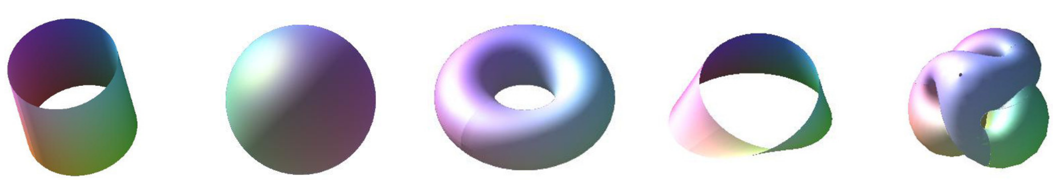
MAS61015 ALGEBRAIC TOPOLOGY

Contents

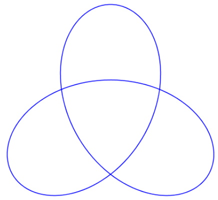
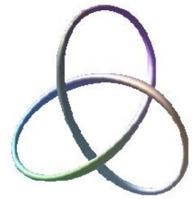
1. Introduction

This course is about the topological structure of spaces. We start by discussing some examples. Here are some that you should remember from the Knots and Surfaces course: the cylinder, the sphere, the torus, the Möbius strip and the real projective plane.



Interactive demo

The last picture actually shows a set called Boy's surface, which crosses over itself. To understand how this relates to the real projective plane, consider (as an analogy) the following pictures:

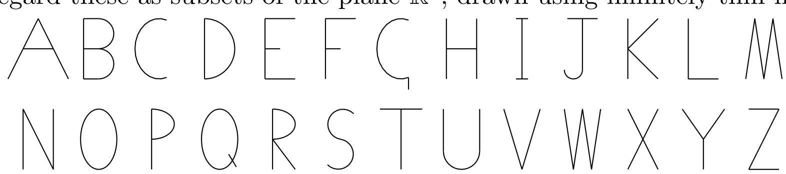



Interactive demo

The right hand picture shows the trefoil knot in $\mathbb{R}^3$, which does not intersect itself. The left hand picture is an attempt to represent the knot in two dimensions, but that is not possible without introducing self-intersections. Similarly, the real projective plane lives naturally in four dimensional space, where it has no self-intersections, and Boy's surface is an imperfect three-dimensional representation.

In this course, we will consider many examples of spaces of dimension three or less, because that makes it easier to draw pictures. However, you should be familiar with the idea that any problem with n variables can lead you to consider n-dimensional linear algebra, and thus sometimes to think about the geometry of lines, planes and so on in $\mathbb{R}^n$ and how they intersect. If we have nonlinear equations in n variables, we may also need to consider nonlinear subspaces of $\mathbb{R}^n$ and their geometry and topology, and this can often have meaning in the real world even when $n > 4$. In fact, in the recently developed field of Topological Data Analysis, it is common to have very large values of n. For example, there is active work on applications of TDA to neuroscience, in which n is the number of neurons that one is modelling or monitoring. For large values of n we cannot hope to draw meaningful pictures or rely on intuition. Instead, we need methods that convert problems from topology into more tractable questions in algebra. This is the main goal of Algebraic Topology.

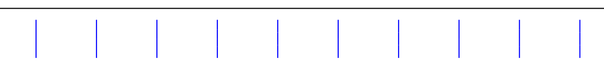
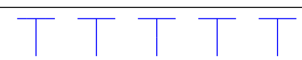
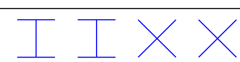
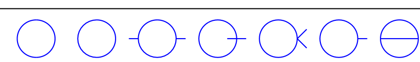
Example 1.1. One family of examples that we can use to illustrate this goal consists of the letters of the alphabet. We will regard these as subsets of the plane $\mathbb{R}^2$, drawn using infinitely thin lines as follows:



Many of these (such as C, L and W) can be straightened out to just give a line segment. With a bit more bending and stretching we can make the E, F, J and Y look the same as the T.

Interactive demo

We can redraw the table as follows:

C G L M N S U V W Z	
E F J T Y	
H I K X	
O D A Q R P B	

Interactive demo

On the left, we have drawn all the letters again in red, grouping together letters that have similar properties. On the right, we have drawn a straightened-out form for each letter. It seems that all the letters in the first group are topologically equivalent, as are all the letters in the second group. However, the letters in the first group do not seem to be equivalent to those in the second group. How can we make this precise? One approach uses the following definition:

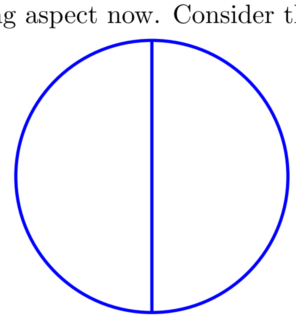
Definition 1.2. A *cut point* of X is a point $x \in X$ such that $X \setminus \{x\}$ is disconnected. We define $d(X)$ to be the number of points that are not cut points.

For the first three groups of letters, almost all the points are cut points; the only exceptions are the tips of the various branches. Thus, for the first group of letters we have $d = 2$, for the second group we have $d = 3$, and for the third group we have $d = 4$. However, for the letters D, O and B there are no cut points, so $d = \infty$. For the remaining letters the situation is more complicated, but there are still infinitely many points where you can cut without disconnecting the space, so again $d = \infty$.

