{f}_2 = (p|_{U_{\alpha_2}})^{-1} \circ f$. Thus if $\tilde{f}_1(x) = \tilde{f}_2(x)$ then $U_{\alpha_1} = U_{\alpha_2}$, and $\tilde{f}_1$ and $\tilde{f}_2$ agree on the open set $N_1 \cap N_2$. Thus $x \in N_1 \cap N_2 \subset A$, so A is open. On the other hand, if $\tilde{f}_1(x) \neq \tilde{f}_2(x)$ then $U_{\alpha_1} \neq U_{\alpha_2}$ and $\tilde{f}_1$ and $\tilde{f}_2$ disagree on all of $N_1 \cap N_2$. This shows that $X \setminus A$ is open. □

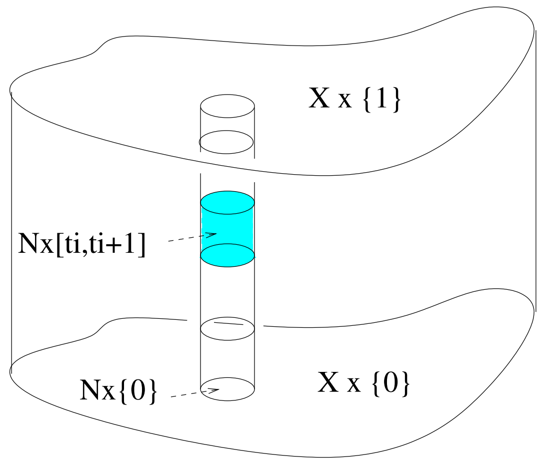
Theorem 5.0.5. (Unique homotopy extension/lifting) *Suppose that $p : \tilde{Y} \to Y$ is a covering map, that $f : X \to Y$ is a continuous map with a lift $\tilde{f}$, and that $F : X \times [0,1] \to Y$ is a homotopy with $F(x,0) = f(x)$. Then then there is a unique homotopy $\tilde{F} : X \times [0,1] \to \tilde{Y}$ lifting F and extending $\tilde{f}$ (i.e. such that $\tilde{F}(x,0) = \tilde{f}(x)$ for all $x \in X$).*

The statement can be summarised by the diagram

$$\begin{tikzcd}
& & \tilde{Y} \arrow[d, "p"] \\
X \times \{0\} \arrow[r] \arrow[urr, "\tilde{f}"] & X \times [0,1] \arrow[r, "F"'] \arrow[ur, dashed, "\exists! \tilde{F}"'] & Y
\end{tikzcd} \tag{5.0.1}$$

Proof. Choose a collection $\{V_\beta\}$ of well-covered open sets whose union covers Y. Let x_0 be a point of X. First we will construct a lift $\tilde{F}_N$ of F on some open set of the form $N \times [0,1]$, where N is a neighbourhood of x_0 in X.

Since the V_β cover Y, their preimages under F cover $X \times [0,1]$. For each $t \in [0,1]$, we can choose a neighbourhood N_t of x_0 and $a_t, b_t \in [0,1]$ so that $F(N_t \times (a_t, b_t))$ is contained in some V_β (if $t = 0$ or 1, replace (a_t, b_t) by $[0, b_t)$ or $(a_t, 1]$ respectively). A finite number of the intervals $[0, b_t), (a_t, b_t)$ and $(a_t, 1]$ cover $[0,1]$; by intersecting the corresponding neighbourhoods N_t of x_0, we obtain an open neighbourhood N of x_0, and $0 = t_0 < t_1 < \cdots < t_n = 1$ such that for each $i = 0, \ldots, n$, $[t_i, t_{i+1}]$ is contained in one of these intervals (a_t, b_t) or $[0, b_t)$ or $(a_t, 1]$, so that $F(N \times [t_i, t_{i+1}])$ is contained in some V_β. Denote this V_β by V_i.



We begin by defining $\tilde{F}_N$ on $N \times [0, t_1]$. This is easy. As $F(N \times [0, t_1]) \subset V_1$, and as V_1 is well-covered by p, we know that $p^{-1}(V_1)$ is a union of open sets, each of which is mapped homeomorphically to V_1. Exactly one of these open sets must contain $\tilde{f}(x_0)$. Call it U_1. By shrinking N (replacing it by $N \cap \tilde{f}^{-1}(U_1)$, which is still a neighbourhood of x_0), we guarantee that $\tilde{f}(N) \subset U_1$. Let $(p|_{U_1})^{-1}$ be the inverse of $p| : U_1 \to V_1$. Define $\tilde{F}_N$ on $N \times [0, t_1]$ by $\tilde{F}_N = (p|_{U_1})^{-1} \circ F$.

Now suppose inductively that $\tilde{F}_N$ has been constructed on $N \times [0, t_i]$. Essentially the same argument we have just used, extends $\tilde{F}_N$ to $N \times [0, t_{i+1}]$: as $F(N \times [t_i, t_{i+1}])$ is contained in V_i, and V_i is well-covered, there is a unique open set $U_i \subset \tilde{Y}$ mapped homeomorphically to V_i by p, and such that $\tilde{F}_N(x_0, t_i) \in U_i$. As before, we shrink N so that $\tilde{F}(N \times \{t_i\}) \subset U_i$ (simply replace N by the projection to X of $(N \times \{t_i\}) \cap \tilde{F}^{-1}(U_i)$). Define $\tilde{F}_N$ on $N \times [t_i, t_{i+1}]$ by $\tilde{F}_N = (p|_{U_i})^{-1} \circ F$. By construction, this definition agrees with the $\tilde{F}_N$ already defined on $N \times [0, t_i]$, on the intersection $N \times \{t_i\}$ of their domains.

Finitely many steps of this kind define $\tilde{F}_N$ on all of $N \times [0, 1]$. Each step has involved shrinking N to a possibly smaller neighbourhood of x_0, but as this happens only a finite number of times, the N we end up with is still a neighbourhood of x_0.

Now we show that the $\tilde{F}_N$ piece together to define a lift $\tilde{F}$ on all of $X \times [0, 1]$. For this, we again use Lemma 5.0.4. Suppose that M and N are open sets in X, and $\tilde{F}_N : N \times [0, 1]$ and $\tilde{F}_M : M \times [0, 1]$ are lifts of $F|_{M \times [0,1]}$, and $F|_{N \times [0,1]}$, satisfying $\tilde{F}_N(x, 0) = \tilde{f}(x)$ for $x \in N$, and similarly $\tilde{F}_M(x, 0) = \tilde{f}(x)$ for $x \in M$. If $x \in M \cap N$, then $\tilde{F}_M$ and $\tilde{F}_N$ agree at $(x, 0)$, since both take the value $\tilde{f}(x)$ there. Hence, by Lemma 5.0.4 applied to their restrictions to $\{x\} \times [0,1]$, they coincide on all of $\{x\} \times [0,1]$ and therefore on all of $(M \cap N) \times [0,1]$.

Thus, all of our local lifts $\tilde{F}_N$ coincide on the intersections of their domains. This means that we obtain a well-defined and continuous lift $\tilde{F}$ simply by setting

$$\tilde{F}(x,t) = \tilde{F}_N(x,t)$$

for any $\tilde{F}_N$ such that x lies in the open set N.

Uniqueness of the lift $\tilde{F}$ follows from Lemma 5.0.4. □

Corollary 5.0.6. *Let $p : \tilde{Y} \to Y$ be a covering map. If $a : [0,1] \to Y$ is a path and $p(\tilde{y}_0) = a(0)$, then there exists a unique lift $\tilde{a}$ of a with $\tilde{a}(0) = \tilde{y}_0$.*

Proof. Apply the Theorem in the special case where $X = \{0\}$.

$$\begin{tikzcd}
& & \tilde{Y} \arrow[d, "p"] \\
\{0\} \arrow[r] \arrow[urr] & {[0,1]} \arrow[r, "a"'] \arrow[ur, dashed, "\tilde{a}"'] & Y
\end{tikzcd} \tag{5.0.2}$$

□

5.0.3 Applications of Unique Homotopy Lifting

For brevity, if a is a path in Y with $a(0) = y_0$, we will refer to it as a *path in* (Y, y_0), and we will refer to its lift to $\tilde{Y}$, starting at $\tilde{y}_0$, as its *lift to* $(\tilde{Y}, \tilde{y}_0)$.

Corollary 5.0.7. *Let $p : \tilde{Y} \to Y$ be a covering map, and let y_0, y_1 be points in Y. Each path in Y from y_0 to y_1 determines a bijection from $p^{-1}(y_0)$ to $p^{-1}(y_1)$.*

Proof. Let a be such a path. For each point $\tilde{y}_0 \in p^{-1}(y_0)$ there is a unique lift $\tilde{a}_{\tilde{y}_0}$ of a to $(\tilde{Y}, \tilde{y}_0)$. We must have $\tilde{a}_{\tilde{y}_0}(1) \in p^{-1}(y_1)$. Thus

$$\tilde{y}_0 \mapsto \tilde{a}_{\tilde{y}_0}(1)$$

determines a map

$$\tilde{c}_a : p^{-1}(y_0) \to p^{-1}(y_1).$$

It has inverse $\tilde{c}_{a^{-1}} : p^{-1}(y_1) \to p^{-1}(y_0)$. This is the inverse because if $\tilde{a}_{\tilde{y}_0}$ is a lift of a to $(Y, \tilde{y}_0)$ then $(\tilde{a}_{\tilde{y}_0})^{-1}$ is a lift of a^{-1} to $(\tilde{Y}, c_a(\tilde{y}_0))$, and therefore

$$(\tilde{a}_{\tilde{y}_0})^{-1} = \widetilde{a^{-1}}_{c_a(\tilde{y}_0)}$$

(we are using the uniqueness in the statement of 5.0.6), from which

$$\tilde{y}_0 = (\tilde{a}_{\tilde{y}_0})^{-1}(0) = \widetilde{a^{-1}}_{c_a(\tilde{y}_0)}(0) = c_{a^{-1}}(c_a(\tilde{y}_0)).$$

□

Note that if a is a loop in (Y, y_0), its lift to $(\tilde{Y}, \tilde{y}_0)$ may not be a loop. This is exactly the point of our definition of the degree of a loop in S^1: the end points of a lift to $\mathbb{R}$ of a loop in S^1 may differ by an integer. However, whether or not a loop lifts to a loop depends only on its epph class.

Corollary 5.0.8. *(i) In the situation of Theorem 5.0.5, suppose a_0 and a_1 are paths in Y based at y_0 and are epph. Then their lifts to $(\tilde{Y}, \tilde{y}_0)$ are epph.*
(ii) If a_0 and a_1 are loops in (Y, y_0), and are epph to one another, and if the lift of a_0 to $(\tilde{Y}, \tilde{y}_0)$ is a loop, then so is the lift of a_1 to $(\tilde{Y}, \tilde{y}_0)$.
(iii) If the loop a in (Y, y_0) is trivial (i.e. epph to the constant loop) then its lift to $(\tilde{Y}, \tilde{y}_0)$ is a loop.