What can we take away from this discussion?

(a) We have used some ideas about connected and disconnected spaces that seem very reasonable, but we still have not made precise definitions or proved any theorems, which will be essential if we want to make sure that everything will work in higher dimensions.
(b) We have implicitly assumed that topologically equivalent spaces have the same value of $d(X)$. This is true if we use the most obvious version of topological equivalence, which is called *homeomorphism*, but we need to give an actual proof of that. We will also spend a lot of time discussing a different notion called *homotopy equivalence*. It will turn out that homotopy equivalent spaces need not have the same value of $d(X)$, which emphasises why we need to be careful with definitions and proofs.
(c) We have defined a number $d(X)$, such that $d(X) = d(Y)$ whenever X and Y are homeomorphic. In other words, $d(X)$ is a *numerical homeomorphism invariant* of X. The letters B and C have different values of d, so they are not homeomorphic. (This is visually obvious, but now we have a method of proof that we can hope to apply in cases that are not visually obvious.)
(d) However, this logic does not work backwards. The letters H and X both have $d = 4$, but we cannot conclude that these letters are homeomorphic. In fact, we can prove that they are not homeomorphic: if we remove the central point from X then the remaining space breaks into 4 connected pieces, but there is no way to break H into more than 3 pieces by removing a single point.
(e) Moreover, this technique becomes very ineffective if we consider more complicated spaces. For example, we can remove any finite set of points from the sphere and it will still remain connected, and the torus has the same property, so this approach has no chance of detecting the difference between the sphere and the torus. We will need other techniques that need much more work to set up.

Now consider again the letters B, O and C. Although we can distinguish between them by methods similar to those described above, this is in some sense missing the most obvious point: B has two holes, O has one hole, and C has none. Unfortunately, it will take a great deal of work to give a proper mathematical formulation of this point, and we will not achieve that until a long way into the course. However, we will explain one interesting aspect now. Consider the following pictures:



Interactive demo

The right hand picture shows a space X in three dimensions, consisting of three lines of longitude joining the north and south poles of the sphere S^2. The left hand picture is the same space, flattened out into the plane. Looking at the right hand picture, it seems natural to say that there are three holes, arranged symmetrically around the z-axis. Looking at the left hand picture, it seems natural to say that there are only two holes. Which is correct? The key point here is that we should not think of the holes as just giving a number; instead, there is an abelian group of holes, called $H_1(X)$, the first homology group of X. There really are three symmetrically arranged holes, which we can call a, b and c, but they satisfy $a+b+c=0$. We can therefore use the relation $c=-a-b$ to express every element of $H_1(X)$ in the form $na+mb$ for some $(n,m)\in\mathbb{Z}^2$, so the group $H_1(X)$ is isomorphic to $\mathbb{Z}^2$. This is the sense in which there are really only two basic holes. The notion of addition and subtraction used here is not obvious, but it will emerge naturally when we give the formal definitions.

This should hopefully motivate the idea that our invariants should not just be numbers, but should instead be algebraic structures such as groups or rings. We will eventually define abelian group denoted by $H_n(X)$ for all $n\geq 0$, which are again called homology groups. For example, the torus T has $H_0(T)\simeq H_2(T)\simeq\mathbb{Z}$ and $H_1(T)\simeq\mathbb{Z}^2$ and $H_n(T)=0$ for $n>2$. We will also write $H_*(X)$ for the sequence of all homology groups of X, so $H_*(T)=(\mathbb{Z},\mathbb{Z}^2,\mathbb{Z},0,0,\dots)$ for example. We can use these groups to prove many interesting topological facts that are completely inaccessible by other methods.

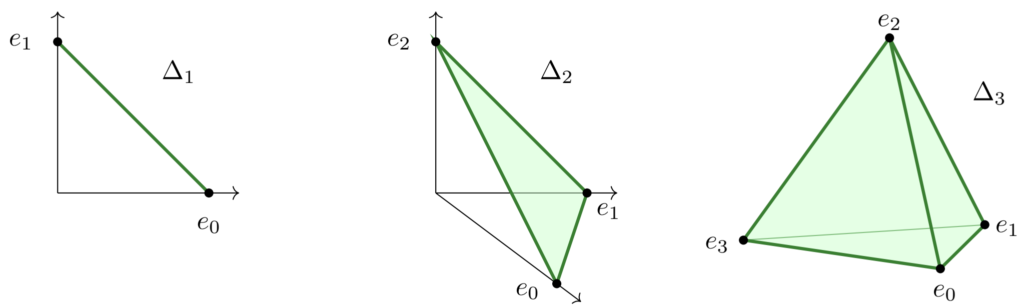
Our approach to homology groups will involve simplices and simplicial complexes.

Definition 1.3. The *standard n-simplex* is the space

$$\Delta_n=\{(x_0,\dots,x_n)\in\mathbb{R}^{n+1}\mid x_i\geq 0 \text{ for all } i \text{ and } \sum_i x_i=1\}.$$

The *vertices* of Δ_n are just the standard basis vectors $e_0,\dots,e_n$, so $e_0=(1,0,\dots,0)$ and $e_1=(0,1,0,\dots,0)$ and $e_n=(0,\dots,0,1)$ and so on.

The standard 0-simplex Δ_0 is just a single point, and Δ_1 is a line segment, and Δ_2 is a triangle, and Δ_3 is a tetrahedron. We can draw them as follows:



Interactive demo

A key technique will be to study spaces by dividing them up into simplices. Consider the following pictures:

Interactive demo

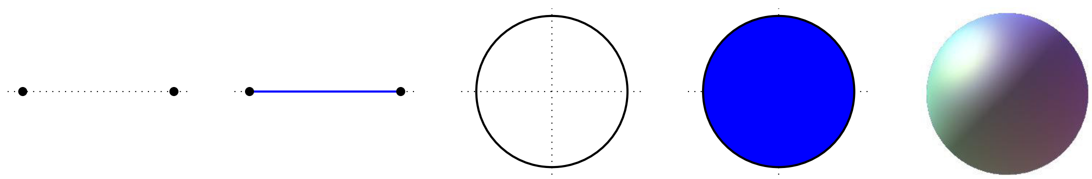
The middle picture shows the undivided sphere S^2, which is what we really want to study. The second picture shows S^2 divided into curved triangles in an octahedral pattern, and the fourth picture shows S^2 divided into curved triangles in an icosahedral pattern. The first picture is a genuine octahedron with flat faces, and the last picture is a genuine icosahedron with flat faces. These are examples of *simplicial complexes*. The sphere is homeomorphic to both the octahedron and the icosahedron. We will eventually prove a theorem that will allow us to use the combinatorial structure of either of these complexes to compute the homology of S^2 (although in this example, other methods of computing homology are easier).

Example 1.4. Here are some further examples of higher-dimensional spaces that we will consider later.

(a) The most basic example is $\mathbb{R}^n$. Although it might seem obvious, it is already difficult to prove that $\mathbb{R}^n$ and $\mathbb{R}^m$ are not homeomorphic when $n \neq m$, but we will achieve that by the end of the course.

(b) Inside $\mathbb{R}^{n+1}$ we have the unit ball and the unit sphere:

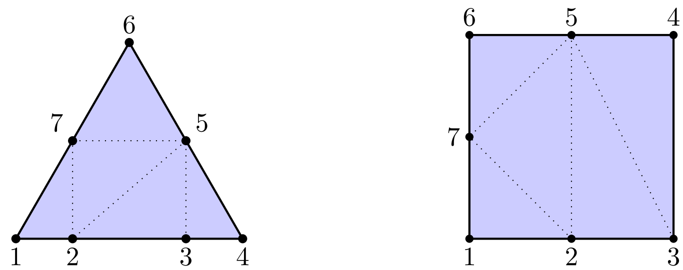
$$B^{n+1} = \{x \in \mathbb{R}^{n+1} \mid \|x\| \leq 1\}$$
$$S^n = \{x \in \mathbb{R}^{n+1} \mid \|x\| = 1\}.$$



S^0 $\quad$ B^1 $\quad$ S^1 $\quad$ B^2 $\quad$ S^2

(The last picture is supposed to depict a hollow shell; the space inside does not count as part of S^2.)

(c) We have already seen the simplex Δ_n. This is actually homeomorphic to the ball B^n, or to the cube $[0,1]^n$. Later we will see a nice general theorem that makes it easy to prove that something is homeomorphic to $[0,1]^n$. Alternatively, we can draw a picture for $n = 2$: there is a homeomorphism sending the numbered points to the numbered points and the dotted edges to the dotted edges.

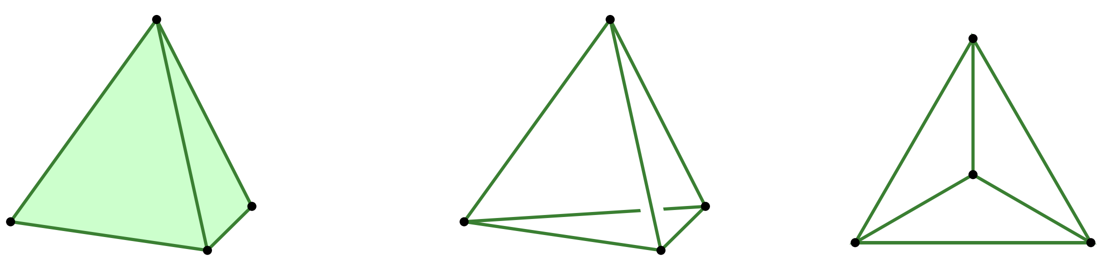


Interactive demo

We will also be interested in skeleta of simplices:

$$\mathrm{skel}^k(\Delta_n) = \{x \in \Delta_n \mid \text{ at least } n-k \text{ coordinates are zero } \}$$

For example, the 1-skeleton of Δ_n consists of all the vertices and edges of Δ_n, but nothing else.



$\Delta^3 =$ solid tetrahedron $\quad$ $\mathrm{skel}^1\Delta^3 =$ just the edges $\quad$ flattened version

Interactive demo

(d) The n-dimensional torus is $S^1 \times \cdots \times S^1 = (S^1)^n$. The case 0-dimensional torus is just a point, the 1-dimensional torus is a circle, and the 2-dimensional torus is what we normally just call the torus.