Proof. (i) Let $F : [0,1] \times [0,1] \to Y$ be an epph between a_0 and a_1, and let $\tilde{a}_0$ be the lift of a_0 to $(\tilde{Y}, \tilde{y}_0)$. By the Theorem, there is a lift $\tilde{F}$ of F, extending the lift $\tilde{a}_0$; note that $t \mapsto \tilde{F}(t, 1)$ is the lift of a_1 to $(\tilde{Y}, y_0)$.
The map $\tilde{F}$ must be an epph: since $F(0, u) = y_0$ for all u, $\tilde{F}(0, u)$ must lie in the fibre $p^{-1}(y_0)$ for all u, and is therefore constant. Similarly, $F(1, u) = a_0(1) = a_1(1)$ is constant, so $\tilde{F}(1, u) \in p^{-1}(a_0(1))$ for all u, and is also constant.
(ii) This is just the special case of (i) where $a_0(0) = a_0(1)$. By the argument of (i), $\tilde{F}(0, u)$ and $\tilde{F}(1, u)$ must be equal to $\tilde{y}_0$ for all u. Thus, $t \mapsto \tilde{F}(t, 1)$ is a loop in $(\tilde{Y}, \tilde{y}_0)$.

(iii) This is a special case of (ii): the lift to $(\tilde{Y}, \tilde{y}_0)$ of the constant loop at y_0 is the constant loop at $\tilde{y}_0$. In particular it is a loop. So if a is epph to the constant loop, its lift is a loop. □

Corollary 5.0.9. *For any covering map $p : \tilde{Y} \to Y$ with $p(\tilde{y}_0) = y_0$,*
(i) The homomorphism $p_* : \pi_1(\tilde{Y}, \tilde{y}_0) \to \pi_1(Y, y_0)$ is injective, and
(ii) $p_*\big(\pi_1(\tilde{Y}, \tilde{y}_0)\big) = \{[a] \in \pi_1(Y, y_0) :$ the lift of a to $(\tilde{Y}, \tilde{y}_0)$ is a loop.$\}$

Proof. (i) An epph class $[\tilde{a}] \in \pi_1(\tilde{Y}, \tilde{y}_0)$ lies in the kernel of p_* if there is an epph $F : [0,1] \times [0,1]$ such that $F(t, 0) = p \circ \tilde{a}(t), F(t, 1) = y_0$. By Theorem 5.0.5, any such F lifts to an epph $\tilde{F}$ between $\tilde{a}$ and the constant loop based at $\tilde{y}_0$. So $[\tilde{a}]$ is the epph class of the constant loop.
(ii) Note that the right hand side is well-defined, by Corollary 5.0.8(ii). To

prove the equality, it's obvious that the left hand side contains the right hand side: if $[a] \in \pi_1(Y, y_0)$ and a lifts to a loop $\tilde{a}$ based at $\tilde{y}_0$, then $[a] = p_*([\tilde{a}]) \in p_*(\pi_1(\tilde{Y}, \tilde{y}_0))$. Conversely, if $[a] = p_*[\tilde{b}]$ with $[a] \in \pi_1(Y, y_0)$ and $[\tilde{b}]$ in $\pi_1(\tilde{Y}, \tilde{y}_0)$ then a is epph to $p \circ \tilde{b}$; so by 5.0.5, its lift $\tilde{a}$ to $(\tilde{Y}, \tilde{y}_0)$ is epph to $\tilde{b}$, and is therefore a loop. □

5.0.4 Solving the lifting problem

Suppose that $p : \tilde{Y} \to Y$ is a covering map and that $f : X \to Y$ is a continuous map, and pick basepoints $\tilde{y}_0 \in \tilde{Y}, y_0 \in Y$ and $x_0 \in X$ such that $p(\tilde{y}_0) = y_0 = f(x_0)$. Suppose there is a lift $\tilde{f}$ of f, with $\tilde{f}(x_0) = \tilde{y}_0$. Then there are commutative diagrams

$$\begin{tikzcd} & (\tilde{Y}, \tilde{y}_0) \arrow[d, "p"] \\ (X, x_0) \arrow[ur, "\tilde{f}"] \arrow[r, "f"'] & (Y, y_0) \end{tikzcd} \qquad \begin{tikzcd} & \pi_1(\tilde{Y}, \tilde{y}_0) \arrow[d, "p_*"] \\ \pi_1(X, x_0) \arrow[ur, "\tilde{f}_*"] \arrow[r, "f_*"'] & \pi_1(Y, y_0) \end{tikzcd} \tag{5.0.3}$$

From the second diagram it is clear that

$$f_*(\pi_1(X, x_0)) \subset p_*(\pi_1(\tilde{Y}, \tilde{y}_0)). \tag{5.0.4}$$

This is thus a necessary condition for the existence of the lift $\tilde{f}$. For instance, when $p : \tilde{Y} \to Y$ is $\exp : \mathbb{R} \to S^1$, the right hand side of the inclusion (5.0.4) is the trivial group 1. Thus if $f_*(\pi_1(X, x_0)$ is not trivial then $f : X \to S^1$ cannot be lifted.

A space X is *locally path connected* if for each $x \in X$, every open neighbourhood U of x contains a path connected neighbourhood of x. For example, every open set in $\mathbb{R}^n$ is locally path connected. It follows that the same is true of manifolds (Definition 2.2.3), in many respects the most important spaces in geometric topology. Path-connectedness does not imply local path-connectedness, nor does the opposite implication hold. An exercise in Exercises 4 Section C constructs an example of a path-connected space which is not locally path connected. In the theory of covering maps, local path-connectedness is almost always assumed.

Exercise 5.0.10. *Show that if $p : \tilde{Y} \to Y$ is a covering map and Y is locally path-connected then so is $\tilde{Y}$.*

Theorem 5.0.11. *Suppose X is path-connected and locally path connected. Then* (5.0.4) *is also a sufficient condition for the existence of a lift $\tilde{f}$ sending x_0 to $\tilde{y}_0$.*

Proof. We will define $\tilde{f}$ as follows: let $x \in X$; pick any path a in X from x_0 to x, let $\widetilde{f \circ a}$ be the lift of $f \circ a$ to $(\tilde{Y}, \tilde{y}_0)$, and set $\tilde{f}_a(x) = \tilde{a}(1)$. We will show that for any other path b from x_0 to x,

$$\tilde{f}_a(x) = \tilde{f}_b(x), \tag{5.0.5}$$

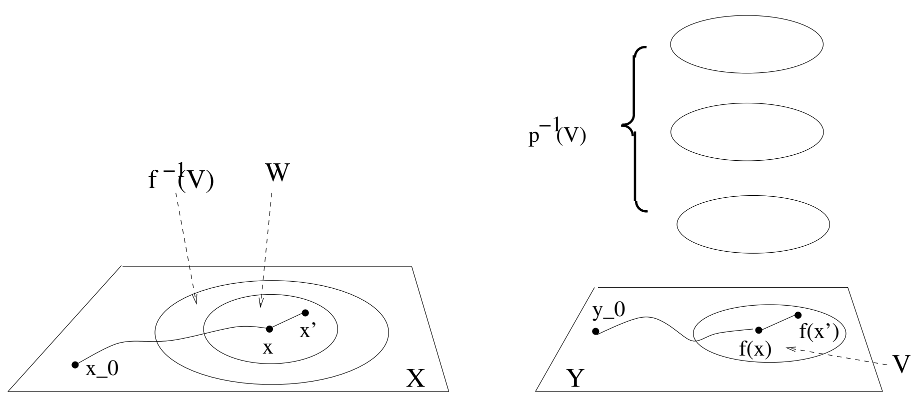
i.e. the end point of the lift of $f \circ a$ to $(\tilde{Y}, \tilde{y}_0)$ is independent of the choice of path a, and depends only on its endpoint, x. We will then define $\tilde{f}(x) = \tilde{f}_a(x)$.

To prove (5.0.5), observe that $a * b^{-1}$ is a loop in X based at x_0, and $f \circ (a * b^{-1}) = (f \circ a) * (f \circ b^{-1})$ is a loop in Y based at y_0. As $[f \circ (a * b^{-1})] = f_*[a * b^{-1}] \in f_*(\pi_1(X, x_0)) \subset p_*(\pi_1(\tilde{Y}, \tilde{y}_0))$, it follows by 5.0.9(ii) that the lift of $f \circ (a * b^{-1})$ to $(\tilde{Y}, \tilde{y}_0)$ is a loop. Let $\widetilde{f \circ b^{-1}}$ be the lift of $f \circ b^{-1}$ to $(\tilde{Y}, \widetilde{f \circ a}(1))$. We have

$$(f \circ a)\widetilde{* (f \circ} b^{-1}) = \widetilde{f \circ a} * \widetilde{f \circ b^{-1}},$$

by uniqueness of path-lifting. The fact that this is a loop means that $\widetilde{f \circ b^{-1}}$ is a path from $\widetilde{f \circ a}(1)$ to $\tilde{y}_0$. But then this same path, traversed backwards, is a path from $\tilde{y}_0$ to $\widetilde{f \circ a}(1)$. Thus, the lift of $f \circ b$ to $(\tilde{Y}, \tilde{y}_0)$ ends at the same point as the lift of $f \circ a$ to $(\tilde{Y}, \tilde{y}_0)$, and (5.0.5) is proved.

Continuity is easy to show – and is the point at which we use local path connectedness of X. Given x, pick a path a from x_0 to x, pick a well-covered open set $V \subset Y$ containing $f(x)$, and then pick a path-connected open neighbourhood W of x with $W \subset f^{-1}(V)$.



Each point x' in W can be joined to x_0 by a path of the form $a * b$, where b is a path in W from x to x'. The point $\tilde{f}(x)$ lies in one of the homeomorphic preimages U_α of V in $\tilde{Y}$. Then the lift of $f \circ b$ to $(\tilde{Y}, \tilde{f}(x))$ coincides with

$(p|_{U_\alpha})^{-1} \circ f \circ b$, and thus on W, $\tilde{f} = (p|_{U_\alpha})^{-1} \circ f$. This shows that $\tilde{f}$ is continuous at x.

□

Corollary 5.0.12. *Suppose that X is simply connected and locally path-connected. Given a continuous map $f : (X, x_0) \to (Y, y_0)$, a covering map $p : \tilde{Y} \to Y$, and a point $\tilde{y}_0 \in p^{-1}(y_0)$, there exists a unique lift $\tilde{f} : (X, x_0) \to (\tilde{Y}, \tilde{y}_0)$ of f.*

Proof. As $\pi_1(X, x_0) = \{1\}$, condition (5.0.4) holds. □

Notice that Proposition 4.1.1 and Corollary 4.1.3 are a special case of this.

5.1 Classification and Construction of Covering Spaces

A covering map $p : \tilde{Y} \to Y$ is *trivial* if, up to homeomorphism, $\tilde{Y}$ is just the product of Y with a fibre $p^{-1}(y_0)$. That is, if there is a homeomorphism Φ making the diagram

$$\begin{tikzcd} \tilde{Y} \arrow[rr, "\Phi"] \arrow[dr, "p"'] & & Y \times p^{-1}(y_0) \arrow[dl, "\mathrm{pr}"] \\ & Y & \end{tikzcd} \tag{5.1.1}$$

commute. In this case $\tilde{Y}$ is just a collection of $|p^{-1}(y_0)|$ disjoint copies of Y.