(e) Let $M_n(\mathbb{R})$ be the space of $n \times n$ matrices over the real numbers. This can be identified with the space $\mathbb{R}^{n^2}$ which we have considered already. Inside $M_n(\mathbb{R})$, we can consider the subspace $GL_n(\mathbb{R})$ of invertible matrices, and the subspace O_n of orthogonal matrices (satisfying $A^T A = I$), and the subspace SO_n of special orthogonal matrices (satisfying $\det(A) = 1$ as well as $A^T A = I$). All of these spaces have interesting topology. For example, $GL_2(\mathbb{R})$ is homeomorphic to $\mathbb{R}^3 \times S^1 \times \{1, -1\}$ (as shown in Example 4.11), but the answers are more complicated for $n > 2$. There are also similar examples involving complex matrices.

(e) The n-dimensional real projective space $\mathbb{R}P^n$ is obtained from the sphere S^n by identifying x with $-x$ for all x. This is a kind of quotient construction, for which we will need to recall various ideas about equivalence relations and study how they interact with topology. (This is also required for a rigorous treatment of the kind of gluing constructions that you will have seen in the Knots and Surfaces course.) It works out that $\mathbb{R}P^1$ is homeomorphic to S^1 and $\mathbb{R}P^3$ is homeomorphic to the matrix group SO_3. There is also a different way to describe $\mathbb{R}P^n$ as a space of matrices: it turns out that $\mathbb{R}P^n$ is homeomorphic to the space

$$P_n = \{A \in M_{n+1}(\mathbb{R}) \mid A^2 = A^T = A, \ \mathrm{trace}(A) = 1\}.$$

(f) We can also consider the complex projective space $\mathbb{C}P^n$, for this, we recall that S^{2n+1} is the unit sphere in $\mathbb{R}^{2n+2}$, but $\mathbb{R}^{2n+2}$ can be identified with $\mathbb{C}^{n+1}$, so the points of S^{2n+1} can be regarded as complex vectors. This lets us form a quotient space in which x is identified with zx whenever z is a complex number with $|z| = 1$. This quotient space is $\mathbb{C}P^n$. It turns out that $\mathbb{C}P^1$ is just S^2, but $\mathbb{C}P^2$ is already quite interesting.

In order to understand all these examples, we need a general theory of topology. The framework of metric spaces is adequate for most, but not all purposes. We will therefore spend some time on the more general framework of topological spaces. This is a large and important topic in its own right, but we will try to cover the minimum that we need without too many distractions.

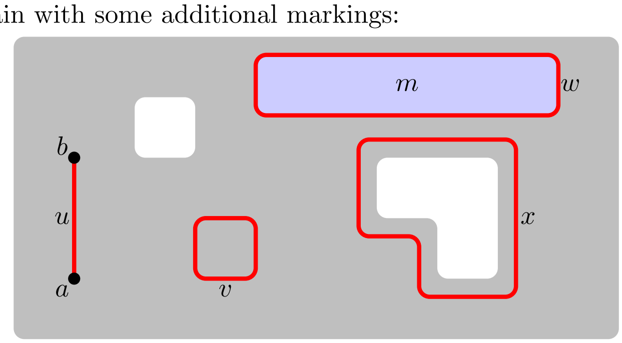
2. The idea of homology

Video

Consider the following space X:

How can we express mathematically the fact that it has two holes? For this we need the concepts of *cycles* and *boundaries*. In this section we will give an imprecise and informal discussion of these ideas. After that we will need to do some foundational work before we can get to a rigorous treatment. In outline: all boundaries are cycles, and many cycles are boundaries, but not all. Cycles that are not boundaries reveal the existence of holes.

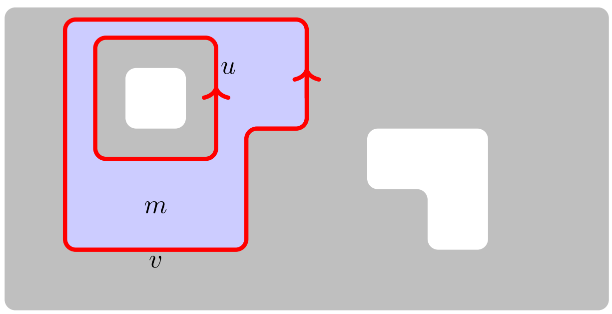
Here is the space X again with some additional markings:



- The boundary of the path u consists of the points a and b.
- The paths v, w and x are closed, so they have no boundary. Objects with no boundary are called *cycles*. Thus, v, w and x are cycles.
- If we look at a small piece of v, then it looks the same as a small piece of $\mathbb{R}^1$, so we regard v as being intrinsically 1-dimensional. Cycles of dimension k are called k-cycles, so v, w and x are 1-cycles.
- The boundary of the region m is the path w, so we can say that w is a boundary. Here m is 2-dimensional and w is 1-dimensional so we say that w is a 1-boundary. Similarly, v is the boundary of an evident region that we have not named, so v is another 1-boundary.
- On the other hand, we cannot fill in the interior of x without using points that are not part of our ambient space X. Thus, with respect to the space X, the cycle x is not a boundary. This reflects the presence of the right-hand hole.
- We regard points as 0-dimensional objects with no boundary, so a and b are 0-cycles.

Video

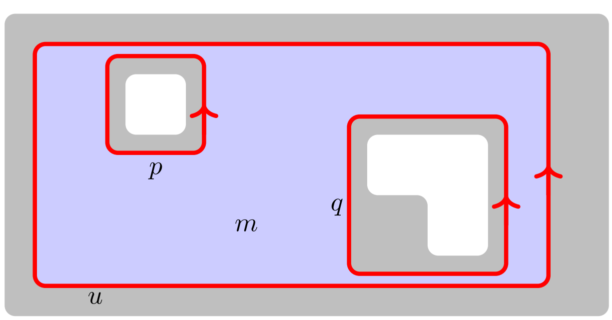
Now consider the following picture:



Both u and v are cycles that are not boundaries, revealing the existence of a hole. However, u and v surround the same hole, so they should not be regarded as interestingly different. This corresponds to the fact that although u and v are not individually boundaries, the difference between them is the boundary of m. In general, cycles should be considered equivalent if the difference between them is a boundary. This suggests that we should set up our detailed definitions so we have an abelian group $Z_1(X)$ of cycles and a subgroup $B_1(X)$ of boundaries, and we should consider the quotient group $Z_1(X)/B_1(X)$. This group will be called $H_1(X)$, and the elements will be called (1-dimensional) homology classes. Each element of $H_1(X)$ is therefore a coset $u + B_1(X)$, where u is a 1-cycle. We also use the notation $[u]$ for $u + B_1(X)$.

You might object that the boundary of m is $v + u$ rather than $v - u$, and that this creates a problem for our interpretation in terms of a quotient group. In fact, if we develop the details using only the ingredients discussed so far, then we end up with a group in which all elements have order 2, so $v + u$ and $v - u$ are the same. However, there is a more refined version in which every path u has a direction, and $-u$ is interpreted as the same path in the opposite direction, and there are similar considerations for objects of dimension greater than one. With appropriate conventions of this type, and with u and v oriented anticlockwise as indicated by the arrows, it works out that the boundary of m is $v - u$, so $[v] = [u]$ in $H_1(X)$.

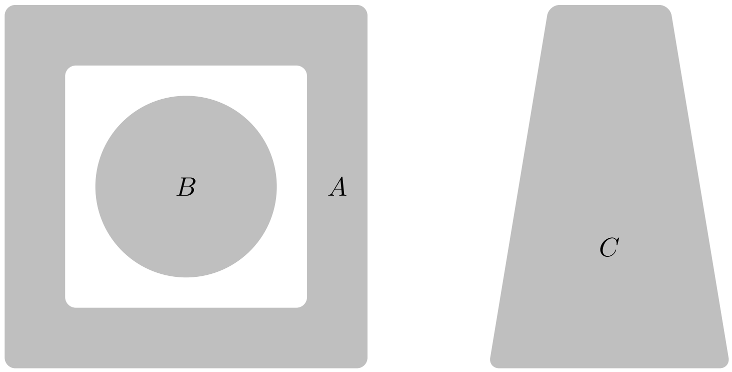
As another example of how this works out, consider the following picture:



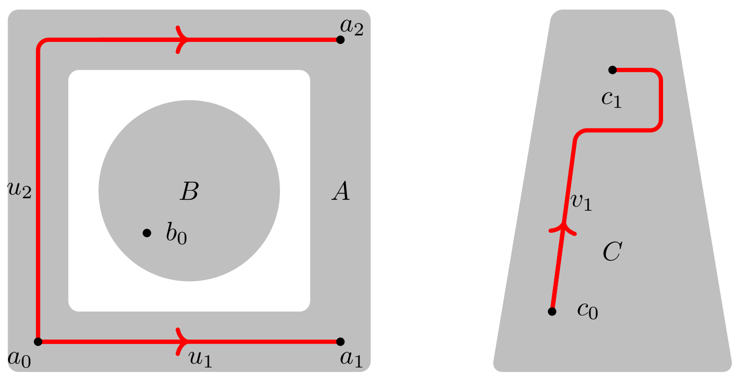
The boundary of m is $u - p - q$, so $[u] = [p] + [q]$ in $H_1(X)$. Similarly, it will work out that if v is any loop that winds i times around the left-hand hole and j times around the right-hand hole then $[u] = i[p] + j[q]$ in $H_1(X)$. Using this we find that $H_1(X)$ is isomorphic to $\mathbb{Z}^2$, with one factor of $\mathbb{Z}$ for each hole.

Video

We can also define a group H_0 along the same lines as H_1. We will illustrate this with reference to the following space Y, which is the disjoint union of three subspaces A, B and C.



The group $H_0(Y)$ is defined as $Z_0(Y)/B_0(Y)$. It will work out that $H_0(Y)$ is isomorphic to $\mathbb{Z}^3$, with one copy of $\mathbb{Z}$ for each of the components A, B and C. To explain this in more detail, we can annotate the picture as follows:



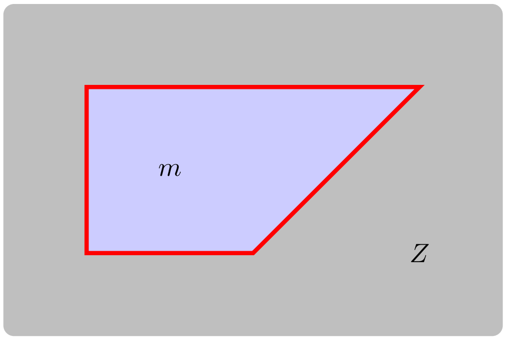
The points a_0, a_1 and a_2 all count as 0-cycles, so they give homology classes $[a_0], [a_1], [a_2] \in H_0(Y)$. However, these are all the same. Indeed, the boundary of the path u_1 is $a_1 - a_0$, so $a_1 - a_0 \in B_0(Y)$, so the cosets $a_1 + B_0(Y)$ and $a_0 + B_0(Y)$ are the same, or in other words $[a_1] = [a_0]$ in $H_0(Y)$. We can use the paths u_1 and v_1 in the same way to see that $[a_2] = [a_0]$ and $[c_1] = [c_0]$. More generally, any point in A has the same homology class as a_0, any point in B has the same homology class as b_0, and any point in C has the same homology class as c_0. Thus, if x is a 0-cycle consisting of i points in A, j points in B and k points in C then $[x] = i[a_0] + j[b_0] + k[c_0]$ in $H_0(Y)$. When we have a more precise set of definitions we will be able to show that the elements $[a_0]$, $[b_0]$ and $[c_0]$ actually give a basis for $H_0(Y)$ over $\mathbb{Z}$, so $H_0(Y) \simeq \mathbb{Z}^3$ as claimed.

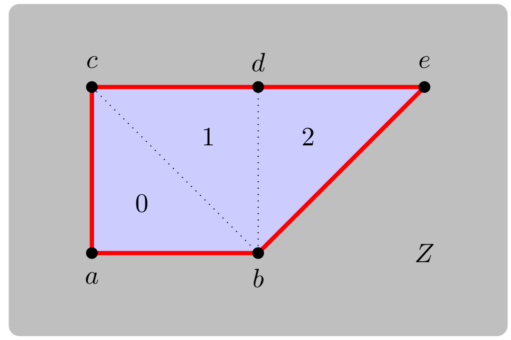
So far we have treated paths in X as subsets of X, but this turns out to be technically awkward, especially if we need to deal with paths that cross over themselves or are fractal or have other unusual behaviour. In our formal definitions, we will instead define paths in X to be continuous maps from the unit interval $[0, 1]$ (or the 1-simplex Δ_1, which is essentially the same) to X.

Similarly, our 2-dimensional objects will not just be subsets of X; instead, they will be continuous maps from the space

$$\Delta_2 = \{(x_0, x_1, x_2) \in \mathbb{R}^3 \mid x_0, x_1, x_2 \geq 0, x_0 + x_1 + x_2 = 1\}$$

to X. For example, consider the pictures below.





In our informal discussion, we might have considered the region $m \subseteq Z$ shown in the left-hand picture. In our formal treatment, we will instead consider the expression $m_0 + m_1 + m_2$, where the maps $m_i : \Delta_2 \to Z$ are defined by

$$\begin{aligned} m_0(x_0, x_1, x_2) &= x_0 a + x_1 b + x_2 c \\ m_1(x_0, x_1, x_2) &= x_0 b + x_1 d + x_2 c \\ m_2(x_0, x_1, x_2) &= x_0 b + x_1 e + x_2 d. \end{aligned}$$

The image $m_i(\Delta_2)$ is the region marked i in the right hand diagram.

3. Topological spaces

We will now change direction, and spend some time building the required theory of topological spaces, which will serve as a foundation for the rigorous definition of homology groups.

Topological spaces are a generalisation of metric spaces. Knowledge of metric spaces is mostly assumed as a prerequisite, but will be reviewed briefly here. References marked [MS] refer to Dr Roxanas's notes for MAS331 (Metric Spaces), but equivalent results can be found in many other sources.

Video (Definition 3.1 to Example 3.3)

Definition 3.1 (MS, page 8)**.** Let X be a set. A *metric* on X is a function $d : X \times X \to \mathbb{R}$ with properties as follows:

(a) For all $x, y \in X$ we have $d(x, y) \geq 0$, and $d(x, y) = 0$ iff $x = y$.
(b) For all $x, y \in X$ we have $d(x, y) = d(y, x)$.
(c) For all $x, y, z \in X$ we have $d(y, z) \leq d(x, y) + d(y, z)$ (the Triangle Inequality).

A *metric space* is a set equipped with a metric.

Example 3.2 (MS, pages 9–10)**.** We can define three different metrics on $\mathbb{R}^n$, as follows:

$$\begin{aligned} d_1(x, y) &= \sum_{i=1}^{n} |x_i - y_i| \\ d_2(x, y) &= \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \\ d_\infty(x, y) &= \max(|x_1 - y_1|, \ldots, |x_n - y_n|). \end{aligned}$$

Later we will explain a sense in which these are essentially the same.

Example 3.3. Consider the set $M_n(\mathbb{R})$ of real $n \times n$ matrices (or any of the various subsets of $M_n(\mathbb{R})$ discussed in Example 1.4). We can define a metric as follows:

$$\mathrm{trace}(A) = \sum_{i=1}^{n} A_{ii}$$
$$\|A\| = \sqrt{\mathrm{trace}(A^T A)}$$
$$d(A, B) = \|A - B\|.$$

This is not really a new example. A little matrix algebra shows that $\mathrm{trace}(A^T A) = \sum_{i,j=1}^{n} A_{ij}^2$, so if we identify $M_n(\mathbb{R})$ with $\mathbb{R}^{n^2}$ in the obvious way, then this new metric is just the same as the standard metric d_2 in the previous example. However, this new formula for the metric makes it easier to combine with other constructions in matrix theory, such as the definition $O_n = \{A \mid A^T A = I\}$ of the orthogonal group.