Proposition 5.1.1. *Suppose that Y is simply connected*[1] *and locally path-connected. Then any covering map $p : \tilde{Y} \to Y$ is trivial.*

Proof. If $\tilde{y} \in \tilde{Y}$, I claim there is just one point in $p^{-1}(y_0)$ to which $\tilde{y}$ can be connected by a path in $\tilde{Y}$. There is at least one point, for $p(y)$ can be joined to $p(y_0)$ by a path a in Y, and a has a unique lift to a path $\tilde{a}$ in $(\tilde{Y}, \tilde{y})$. Clearly $\tilde{a}(1) \in p^{-1}(y_0)$.

Now suppose $\tilde{y}$ can be joined by paths $\tilde{a}$ and $\tilde{b}$ to two distinct points $\tilde{y}_1$ and $\tilde{y}_2$ in $p^{-1}(y_0)$. Then $(p \circ \tilde{a}) * (p \circ \tilde{b}^{-1})$ is a loop in (Y, y_0). This loop is epph to the constant loop, because $\pi_1(Y, y_0)$ is trivial. But its lift, $\tilde{a} * \tilde{b}^{-1}$, is *not* a loop, as its end points, $\tilde{y}_1$ and $\tilde{y}_2$ are distinct. This contradicts Corollary 5.0.8(iii).

[1] Recall that to say that X is simply connected means that it is path connected as well as having trivial fundamental group.

Let $\tilde{y}_\alpha \in p^{-1}(y_0)$, and let $\tilde{Y}_\alpha$ be the set of points to which $\tilde{y}_\alpha$ can be joined by a path in $\tilde{Y}$. That is, $\tilde{Y}_\alpha$ is is the path-component of $\tilde{Y}$ containing $\tilde{y}_a$. It is open in $\tilde{Y}$, because path-components of a locally path-connected space are open (see Exercises 4). I claim that $p| : \tilde{Y}_\alpha \to Y$ is a homeomorphism. To see this, we use Theorem 5.0.11 to construct a right-inverse.

Consider the lifting problem

$$\begin{tikzcd} & \tilde{Y} \arrow[d, "p"] \\ Y \arrow[r, "\mathrm{id}_Y"'] \arrow[ur, dashed] & Y \end{tikzcd} \tag{5.1.2}$$

Because Y is simply connected, the condition (5.0.4) obviously holds; that is,

$$\mathrm{id}_{Y*}(\pi_1(Y, y_0)) \subset p_*(\pi_1(\tilde{Y}, \tilde{y}_\alpha).$$

Hence there is a (unique) lift $\widetilde{\mathrm{id}_Y} : Y \to \tilde{Y}$ of id_Y, sending y_0 to $\tilde{y}_\alpha$. This lift must map Y into $\tilde{Y}_\alpha$, since Y is path connected and it maps y_0 to $\tilde{y}_a \in \tilde{Y}_\alpha$. It must be surjective, for if $\tilde{y} \in \tilde{Y}_\alpha$, then there is a path $\tilde{a}$ in $\tilde{Y}_\alpha$ from $\tilde{y}_\alpha$ to $\tilde{y}$. Both $\widetilde{\mathrm{id}_Y} \circ p \circ \tilde{a}$ and $\tilde{a}$ are lifts of $p \circ \tilde{a}$ starting at $\tilde{y}_\alpha$. It follows by the uniqueness of path-lifting that they are equal, so $\tilde{y}$ is in the image of $\widetilde{\mathrm{id}_Y}$.

Write p_α in place of $p|_{\tilde{Y}_\alpha}$. Because $p_\alpha \circ \widetilde{\mathrm{id}_Y} = \mathrm{id}_Y$ and $\widetilde{\mathrm{id}_Y}$ is surjective, it follows [2] also that that $\widetilde{\mathrm{id}_Y} \circ p_\alpha = \mathrm{id}_{\tilde{Y}_\alpha}$. We have proved that p_α and $\widetilde{\mathrm{id}_Y}$ are mutually inverse. As both are continuous, each is a homeomorphism.

Thus $\tilde{Y}$ is a disjoint union of open sets $\tilde{Y}_\alpha$, each of which is homeomorphic, via p, to Y itself. There is one of these sets for each point of $p^{-1}(y_0)$. This proves the proposition: a homeomorphism $\tilde{Y} \to Y \times p^{-1}(y_0)$ can be defined on each $\tilde{Y}_\alpha$ separately, mapping $\tilde{y} \in \tilde{Y}_\alpha$ to $(p(\tilde{y}), \tilde{y}_\alpha)$. □

Example 5.1.2. If $n > 1$, the n-sphere S^n is simply connected, path connected and locally path connected (the latter because it is a manifold). So by the corollary it has no non-trivial covering. This has some interesting consequences.

1. **Exercise** The 3-sphere S^3 is in fact a group: it can be identified with the unit quaternions

$$\mathbb{H}_1 := \{a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} : a^2 + b^2 + c^2 + d^2 = 1\}$$

---

[2]Given a diagram $\begin{tikzcd} X \arrow[r, bend left, "f"] & Y \arrow[l, bend left, "g"] \end{tikzcd}$, and $f \circ g = \mathrm{id}_Y$, then surjectivity of g implies also $g \circ f = \mathrm{id}_X$.

which is a (multiplicative) subgroup of the multiplicative group of non-zero quaternions $\mathbb{H}$. One might expect that for $n \neq 0$, the map $\mathbb{H}_1 \to \mathbb{H}_1$ sending z to z^n would be a covering map, as it is in the case of S^1. But we have just seen that it can not be. What goes wrong?

2. Let $X \subset \mathbb{R}^3$ be a smooth surface. At each point it has two unit normal vectors. It is *orientable* if it is possible to choose one of them at each point in such a way that the resulting "normal vector field" n varies smoothly (and doesn't suddenly jump from one side of the surface to the other). In fact all compact boundaryless surfaces in $\mathbb{R}^3$ are orientable; the Möbius strip is not. If X is orientable, the map

$$X \to S^2, \ x \mapsto n(x)$$

obtained by choosing a continuous unit normal vector field is called the *Gauss map* of X. The *Gauss curvature* of S at a point x, $\kappa(x)$, is defined to be the determinant of the derivative of this map[3]. If $\kappa(x) \neq 0$, the inverse function theorem tells us that the map n is a local diffeomorphism (and hence a local homeomorphism) at x. From this we deduce

Proposition 5.1.3. *If X is a compact oriented surface in $\mathbb{R}^3$ and $\kappa(x) \neq 0$ for all $x \in X$ then $n : X \to S^2$ is a covering map.*

The proof needs the following lemma from point-set topology:

Lemma 5.1.4. *Suppose that $f : M \to N$ is a continuous map with the property that for all $x \in M$, there are open neighbourhoods U of x and V of $f(x)$ (in M and N respectively) such that $f| : U \to V$ is a homeomorphism. Suppose that M is compact and N is connected and Hausdorff. Then*

1. *f is surjective, and*
2. *f is a covering map.*

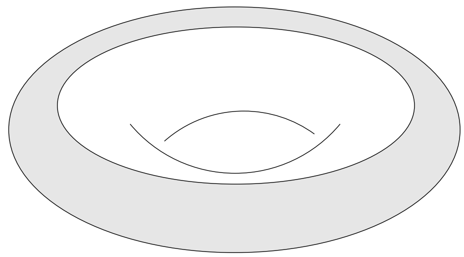
Proof. Exercise; see Exercises 4 Section C for hints. . □

Now applying 5.1.1, we have

[3]Some work is needed to show that this definition makes sense. The key point is that the tangent space to X at x, T_xX, is equal to $n(x)^\perp$ and therefore equal to $T_{n(x)}S^2$. So the derivative at x of the Gauss map, $d_xn : T_xX \to T_{n(x)}S^2$, can be viewed as a map from a vector space to itself. This allows its determinant to be defined: if E is a basis for T_xX and $[d_xn]_E$ is the matrix of $d_xN : T_xX \to T_xX$ written with respect to E, then $\det[d_xN]_E$ is in fact independent of the choice of E.

Corollary 5.1.5. *If $X \subset \mathbb{R}^3$ is a compact connected boundaryless surface whose Gauss curvature κ is nowhere zero then X is homeomorphic to a sphere.*

In fact the homeomorphism provided by 5.1.1 is a diffeomorphism – that is, differentiable with differentiable inverse.



In the shaded region $\kappa > 0$, in the white region $\kappa < 0$

5.1.1 Universal coverings

Throughout this subsection, Y is assumed locally path-connected. Suppose that $p : \tilde{Y} \to Y$ is a covering map, and that $\tilde{Y}$ is simply connected. I claim that $\tilde{Y}$ is at the top of a hierarchy of covering spaces of Y. Let $q : \tilde{Z} \to Y$ be a second covering map, with $\tilde{Z}$ connected. Let us choose base points $\tilde{y}_0 \in \tilde{Y}$, $\tilde{z}_0 \in \tilde{Z}$ and y_0 in Y, with $p(\tilde{y}_0) = y_0 = q(\tilde{z}_0)$. Consider the diagram

$$\begin{tikzcd} & \tilde{Z} \arrow[d, "q"] \\ \tilde{Y} \arrow[r, "p"'] \arrow[ur, dashed, "\tilde{p}"] & Y \end{tikzcd} \tag{5.1.3}$$

Because $\tilde{Y}$ is simply connected, we have

$$p_*(\pi_1(\tilde{Y}, \tilde{y}_0)) \subset q_*(\pi_1(\tilde{Z}, \tilde{z}_0)). \tag{5.1.4}$$

It follows by Theorem 5.0.11 that there is a unique lift of p, $\tilde{p} : (\tilde{Y}, \tilde{y}_0) \to (\tilde{Z}, \tilde{z}_0)$; that is,

$$q \circ \tilde{p} = p, \;\; \tilde{p}(\tilde{y}_0) = \tilde{z}_0. \tag{5.1.5}$$

Proposition 5.1.6. *$\tilde{p}$ is a covering map.*

Proof. We show first that $\tilde{p}$ is surjective. Let $\tilde{z} \in \tilde{Z}$ and let a be a path in $\tilde{Z}$ from $\tilde{z}_0$ to $\tilde{z}$. Then $q \circ a$ is a path in Y from $q(\tilde{z}_0) = y_0$ to $q(\tilde{z})$. This path has a (unique) lift to a path $\widetilde{q \circ a}$ in $(\tilde{Y}, \tilde{y}_0)$ (here we use the fact that p is a covering); that is, such that

$$p \circ \widetilde{q \circ a} = q \circ a, \quad \text{and } \widetilde{q \circ a}(0) = \tilde{y}_0. \tag{5.1.6}$$

Claim: The path $\tilde{p} \circ \widetilde{q \circ a}$ is a lift of $q \circ a$ to $(\tilde{Z}, \tilde{z}_0)$, i.e.

$$q \circ \tilde{p} \circ \widetilde{q \circ a} = q \circ a.$$

To see this, we simply calculate:

$$\begin{aligned} q \circ \tilde{p} \circ \widetilde{q \circ a} &= p \circ \widetilde{q \circ a} \quad \text{by (5.1.5)} \\ &= q \circ a \quad \text{by (5.1.6)} \end{aligned}$$

Clearly $\tilde{p} \circ \widetilde{q \circ a}(0) = \tilde{p}(\tilde{y}_0) = \tilde{z}_0$, so the claim is proved.

Now, the original path a is also a lift of $q \circ a$ to $(\tilde{Z}, \tilde{z}_0)$, and we have

$$a(0) = \tilde{z}_0 = \tilde{p} \circ \widetilde{q \circ a}(1);$$

by uniqueness of path lifting, it follows that $a = \tilde{p} \circ \widetilde{q \circ a}$. In particular,

$$\tilde{z} = a(1) = \tilde{p} \circ \widetilde{q \circ a}(1) \in \text{ image of } \tilde{p}.$$

Next we show that $\tilde{z}$ has a neighbourhood in $\tilde{Z}$ which is well-covered by $\tilde{p}$. Let V_p and V_q be neighbourhoods of $q(\tilde{z})$ in Y which are well-covered by p and q respectively. Let $V \subset V_p \cap V_q$ be a path connected neighbourhood of $p(\tilde{z})$. It is well-covered by both p and q.

Let
$$q^{-1}(V) = \bigcup_{\alpha \in A} U_\alpha, \qquad p^{-1}(V) = \bigcup_{\beta \in B} W_\beta,$$
with $q : U_\alpha \to V$ and $p : W_\beta \to V$ homeomorphisms for all α and β. Consider the restriction of $\tilde{p}$ to some W_β. We have $q \circ \tilde{p}(W_\beta) = p(W_\beta) = V$, so $\tilde{p}(W_\beta) \subset \bigcup_\alpha U_\alpha$. By path connectedness, $\tilde{p}(W_\beta) \subset U_\alpha$ for some α, and hence from $q \circ \tilde{p} = p$ we get

$$\big(q|_{U_\alpha}\big) \circ (\tilde{p}|_{W_\beta}) = p|_{W_\beta},$$

and thus
$$\tilde{p}|_{W_\beta} = \big(q|_{U_\alpha}\big)^{-1} \circ \big(p|_{W_\beta}\big),$$
showing that $\tilde{p} : W_\beta \to U_\alpha$ is a homeomorphism. The surjectivity of $\tilde{p}$, shown in the first part of this proof, shows that for each of the U_α in $q^{-1}(V)$ there is at least one W_β mapped to it by $\tilde{p}$. This completes the proof that $\tilde{p}$ is a covering map. $\square$

So every covering space of Y is a quotient of the simply connected covering space $\tilde{Y}$. For this reason $p : \tilde{Y} \to Y$ is called the *universal covering space.* The definite article "the" is justified because any two such are suitably equivalent:

Proposition 5.1.7. *Suppose that $p_1 : Z_1 \to Y$ and $p_2 : Z_2 \to Y$ are covering maps, and we pick base-points z_1, z_2, y_0 such that $p_1(z_1) = p_2(z_2) = y_0$. If*

$$p_{1*}(\pi_1(Z_1, z_1)) = p_{2*}(\pi_1(Z_2, z_2))$$

then there is a unique homeomorphism $h : (Z_1, z_1) \to (Z_2, z_2)$ such that $p_2 \circ h = p_1$.

Proof. Apply Theorem 5.0.11 to the two diagrams

$$\begin{tikzcd} & (Z_1, z_1) \arrow[d, "p_1"] \\ (Z_2, z_2) \arrow[r, "p_2"'] & (Y, y_0) \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} & (Z_2, z_2) \arrow[d, "p_2"] \\ (Z_1, z_1) \arrow[r, "p_1"'] & (Y, y_0) \end{tikzcd}$$

obtaining lifts $\tilde{p}_2 : (Z_2, z_2) \to (Z_1, z_1)$ and $\tilde{p}_1 : (Z_1, z_1) \to (Z_2, z_2)$ respectively. The composites of the two lifts fit in the diagrams

$$\begin{tikzcd} & (Z_1, z_1) \arrow[d, "p_1"] \\ (Z_1, z_1) \arrow[ur, "\tilde{p}_2 \circ \tilde{p}_1"] \arrow[r, "p_1"'] & (Y, y_0) \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} & (Z_2, z_2) \arrow[d, "p_2"] \\ (Z_2, z_2) \arrow[ur, "\tilde{p}_1 \circ \tilde{p}_2"] \arrow[r, "p_2"'] & (Y, y_0) \end{tikzcd}$$

Since id_{Z_1} and id_{Z_2} also fit there, uniqueness of lifting implies that $\tilde{p}_2 \circ \tilde{p}_1 = \mathrm{id}_{Z_1}$ and $\tilde{p}_1 \circ \tilde{p}_2 = \mathrm{id}_{Z_2}$. □

We call two coverings of the same space *isomorphic* if there exists a homeomorphism h as in the Proposition.

Example 5.1.8.

1. $\exp : \mathbb{R} \to S^1$ is the universal covering of S^1, and the map $\mathbb{R}^2 \to T$ obtained by composing $\exp \times \exp$ with a homeomorphism $S^1 \times S^1 \to T$ is the universal covering of the torus T.

2. The quotient map $q : S^n \to \mathbb{R}\mathbb{P}^n$ is the universal covering of $\mathbb{R}\mathbb{P}^n$ for $n \geq 2$.

3. What is the universal cover of an oriented surface X_g of genus $g \geq 2$? The answer is rather striking: there is a group G_g of hyperbolic isometries acting in the hyperbolic plane $\mathbb{H}^2$ such that the quotient of $\mathbb{H}^2$ by (the equivalence relation induced by the action of) G_g is X_g. The construction is rather down-to-earth but ingenious, and the following brief summary omits a number of quite delicate details. First find a polygon P_g such that X_g is the result of identifying some of the sides of P_g with one another (in the same way that the torus, Klein bottle and projective plane can all be obtained as quotients of a square). See the Appendix to Chapter 6, which can be read without reference to the rest of the chapter. We then situate P_g in $\mathbb{H}^2$ with (hyperbolic) geodesics for its edges. The following diagram shows a regular hyperbolic 8-gon in the Poincaré Disc model of the hyperbolic plane. Here the geodesics are arcs of circles meeting the boundary at right angles.

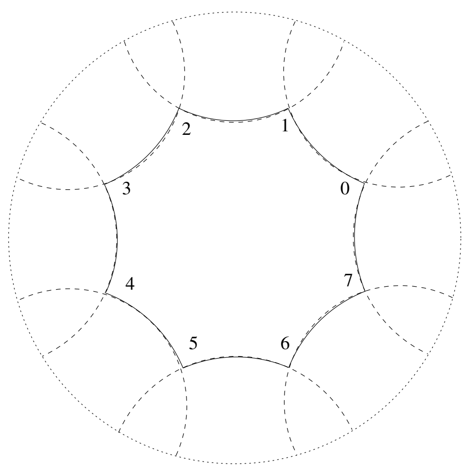


Figure 6.7

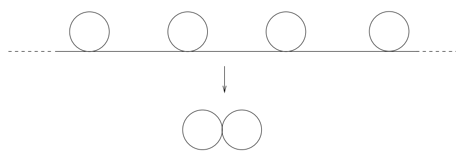
The hyperbolic reflections in the edges of P_g now generate a group G_g of hyperbolic isometries. By adjusting the angles at the vertices (by dilating or shrinking P_g) we can ensure that the group action carries out the required edge-identifications; once this is so, the translates of P_g by the group action tessellate $\mathbb{H}^2$ (cover it without overlaps). The quotient map $\mathbb{H}^2 \to \mathbb{H}^2/G_g \simeq X_g$ is then a covering map. Since $\mathbb{H}^2$ is homeomorphic to $\mathbb{R}^2$, it is simply connected and thus is the universal covering space of X_g.

This construction does more than just find the universal cover of X_g: since the group G_g acts by hyperbolic isometries, it shows one can endow X_g with a hyperbolic metric giving it constant curvature -1 (like $\mathbb{H}^2$).

4. The quotient map
$$[0,1] \times [0,1] \to T$$
extends to a covering map $\mathbb{R}^2 \to T$. Does the quotient map $[0,1] \times [0,1] \to \mathbb{R}\mathbb{P}^2$ also extend to a covering map $\mathbb{R}^2 \to T$?

5. What about the quotient map $[0,1] \times [0,1] \to K = ([0,1] \times [0,1])/\sim$ determined by the fifth square in Example (5) of Chapter 3?

6. There is a covering map over the figure 8,



whose domain is a line with a circle added at each integer point. Provide the details.

7. Draw a simply connected covering space over the figure 8.

8. Draw a covering of finite degree over the figure 8.

We end this section with a proof that reasonable spaces have universal coverings. We need a new definition:

Definition 5.1.9. The space X is *semi-locally simply connected* if each point $x_0 \in X$ has a neighbourhood U such that the homomorphism

$$\pi_1(U, x_0) \to \pi_1(X, x_0)$$

induced by the inclusion of U in X is trivial – i.e. sends all of $\pi_1(U, x_0)$ to the neutral element.

Once again, open sets in $\mathbb{R}^n$ have this property (indeed, here every point has a simply connected neighbourhood), and so manifolds have it also. It is easy to see that it is a necessary condition for the existence of a simply connected covering space. For if $p : \tilde{X} \to X$ is a covering map with $\tilde{X}$

simply connected, if $x \in X$, and if V is a well-covered neighbvourhood of x in X, let $U_\alpha \subset \tilde{X}$ be mapped homeomorphically to V by p. There is a commutative diagram of inclusions and covering maps

$$\begin{tikzcd} U_\alpha \arrow[r] \arrow[d, "p|"', "\simeq"] & \tilde{X} \arrow[d, "p"] \\ V \arrow[r] & X \end{tikzcd} \qquad \text{inducing} \qquad \begin{tikzcd} \pi_1(U, \tilde{x}) \arrow[r] \arrow[d, "p|_*"', "\simeq"] & \pi_1(\tilde{X}, \tilde{x}) = \{1\} \arrow[d, "p_*"] \\ \pi_1(V, x) \arrow[r] & \pi_1(X, x) \end{tikzcd} .$$

This shows that the homomorphism $\pi_1(V, x) \to \pi_1(X, x)$ is trivial.

Declare an open set $V \subset X$ to be *worthy* (for the purposes of the forthcoming theorem) if V is path-connected and for some $x \in V$, $\pi_1(V, x) \to \pi_1(X, x)$ is trivial.

Exercise 5.1.10. 1. Use Proposition 4.2.3 to show that if V is path connected and $\pi_1(V, x) \to \pi_1(X, x)$ is trivial for *some* $x \in V$ then $\pi_1(V, x) \to \pi_1(X, x)$ is trivial for *all* $x \in V$.