We will not need any examples that are much more exotic than these.

Video (Definition 3.4 to Proposition 3.12)

Definition 3.4 (MS, page 51)**.** Let X and Y be metric spaces, with metrics d_X and d_Y. Let f be a function from X to Y. We say that f is *continuous* if it has the following property:

> For all $x \in X$ and $\epsilon > 0$, there exists $\delta > 0$ such that for all $x' \in X$ with $d_X(x, x') < \delta$, we have $d_Y(f(x), f(x')) < \epsilon$.

Lemma 3.5. *Suppose that f satisfies $d_Y(f(x), f(x')) \leq d_X(x, x')$ for all $x, x' \in X$. Then f is continuous.*

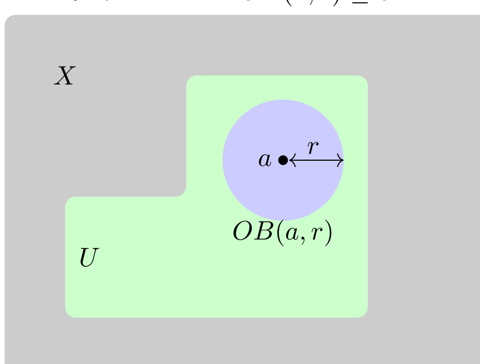
Proof. Suppose we are given $x \in X$ and $\epsilon > 0$. We need to provide a number $\delta > 0$ and check that it has a certain property. We just take $\delta = \epsilon$. Suppose that we have $x' \in X$ with $d(x, x') < \delta = \epsilon$. By our assumptions on f, x and x', we then have $d(f(x), f(x')) \leq d(x, x') < \epsilon$. This is the required property of δ. □

Definition 3.6 (MS, page 20)**.** Let X be a metric space, let a be a point of X, and let r be a positive real number. We put

$$OB(a, r) = \{x \in X \mid d(a, x) < r\},$$

and call this the *open ball of radius r centred at a*.

Definition 3.7 (MS, page 42)**.** Let X be a metric space, and let U be a subset of X. We say that U is *open* if for every point $a \in U$, there exists $r > 0$ such that $OB(a, r) \subseteq U$.



We also say that a subset $F \subseteq X$ is *closed* if the complement $X \setminus F$ is open.

Interactive demo

Remark 3.8. Our definition of closed sets is not the same as in [MS, page 40], but it is equivalent to that definition, as proved in [MS, page 43].

Proposition 3.9 (MS, page 44)**.** *In any metric space X:*

(a) *The empty set and the whole set X are open.*
(b) *The union of any collection of open sets is open.*
(c) *The intersection of any finite collection of open sets is open.*

One of the main reasons for introducing metric spaces is to formalise the notion of a continuous map. The following result shows that we can do that using only the system of open sets, we do not actually need the metric.

Definition 3.10 (MS, page 56)**.** Let $f\colon X \to Y$ be any map of sets, and let B be a subset of Y. We put

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\},$$

and we call this the *preimage* of B under f.

Lemma 3.11. *Let $f\colon X \to Y$ be any map of sets.*

(a) *For any family of subsets $B_i \subseteq Y$, we have $f^{-1}(\bigcap_i B_i) = \bigcap_i f^{-1}(B_i)$ and $f^{-1}(\bigcup_i B_i) = \bigcup_i f^{-1}(B_i)$.*
(b) *For any subset $B \subseteq Y$ we have $f^{-1}(Y \setminus B) = X \setminus f^{-1}(B)$.*
(c) *If $g\colon Y \to Z$ is another map of sets, and $C \subseteq Z$, then we have $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$.*

Proof. For the first claim, we have $x \in f^{-1}(\bigcap_i B_i)$ iff $f(x) \in \bigcap_i B_i$ iff ($f(x) \in B_i$ for all i) iff ($x \in f^{-1}(B_i)$ for all i) iff $x \in \bigcap_i f^{-1}(B_i)$. All the other claims can be proved in a similar way. $\square$

Proposition 3.12 (MS, page 56)**.** *Let $f\colon X \to Y$ be a function between metric spaces. Then the following are equivalent:*

(a) *f is continuous.*
(b) *For every open set $V \subseteq Y$, the preimage $f^{-1}(V) \subseteq X$ is an open subset of X.*
(c) *For every closed set $G \subseteq Y$, the preimage $f^{-1}(G) \subseteq X$ is a closed subset of X.*

We can now introduce the theory of topological spaces, which was mentioned briefly at the end of [MS, Section 3].

Video (Definition 3.13 to Example 3.15)

Definition 3.13. Let X be a set. A *topology* on X is a collection τ of subsets of X (which are called *open* sets) with the following properties:

(a) The empty set and the whole set X are open.
(b) The union of any collection of open sets is open.
(c) The intersection of any finite collection of open sets is open.

Example 3.14. If X is a metric space, we can define open sets as in Definition 3.7. Proposition 3.9 then tells us that these open sets satisfy the axioms in Definition 3.13, so they give a topology on X, which we call the *metric topology*.