2. Show that if V is worthy and $V' \subset V$ is path connected then V' is worthy also.

Theorem 5.1.11. *Let X be path connected, locally path connected and semi-locally simply connected. Then X has a universal cover $\tilde{X}$. That is, there is a covering map $\tilde{X} \to X$ with $\tilde{X}$ simply connected.*

The idea of the proof is simple. Fix a base-point $x_0 \in X$. Every point x of X can be reached from x_0 by a path. Indeed, there will be many paths from x_0 to x. We want to regard as different the point x reached by path a and the point x reached by path b, *if a and b are not epph.*

This idea can be understood clearly in the case of the circle S^1. Take $x_0 = 1$ as basepoint. We can distinguish between the point -1, arrived at by rotating through π (in the positive direction), and the point -1 arrived at by rotating through $-\pi$, by denoting the first point by $(-1, \pi)$ and the second by $(-1, -\pi)$. Similarly, we can distinguish between the point 1 arrived at by staying put, and the point 1 arrived at by rotating though 2π, by writing $(1, 0)$ for the first and $(1, 2\pi)$ for the second. In this way we get many "points" (x, θ) from each each point $x \in S^1$, with θ coordinates differing by multiples of 2π. In fact, the second coordinate of each "point" (x, θ) is the end point of the lift (with respect to $\exp : \mathbb{R} \to S^1$) of the path in S^1 (the rotation) undertaken to reach the point[4] . Since the value of θ determines

[4]Except, that is, for a factor of 2π, because of the fact that we use the version of exp defined by $\exp(t) = e^{2\pi i t}$ instead of e^{it}.

the point x, we can dispense with the notation (x, θ) and simply denote the new points by their θ coordinate.

Note that the angle θ does not determine the path taken from 1 to x, but just its epph class.

In the proof of the theorem, we therefore take, as the set $\tilde{X}$, the set of epph classes of paths based at x_0. An epph class $[a]$ has a well-defined end-point, which we denote by $p([a])$, so there is a well-defined map $p : \tilde{X} \to X$, evidently surjective given the path connectdness of X. Note that as a consequence of this definition, $p^{-1}(x_0)$ is in bijection with $\pi_1(X, x_0)$.

Proof. of Theorem 5.1.11
We have to give the set $\tilde{X}$ a topology. Let $x \in X$, let a be a path in X from x_0 to x, and let V be a path connected neighbourhood of x. For each $x' \in V$, pick a path b in V from x to x'. Then $a * b$ is a path in X from x_0 to x'. Let $V_{[a]}$ be the set of epph classes of all such paths,

$$V_{[a]} = \{[a * b] : b \text{ is a path in } V \text{ with } b(0) = x\} \tag{5.1.7}$$

We call such a set $V_{[a]}$ *preworthy* if V is worthy.

We define a subset U of $\tilde{X}$ to be open if for each $[c] \in U$ there is a preworthy $V_{[a]}$ containing $[c]$ and contained in U. It is not hard to check that this defines a topology on $\tilde{X}$.

It remains to show that p is continuous, that each point in X has a well covered neighbourhood (so that p is a covering map), and that $\tilde{X}$ is simply connected.

Continuity: Let $[a] \in \tilde{X}$, and let $p([a]) = x$ (so x is the endpoint of the path a.) Let W be a neighbourhood of x in X. Let V be a worthy neighbourhood of x contained in W. Then

$$[a] \in V_{[a]} \subset p^{-1}(W).$$

This shows that $p^{-1}(W)$ is open in $\tilde{X}$.

Existence of well-covered neighbourhoods: If $V \subset X$ is worthy then V is well-covered. In fact

$$p^{-1}(V) = \bigcup V_{[a]}$$

where the union is over homotopy classes of paths in X from x_0 to a point $x \in V$. If a and b are paths from x_0 to points x_1 and x_2 of V, then either b is epph to $a * c$, where c is a path in V from x_1 to x_2, (it doesn't matter which, by the worthiness of V), or not. In the first case, $V_{[a]} = V_{[b]}$. In the second case, $V_{[a]} \cap V_{[b]} = \emptyset$. For each set $V_{[a]}$, $p| : V_{[a]} \to V$ is a homeomorphism, with inverse sending $x' \in V$ to $[a * b]$, where b is a path in V from x to x'.

Simple-connectedness of $\tilde{X}$: For $[c] \in \tilde{X}$, let c_t be the path in X defined by $c_t(u) = c(tu)$ (for $u \in [0,1]$). The map

$$t \mapsto [c_t]$$

is a path in $\tilde{X}$ lifting c, which starts at $[x_0]$, the epph class of the constant path at x_0, and which ends at $[c]$. Since $[c]$ is an arbitrary point of $\tilde{X}$, this shows that $\tilde{X}$ is path connected. To show that $\pi_1(\tilde{X}, [x_0]) = \{1\}$, it is enough to show that $p_*(\pi_1(\tilde{X}, [x_0]) = \{1\}$, since p_* is injective. Elements in the image of p_* are represented by loops c in (X, x_0) that lift to loops $\tilde{c}$ in $(\tilde{X}, [x_0])$. We have seen that the path $t \mapsto [c_t]$ is a lift of $[c]$ to $(\tilde{X}, [x_0])$. For this lifted path to be a loop means that $[c_1] = [x_0]$. Since $c_1 = c$, this says that $[c] = [x_0]$. That is, $[c]$ lifts to a loop only if it is epph to the constant loop $[x_0]$. Hence $p_*(\pi_1(\tilde{X}, [x_0]) = \{1\}$. $\square$

Exercise 5.1.12. Show that the definition of a topology on $\tilde{X}$ given in the proof really does define a topology, by showing that

1. the intersection of two worthy sets is open according to the definition, and

2. that this is all that is needed.

5.1.2 Deck transformations

Much of the material in this section is set out as a series of worked exercises.

If $p : \tilde{X} \to X$ is a covering map, a *deck transformation* is a homeomorphism $\tilde{X} \to \tilde{X}$ such that $p \circ h = p$ – i.e. giving a commutative diagram

$$\begin{tikzcd} \tilde{X} \arrow[rr, "h"] \arrow[dr, "p"'] & & \tilde{X} \arrow[dl, "p"] \\ & X & \end{tikzcd}$$

Leaning this diagram over on its side, as

$$\begin{tikzcd} & \tilde{X} \arrow[d, "p"] \\ \tilde{X} \arrow[ur, "h"] \arrow[r, "p"'] & X \end{tikzcd} \tag{5.1.8}$$

we see that a deck transformation is nothing but a lift of p.

Example 5.1.13. *1. For the covering $\mathbb{R}^2 \to T$ of the torus, all integer translations $h(x_1, x_2) = (x_1 + n_1, x_2 + n_2)$ ($n_1, n_2 \in \mathbb{Z}$) are deck transformations.*

2. For the quotient map $S^n \to \mathbb{R}\mathbb{P}^n$, the antipodal map is a deck transformation.

Exercise 5.1.14. 1. For any $x_0 \in X$, any deck transformation maps $p^{-1}(x_0)$ to $p^{-1}(x_0)$.

2. The set of deck transformations is a group under the binary operation of composition. We will denote this group by Δ.

From now on, assume

- **$\tilde{X}$ is path-connected**
- **X is path connected, locally path-connected**
- **X is semi-locally simply connected**

3. Any two deck transformations agreeing on one point must agree everywhere.

4. A deck transformation induces a bijection $p^{-1}(x_0) \to p^{-1}(x_0)$.

5. Let $\tilde{x}_0, \tilde{x}_1 \in p^{-1}(x_0)$. By Theorem 5.0.11, a necessary and sufficient condition for the existence of a deck transformation sending $\tilde{x}_1$ to $\tilde{x}_0$ is that
$$p_*(\pi_1(\tilde{X}, \tilde{x}_1)) \subset p_*(\pi_1(\tilde{X}, \tilde{x}_0)). \tag{5.1.9}$$
Show that if $\tilde{X} \to X$ is the universal cover (i.e. $\tilde{X}$ is simply connected) then for every $\tilde{x}_\alpha, \tilde{x}_\beta \in p^{-1}(x_0)$, there is a deck transformation sending $\tilde{x}_\alpha$ to $\tilde{x}_\beta$. Hint: apply Theorem 5.0.11 to get maps $h_\alpha : (\tilde{X}, \tilde{x}_\alpha) \to (\tilde{X}, \tilde{x}_\beta)$ and $k_\alpha : (\tilde{X}, \tilde{x}_\beta) \to (\tilde{X}, \tilde{x}_\alpha)$ such that $p \circ h_\alpha = p$ and $p \circ k_\alpha = p$. Then we must have $h_\alpha \circ k_\alpha = \mathrm{id}_{\tilde{Y}} = k_\alpha \circ h_\alpha$ (Why?) so h_α and k_α are mutually inverse homeomorphisms.

6. Now relinquish the supposition that $\tilde{X}$ is simply connected. Let a be a path in $\tilde{X}$ from x_1 to x_0. Deduce from Proposition 4.2.3 that (5.1.9) holds if and only if
$$[\gamma] * (p_*(\pi_1(\tilde{X}, \tilde{x}_0)) * [\gamma]^{-1} \subset p_*(\pi_1(\tilde{X}, \tilde{x}_0)) \tag{5.1.10}$$
where γ is the loop $p \circ a$.

7. Deduce that if $p_*(\pi_1(\tilde{X}, \tilde{x}_0))$ is a normal subgroup of $\pi_1(X, x_0)$ then for each $\tilde{x}_1 \in p^{-1}(x_0)$, there is a unique deck transformation sending $\tilde{x}_1$ to $\tilde{x}_0$.

A covering $p : (\tilde{X}, \tilde{x}_0) \to (X, x_0$ is said to be **normal** *if $p_*(\pi_1(\tilde{X}, \tilde{x}_0))$ is a normal subgroup of $\pi_1(X, x_0)$.*

8. Whether or not the covering is normal does not depend on the choice of base-point $\tilde{x}_0$.

9. The universal covering of $\tilde{X}$ is normal.

10. For each loop γ in (X, x_0), let $\tilde{\gamma}$ be a lift to $(\tilde{X}, \tilde{x}_0)$. The procedure
$$[\gamma] \mapsto \tilde{\gamma}(1)$$
determines a map $\pi_1(X, x_0) \to p^{-1}(x_0)$. Show that it is surjective.

11. Show that if $\tilde{X}$ is simply connected (i.e. if $\tilde{X} \to X$ is the universal covering) then the mapping $\pi_1(X, x_0) \to p^{-1}(x_0)$ just described is also injective, and the composite
$$\begin{array}{ccccc} \pi_1(X, x_0) & \to & p^{-1}(x_0) & \to & \Delta \\ [\gamma] & \mapsto & \tilde{\gamma}(1) & \mapsto & \text{deck trans. sending } \tilde{x}_0 \text{ to } \tilde{x}_1 \end{array}$$
is a bijection. In fact this was already made explicit in the construction of the universal covering – see the remark just before the proof of Theorem 5.1.11. The inverse of this map is easy to understand: a deck transformation h sends $\tilde{x}_0$ to some other point of $p^{-1}(x_0)$. Join $\tilde{x}_0$ to $h(\tilde{x}_0)$ by a path γ in $\tilde{X}$. As $\tilde{X}$ is simply connected, any two such paths are epph in $\tilde{X}$. Then $p \circ \gamma$ is a loop in (X, x_0), and determines an element of $\pi_1(X, x_0)$. Denote the map $\Delta \to \pi_1(X, x_0)$ just defined by θ.

12. θ is an isomorphism of groups. Let h and k be deck transformations, and let γ and σ be paths in $\tilde{X}$ from $\tilde{x}_0$ to $h(\tilde{x}_0)$ and $k(\tilde{x}_0)$ respectively. Then $h \circ \sigma$ is a path in $\tilde{X}$ from $h(\tilde{x}_0)$ to $h(k(\tilde{x}_0))$. Thus $\gamma * h \circ \sigma$ is a path from $\tilde{x}_0$ to $h(k(\tilde{x}_0))$, and

$$\theta(h \circ k) = [p \circ (\gamma * (h \circ \sigma))] = [p \circ \gamma] * [p \circ h \circ \sigma] = [p \circ \gamma] * [p \circ \sigma].$$

The last statement of the exercise is worth restating as a theorem.

Theorem 5.1.15. *Suppose that X is path-connected, locally path connected and semi-locally simply connected. If $p : (\tilde{X}, \tilde{x}_0) \to (X, x_0)$ is the universal covering map then $\theta : \Delta \to \pi_1(X, x_0)$ is an isomorphism.*

Example 5.1.16.

1. The covering $\exp : \mathbb{R} \to S^1$ is normal; the deck transformations are integer translations.

2. All coverings of the torus are normal, because $\pi_1(T, x_0)$ is abelian. So for every covering of T, there are plenty of deck transformations.

3. The universal cover of $\mathbb{R}\,\mathbb{P}^2$ is S^2, since the latter is simply connected. The group of deck transformations consists just of the identity and the antipodal map, so is isomorphic to $\mathbb{Z}/2\,\mathbb{Z}$. Hence we have another proof that $\pi_1(\mathbb{R}\,\mathbb{P}^n, x_0) \simeq \mathbb{Z}/2\,\mathbb{Z}$.

4. Find a non-normal covering of the figure 8. (See Hatcher, page 71, first paragraph, for ideas).

5.1.3 Further developments

A covering for every subgroup of $\pi_1(X, x_0)$

If X meets the conditions of Theorem 5.1.11 and therefore has a simply connected covering space, then the map

$$\{\text{covering spaces over } (X, x_0)\} \to \{\text{subgroups of } \pi_1(X, x_0)\}$$

defined by mapping the covering $p : (\tilde{X}, \tilde{x}_0) \to (X, x_0)$ to $p_*(\pi_1(\tilde{X}, \tilde{x}_0))$ determines a bijection between isomorphism classes of coverings of (X, x_0) (see the definition after Proposition 5.1.7) and subgroups of $\pi_1(X, x_0)$. Injectivity was already shown in Proposition 5.1.7. To prove surjectivity it is necessary to construct a covering for each subgroup of $\pi_1(X, x_0)$. The construction of the universal cover can be viewed simply as the case of this construction for the subgroup $\{1\}$, since $p : \tilde{X} \to X$ is a universal cover if and only if $p_*(\pi_1(\tilde{X}, \tilde{x}_0)) = \{1\}$ (by Corollary 5.0.9(i)). In fact the construction for the remaining cases is an easy step onwards from this special case (see Exercises 5). See Hatcher, page 67 for details.

Higher homotopy groups

There are analogues of $\pi_1(X, x_0)$, with higher dimensional spheres in place of S^1. These higher homotopy groups are surprisingly difficult to calculate, and there remain many open questions. Even the higher homotopy groups of spheres, $\pi_j(S^k)$ for $j > k > 1$, have not yet been determined for all j and k.

Chapter 6

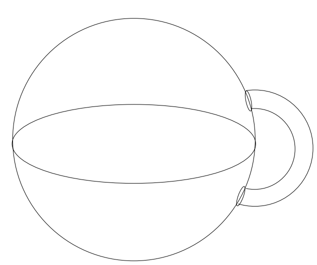
Surfaces

A surface is a 2-dimensional manifold – that is, a Hausdorff topological space in which each point has a neighbourhood homeomorphic to an open set in $\mathbb{R}^2$. We will be interested in compact surfaces. Note that the Möbius strip is only a surface, according to this definition, if we exclude its boundary. A boundary point has no neighbourhood homeomorphic to an open set in $\mathbb{R}^2$. If we want to consider the Möbius strip with its boundary, we need the definition of "manifold with boundary", which we will not go into here.

Surfaces are divided into two natural classes: orientable and non-orientable. In the next section we skirt the notion of orientability and give an *ad hoc* definition which is sufficient for our purposes. The main theorem we will prove in this chapter is

Theorem 6.0.17. *Every compact orientable surface is homeomorphic to a sphere with g handles glued on, for some $g \in \mathbb{N}$.*

The sphere with one handle is the torus T; gluing on g handles gives the genus surface X_g described in Example 2.2.



6.1 Orientation and orientability

In Chapter 5 we defined orientability for surfaces embedded in $\mathbb{R}^3$, as the property of two-sidedness: it is possible to choose a continuous unit normal vector field on the whole surface. The two sides of the surface are distinguished by whether the normal vector points into or away from them. The 2-sphere, and all of the surfaces shown in Figure 3 on page 12, are all orientable. Indeed, any surface which separates $\mathbb{R}^3$ into two connected components, as do all of these, is orientable; choosing one of the two components determines a continuous normal vector field: the vectors pointing into the chosen component.

An *orientation* of a surface is the choice of one of these two vector fields. Evidently every orientable surface $S \subset \mathbb{R}^3$ has two orientations.

The Möbius strip is not orientable.

An intrinsic definition, not making any reference to an embedding in $\mathbb{R}^3$, or in any other space, requires more thought. The two most widely used definitions, by means of an orientation of the tangent space at each point and by means of a fundamental class in the top-dimensional homology group, are both out of reach. So we use the following *ad hoc* definition:

Definition 6.1.1. *A surface S is orientable if it does not contain an open set homeomorphic to a Möbius strip.*

This is rather a cheat, but will have to do for now.

6.2 Triangulations

In order to progress towards the classification of surfaces, we need to be able to triangulate each surface: to subdivide it into triangles, in such a way that any two triangles meet along a common edge, or at a comon vertex, or not at all. We formalise this by the notion of *simplicial complex*.

Definition 6.2.1.

1. An k-simplex in $\mathbb{R}^n$ is the convex hull of $k+1$ points $x_0, \ldots, x_k \in \mathbb{R}^n$ in general position. We call it the *simplex spanned by the points* $x_0, \ldots, x_k$, and denote it by $(x_0, \ldots, x_k)$. These points are called its *vertices*. Thus a 0-simplex is a point, a 1-simplex is a line segment, and a 2-simplex is a triangle (the convex hull of three non-collinear points).

2. Simplices have faces in a natural way: if A and B are simplices, and the vertices of A form a subset of the vertices of B, we say that A is a *face* of B and write $A < B$.

3. A finite collection of simplices in some $\mathbb{R}^n$ is a *simplicial complex* if whenever a simplex lies in this collection, so do all of its faces, and whenever two simplices of the collection meet, their intersection is a face of each.

4. If K is a simplicial complex, the union of all of its simplices is denoted by $|K|$ and called the *underlying polyhedron* of the complex.

5. A space X is *triangulable* if it is homeomorphic to the underlying polyhedron $|K|$ of some simplicial complex.

Theorem 6.2.2. *Every compact surface is triangulable.*

We will not prove this. It is easily seen to be true for the oriented surfaces in our list (the sphere with g handles), but our proof that every compact orientable surface S is homeomorphic to a member of this list requires a triangulation of S as a starting point of the argument.

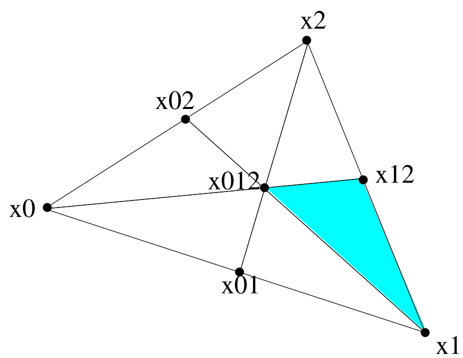
A *triangulation* of a surface S is a simplicial complex K together with a homeomorphism $|K| \to S$. The same surface has many different triangulations. In particular, given any triangulation one can subdivide it to obtain a new triangulation.

Definition 6.2.3. 1. The *barycentre* of a simplex $(v_0, \ldots, v_k) \subset \mathbb{R}^n$ is the point $\frac{1}{k+1} \sum_{i=0}^{k} x_i$. (The term "barycentre" is used because the barycentre of a simplex is the centre of mass of the object obtained by situating a unit mass at each of its vertices.)

2. The *barycentric subdivision* of a k simplex is obtained by taking, as extra vertices, the barycentres of all of its faces, and then taking, as new simplices, all simplices like the one shown shaded in the diagram, where

   - v_0 is a vertex of σ
   - v_1 is the barycentre of a 1-dimensional face F_1 of σ having v_0 as a vertex
   - v_2 is the barycentre of a 2-dimensional face F_2 of σ having F_1 as an edge, etc,

   together with all of their faces.



3. The barycentric subdivision of a simplicial complex K is the new complex formed from K by taking the barycentric subdivision of each of its simplices. We will denote it by K^1. Clearly $|K^1| = |K|$.

In the drawing, xij is the barycentre of the 1-simplex (xi, xj) and $x012$ is the barycentre of the simplex (x_0, x_1, x_2). The shaded triangle is the 2-simplex $(x1, x12, x123)$. If the definition is somewhat indigestible, note that in this course we will use only the barycentric subdivision of simplices of dimension ≤ 2.

Repeated barycentric subdivision is a very useful way of refining a triangulation. For instance, denoting the barycentric subdivision of K^1 by K^2 (it is the "second barycentric subdivision" of K), etc., one can prove that for any simplicial complex K

$$\lim_{m \to \infty} \max\{\text{diameter}(\sigma) : \sigma \text{ is a simplex of } K^m\} = 0.$$

A triangulation of a surface allows us to define what turns out to be its most important topological invariant, its Euler characteristic.

Definition 6.2.4. Let K be a simplicial complex, and for each $k \in \mathbb{N}$ let c_k be the number of k-simplices in K. The *Euler characteristic* of K, $\chi(K)$, is defined by

$$\chi(K) = \sum_k (-1)^k c_k. \tag{6.2.1}$$

Theorem 6.2.5. *χ is a topological invariant: if K and L are simplicial complexes with $|K|$ homeomorphic to $|L|$ then $\chi(K) = \chi(L)$.*

The theorem is proved in the module Algebraic Topology, by showing that

$$\chi(K) = \sum_k (-1)^k \text{rank } H_k(|K|),$$

where, for any topological space $H_k(X)$ is its k'th *homology group*, defined in such a way that its topological invariance is evident.

Because $\chi(|K|)$ is really a property of $|K|$, we can refer to the Euler characteristic of a (triangulable) space, and speak of $\chi(S^2)$, $\chi(T)$, etc.

The Euler characteristic appears throughout topology in a multitude of different guises. Euler's rôle in this was the now familiar observation that any convex polyhedron, $\chi = 2$.

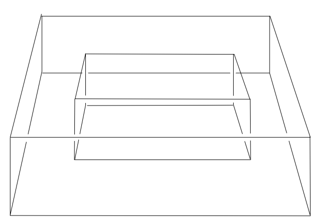
Example 6.2.6.

1. Let K be any triangulation of the circle S^1. Then K consists of an equal number of 0-simplices and 1-simplices, so $\chi(K) = 0$.

2. Let K be any triangulation of the disc $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. Then $\chi(K) = 1$. For this we need to use the fact that $\chi(K)$ depends only on the underlying polyhedron. We can triangulate D, as a single 2-simplex, which gives $\chi(D) = 1$.

3. $\chi(S^2) = 2$, since S^2 is homeomorphic to the underlying polyhedron of a tetrahedron. Check that the alternating sum
$$\#\text{ faces } - \#\text{ edges } + \#\text{ vertices} = 2$$
for each of the regular polyhedra (tetrahedron, cube, octahedron, dodecahedron and icosahedron)- we do not even need to subdivide the faces into triangles for this formula to hold.

4. $\chi(T) = 0$, where T is the torus. To see this we need a triangulation. This diagram shows a surface homeomorphic to the usual curvilinear torus.



You can obtain a triangulation by dividing the plane faces in the diagram into triangles.

5. Let X_g be the genus g compact orientable surface – the sphere with g handles. Then
$$\chi(X_g) = 2 - 2g.$$

We can prove this inductively, using the fact that $\chi(S^2) = 2$ and $\chi(T) = 0$ – the first two cases of our formula. The induction step is achieved by removing a triangle from a triangulation K_g of X_g and from a triangulation K_1 of $T = X_1$, and then joining the two surfaces so obtained, which we will call X'_g and T', along their exposed edges, to obtain X_{g+1}. We have $\chi(X'_g) = \chi(X_g) - 1, \quad \chi(T') = \chi(T) - 1$, since in each case only the number of 2-simplices has changed. Let $c_k(X_{g+1})$ denote the number of k-simplices in the triangulation of X_{g+1} that we obtain. We have

$$\begin{array}{ccccc}
c_2(X_{g+1}) & = & c_2(X'_g) + c_2(T') & = & c_2(X_g) + c_2(T) - 2 \\
c_1(X_{g+1}) & = & c_1(X'_g) + c_1(T') - 3 & = & c_1(X_g) + c_2(T) - 3 \\
c_0(X_{g+1}) & = & c_0(X'_g) + c_0(T') - 3 & = & c_0(X_g) + c_0(T) - 3 \\
\hline
\chi(X_{g+1}) & = & \chi(X'_g) + \chi(T') & = & \chi(X_g) + \chi(T) - 2
\end{array}$$

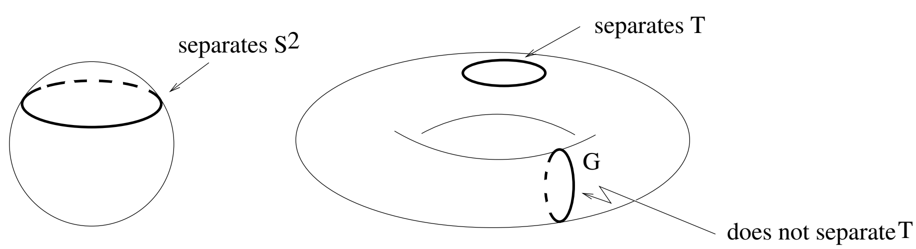
and hence $\chi(X_{g+1}) = 2 - 2(g + 1)$ and the induction is complete.

Thus *the compact oriented surfaces in our list are distinguished by their Euler characteristic.*

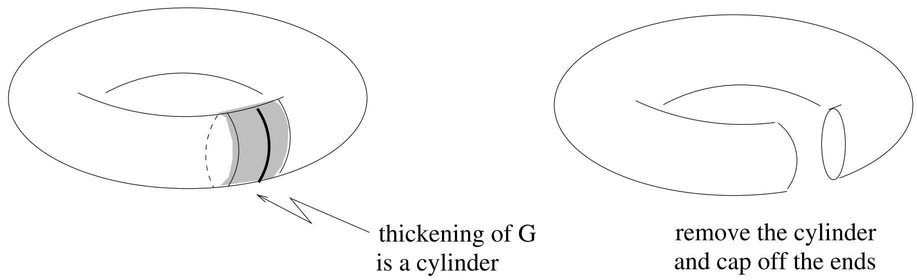
6.3 Sketch proof of the classification theorem

Outline

The proof we sketch is taken from the excellent book of Armstrong, [1], Chapter 7. Suppose that S is a compact oriented surface. The key idea is to look for a simple closed curve G on S which does not separate S into two connected components.



By thickening it we get a cylinder (Lemma 6.3.7); removing the cylinder we get a surface with two holes, S'; capping the holes with two discs, we obtain a new compact connected oriented surface S''.



Lemma 6.3.1. $\chi(S'') = \chi(S) + 2$.

Proof. Step 1 If K and L are simplicial complexes whose intersection is a subcomplex of each, then $K \cup L$ is a simplicial complex and $\chi(K \cup L) = \chi(K) + \chi(L) - \chi(K \cap L)$. To see this, let $c_k(K)$ denote the number of k-simplices in a complex K. Then for each k we have

$$c_k(K \cup L) = c_k(K) + c_k(L) - c_k(K \cap L)$$

as the sum $c_k(K) + c_k(L)$ counts the k-simplices in $K \cap L$ twice. Hence

$$\sum_k (-1)^k c_k(K \cup L) = \sum_k (-1)^k c_k(K) + \sum_k (-1)^k c_k(L) - \sum_k (-1)^k c_k(K \cap L).$$

Step 2 Denote the cylinder by C, and let D_1 and D_2 be the two discs with which we cap the holes in S'. So $S'' = S' \cup D_1 \cup D_2$. We have $S = S' \cup C$ so $\chi(S) = \chi(S') + \chi(C) - \chi(S' \cap C)$. As $S' \cap C$ consists of two disjoint circles, $\chi(S' \cap C) = 0$, so $\chi(S) = \chi(S')$.

Similarly, $S'' = S' \cup D_1 \cup D_2$, so

$$\chi(S'') = \chi(S') + \chi(D_1 \cup D_2) - \chi(S' \cap (D_1 \cup D_2)).$$

Now D_i is a disc, with $\chi = 1$, and $S' \cap D_i$ is a circle, with $\chi = 0$. So this gives $\chi(S'') = \chi(S') + 2 = \chi(S) + 2$. □

If, on the new surface S', there is another curve which does not separate it, repeat the procedure. Each time we do this, the Euler characteristic increases by 2. Now $-\infty < \chi(S) \leq 2$ (the right hand inequality by Lemma 6.3.8 below), so the procedure can be carried out only a finite number of times. Thus, after a certain number of surgeries, say g, there is no simple closed curve which does not separate the surface. We show that when this has occurred, the surface obtained must be a sphere (Proposition 6.3.9).

By reversing the procedure, we construct the original surface S by gluing g handles onto a sphere. This proves that S is in our list.

Slightly more detail

We will need, but will not prove,

Theorem 6.3.2. *(Weak version of Jordan curve theorem) Every simple closed curve separates S^2.*

A connected graph (a 1-dimensional simplicial complex) is a *tree* if it contains no loops.

Lemma 6.3.3. *Let G be a connected graph. Then $\chi(G) \leq 1$ with equality if and only if G is a tree.*

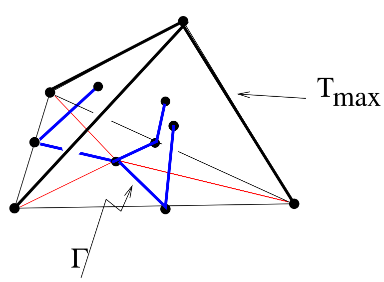
Proof. Suppose that G is a tree. It can be built up beginning with one vertex and no edges. At this stage $\chi = 1$. To add a new edge we have to add a new vertex. So χ is unchanged. Thus if G is a tree, $\chi(G) = 1$. If G has any loop, then it is possible to remove an edge without losing connectivity and without reducing the number of vertices. Thus χ increases by 1. By doing this repeatedly, we eventually get a tree, for which $\chi = 1$. □

Let K be a connected combinatorial surface (that is, a connected simplicial complex such that $|K|$ is a compact surface. A *graph in K* is a graph made up of vertices and edges of K. A *maximal tree in K* is a graph which is a tree, and which cannot be added to without ceasing to be a tree (i.e. without acquiring a loop).

Lemma 6.3.4. *A maximal tree in K contains every vertex of K.*

Proof. Almost obvious. □

Let $T_{\max}$ be a maximal tree in K. We now define another graph, the *dual* of $T_{\max}$, Γ, in K^1. Its vertices are the barycentres of the triangles of K. Two vertices v_i and v_j of Γ are joined by an edge if they lie in triangles of K sharing an edge in K, and this edge is not contained in $T_{\max}$. The edge that joins them is in fact the union of two edges of K^1; this subdivision is indicated in the bottom triangle on the tetrahedron show here.

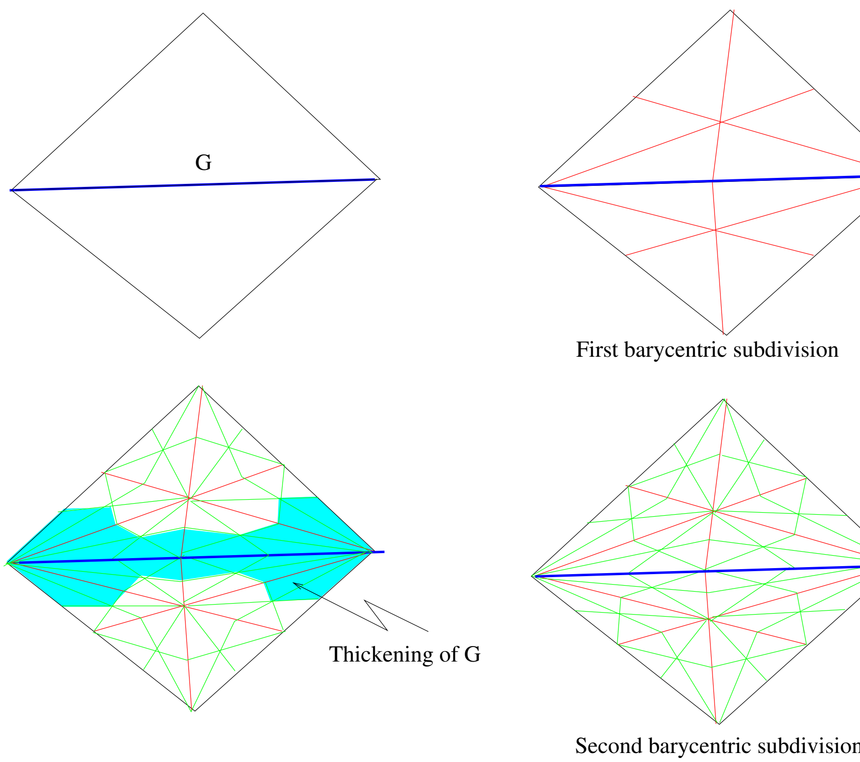


In fact Γ is the union of all simplices in K^1 which do not meet $T_{\max}$.

Lemma 6.3.5. Γ *is connected.*

We prove this by thickening $T_{\max}$ and Γ.

Definition 6.3.6. *Let G be a graph made up of 1-simplices in a combinatorial surface K. The thickening of G is the simplicial complex consisting of all simplices of the second barycentric subdivision K^2 which meet G, together with all of their faces.*

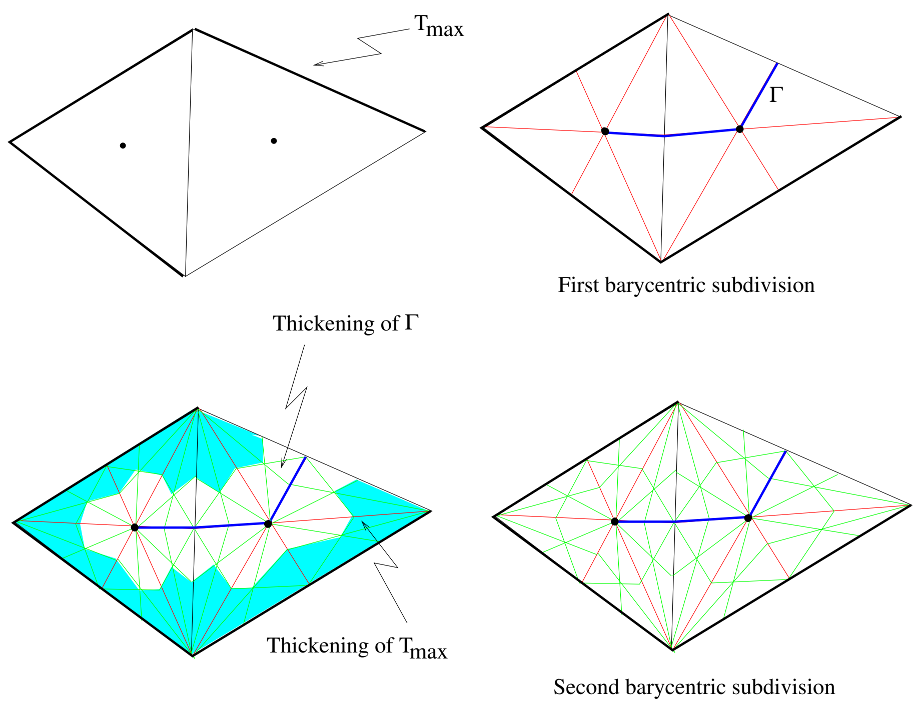


Lemma 6.3.7.
1. *Thickening a tree gives a simplicial complex whose underlying polyhedron is homeomorphic to a disc.*
2. *Thickening a simple closed curve gives a simplicial complex whose underlying polyhedron is homeomorphic either to a cylinder or a Möbius strip.*

Using this lemma we prove that Γ is connected:

Proof. *of Lemma 6.3.5* Thicken $T_{\max}$ and Γ, to complexes $N(T_{\max}$ and $N(\Gamma)$ respectively. Then

1. $|N(T_{\max})| \cup |N_(\Gamma)| = |K|$.
2. $|N(T_{\max})|$ and $|N(\Gamma)|$ intersect in the boundary circle of $|N(T)|$.
3. $N(\Gamma)$ is connected if and only if Γ is connected.



Now any two points x and y of $N(\Gamma)$ can be joined by a path in $|K|$. Let p and q be the first and last points where this path meets the boundary of $|N(T)|$. The part of the path between p and q can be replaced by a path along the boundary of $N(T_{\max})$. This also lies in $N(\Gamma)$. So $N(\Gamma)$ is connected and therefore so is Γ. □

Lemma 6.3.8. $\chi(K) = \chi(T_{\max}) + \chi(\Gamma)$, *and* $\chi(K) \leq 2$.

Proof. All vertices of K lie in $T_{\max}$, Γ has two edges and a vertex for every edge of K not in $T_{\max}$, and there is one vertex in Γ for each triangle of K. The second statement then follows by Lemma 6.3.3. □

Proposition 6.3.9. *Let K be a simplicial complex with $|K|$ a compact oriented surface. The following are equivalent.*

1. *Every simple closed curve in $|K|$, made up of edges of the first barycentric subdivision K^1, separates $|K|$ into two (or more) connected components.*

2. $\chi(K) = 2$

3. *$|K|$ is homeomorphic to S^2.*

Proof. Suppose that (1) holds. Choose a maximal tree $T_{\max}$ in K and let Γ be its dual. I claim that Γ is also a tree. For if not, Γ contains a loop, and by (1), this loop separates $|K|$. But each component of the complement of the loop must contain a vertex of $T_{\max}$, contradicting the fact that $T_{\max}$ is connected and disjoint from Γ. So Γ must be a tree. Hence

$$\chi(K) = \chi(T_{\max}) + \chi(\Gamma) = 1 + 1$$

where the first equality is Lemma 6.3.8 and the second is from Lemma 6.3.3.

Now suppose that (2) holds. By Lemma 6.3.8 and Lemma 6.3.3, Γ must be a tree. It follows that $|N(\Gamma)|$, like $|N(T_{\max})|$, is homeomorphic to a disc. Hence $|K|$ is the union of two discs, glued along their bounding circles, and thus homeomorphic to S^2.

Finally, any simple closed curve separates S^2, (this is Theorem 6.3.2) so (3) $\implies$ (1). $\square$

6.4 Appendix: X_g as a quotient of a regular $4g$-gon

The surface of genus 1 is the quotient of a square with opposite edges identified. Instead of labelling one pair of edges with a double arrow, as previously, we label the horizontal edges "a" and the vertical edges "b". In the gluing, a is glued to a and b to b.

If we remove a corner of the square (shown shaded in the picture below), we get a torus with a hole.

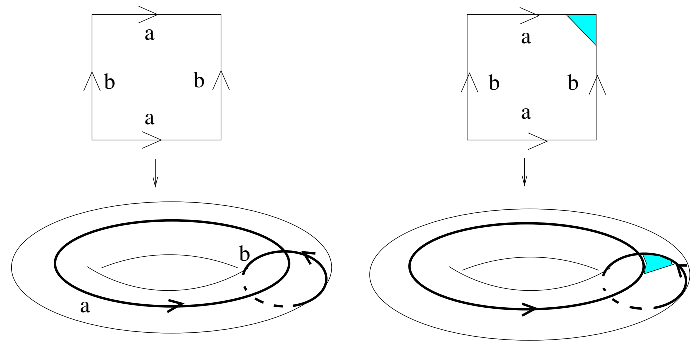


Figure 6.1

Gluing two of these along the edges of the holes gives a genus 2 surface.

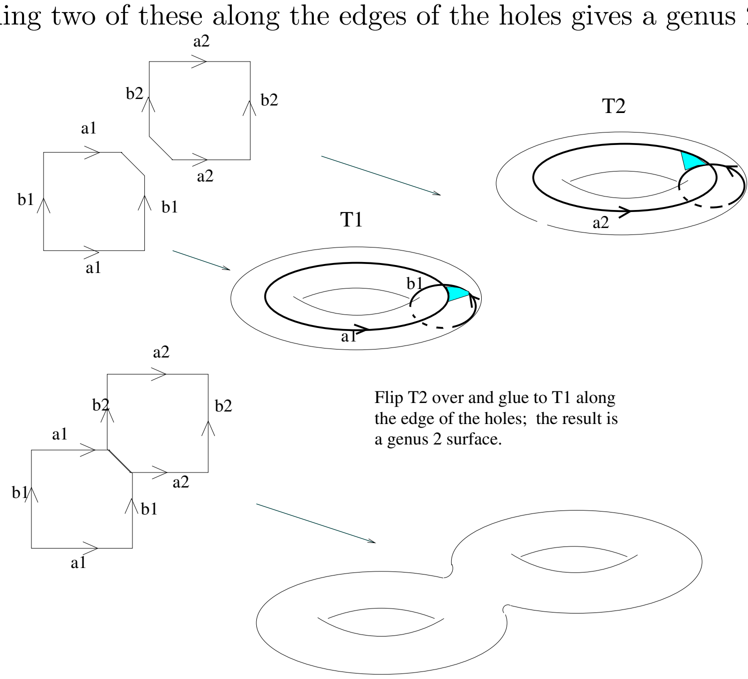


Figure 6.2

The procedure can be repeated to give a genus g surface. The resulting $4g$-gon has $2g$ edges marked a_i and $2g$ edges labelled b_i, for $i = 1, \ldots g$.

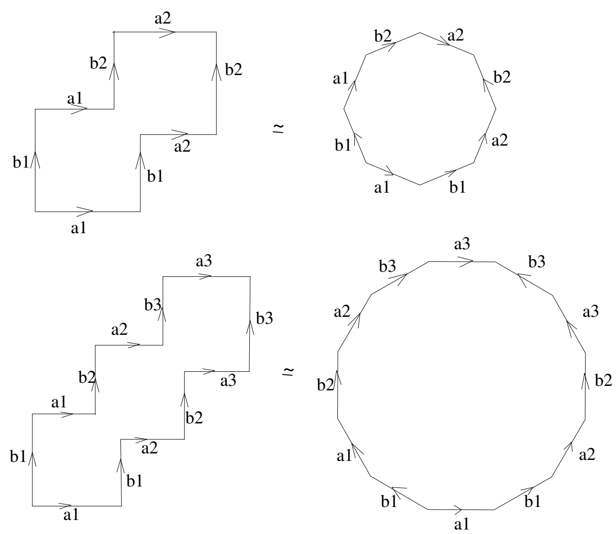


Figure 6.3

We can replace the $4g$-gon obtained by this procedure, by a regular $4g$-gon. The two are clearly homeomorphic, and by imposing the corresponding equivalence relation on the regular $4g$-gon, we obtain as quotient the same genus g surface. The gluing data is easy to understand: starting at any vertex, go anticlockwise labelling the edges successively $a_1, b_1, a_2, b_2, \ldots, a_g, b_g$, and drawing anti-clockwise arrows on them. Now go back to the starting point and travel clockwise, labelling the edges successively $b_1, a_1, b_2, a_2, \ldots, b_g, a_g$, and drawing clockwise arrows on them. Then edges with the same label are glued to one another, in the sense indicated by the arrows.

6.5 Gaps

These notes have not done justice to all of the topics they cover; many are just touched on. Several gaps stand out. The first is the topic of *orientation of manifolds*, and orientability. This is dealt with in Chapters 5 and 6 in a cursory way which sidesteps the difficulties. Orientation is a deep and subtle notion which is only open to students of topology. To define it properly requires either the theory of differentiable manifolds, or homology theory, both beyond the remit of the course. The approach via differential topology is well explained in e.g. the books of Guillemin and Pollack, [2], and of Milnor, [4], while the approach using homology theory is explained in e.g the book [3] of Hatcher.

Details of the proof of the classification of surfaces can be found in the book of Armstrong, [1], though this introduction to topology, like every other that I know of, omits the proof that every surface has a triangulation.

Chapter 1 of Hatcher's book [3] gives many more examples of covering maps, and of other constructions, than we had time for here.

Bibliography

[1] M.A. Armstrong, *Basic Topology*, Undergraduate Texts in Mathematics, Springer, 1983

[2] V. Guillemin and A. Pollack, *Differential Topology*, American Mathematical Society, 2010 (originally published in 1974).

[3] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2001, *available online via* http://www.math.cornell.edu/∼hatcher/#ATI

[4] J. Milnor, *Topology from the diferentiable viewpoint*, Princeton University Press, and freely available online as a pdf

[5] J. Stillwell, *Classical Topology and Combinatorial Group Theory*, Graduate Texts in Mathematics, Springer-Verlag, 1980