Example 3.15. For any set X, we can introduce a topology by declaring that *every* subset is open. This is called the *discrete topology*. Although this is not very interesting, it does occur naturally: it is the most natural topology on $\mathbb{Z}$ for example. At the other extreme, we can introduce a different topology by declaring that only the sets $\emptyset$ and X are open. This is called the *indiscrete topology*. Unlike the discrete topology, this is rarely relevant.

Video (Lemma 3.16 to Corollary 3.17)

Lemma 3.16. *Suppose we have two different metrics on X, say d_1 and d_2. Suppose we also have positive constants c_1 and c_2 such that $d_1(x,y) \leq c_1 d_2(x,y)$ and $d_2(x,y) \leq c_2 d_1(x,y)$. Then a set $U \subseteq X$ is open with respect to d_1 iff it is open with respect to d_2. Thus, the metrics d_1 and d_2 give the same topology.*

Proof. Suppose that U is open with respect to d_1. Consider a point $a \in U$. By assumption, there exists $r > 0$ such that $B_1(a,r) \subseteq U$, where $B_1(a,r)$ is the open ball defined using d_1, or in other words

$$B_1(a,r) = \{x \mid d_1(a,x) < r\}.$$

We claim that $B_2(a, r/c_1)$ is also contained in U. Indeed, if $x \in B_2(a, r/c_1)$, then $d_2(a, x) < r/c_1$, so $d_1(a, x) \le c_1 d_2(a, x) < r$, so $x \in B_1(a, r)$. However, we have $B_1(a, r) \subseteq U$ by assumption, so $x \in U$. This proves that $B_2(a, r/c_1) \subseteq U$ as claimed. As we can do this for any $a \in U$, we see that U is open with respect to d_2. By a symmetrical argument, if U is any set that is open with respect to d_2, then it is also open with respect to d_1. Thus d_1 and d_2 have the same open sets, and thus the same topology. $\square$

Corollary 3.17. *The metrics d_1, d_2 and d_∞ all give the same topology on $\mathbb{R}^n$.*

Proof. In the light of the lemma, it will be enough to prove the inequalities

$$d_\infty(x, y) \le d_2(x, y) \le d_1(x, y) \le n d_\infty(x, y).$$

Put $z_i = |x_i - y_i| \ge 0$, then choose p such that z_p is the largest of all the terms z_i. We then have

$$d_\infty(x, y) = \max(z_1, \ldots, z_n) = z_p,$$

so the required inequalities are

$$z_p \le \sqrt{\sum_i z_i^2} \le \sum_i z_i \le n\, z_p.$$

For the third inequality, the sum contains n terms each of which is less than or equal to z_p, so the total is at most $n\, z_p$, as required. The other two inequalities are equivalent to

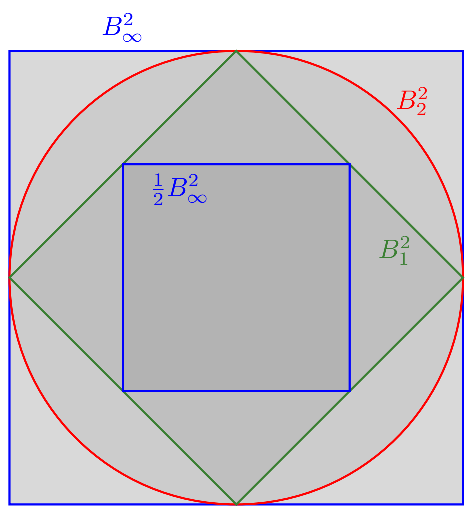
$$z_p^2 \le \sum_i z_i^2 \le (\sum_i z_i)^2 = \sum_{i,j} z_i z_j.$$

The first sum consists of z_p^2 plus some other nonnegative terms. In the second sum, the terms with $i = j$ give the same as the first sum, and there are again some additional nonnegative terms. The claim is clear from this. $\square$

Remark 3.18. For $p \in \{1, 2, \infty\}$ put

$$B_p^n = \{x \in \mathbb{R}^n \mid d_p(0, x) \le 1\}.$$

The inequalities in the above proof are equivalent to the claim that $\frac{1}{n}.B_\infty^n \subseteq B_1^n \subseteq B_2^n \subseteq B_\infty^n$. This can be illustrated in the case $n = 2$ as follows:



Video (Definition 3.19 to Proposition 3.24)

Definition 3.19. Let X and Y be topological spaces, and let f be a function from X to Y. We say that f is *continuous* if for every open set $V \subseteq Y$, the preimage $f^{-1}(V)$ is open in X.

Remark 3.20. Proposition 3.12 shows that when X and Y are metric spaces with the metric topology, this agrees with our earlier definition of continuity.

Example 3.21. We can define a function $f\colon \mathbb{R}^4 \to \mathbb{R}^4$ by

$$f(u,v,w,x) = (e^u\cos(x) + w\,\sin(x),\ -e^u\sin(x) + w\,\cos(x),\ e^v\sin(x),\ e^v\cos(x)).$$

The functions e^x, $\sin(x)$ and $\cos(x)$ are known to be continuous, by basic real analysis. It is also standard that sums and products of continuous functions are continuous, and it follows that each of the four components of f is continuous. From this it follows easily that f itself is continuous. In future, we will not bother to discuss this kind of argument in detail.

Example 3.22. As a very basic example, consider a constant function $f\colon X \to Y$. This means that there is a single point $b \in Y$ such that $f(x) = b$ for all x. Consider an open set $U \subseteq Y$. If $b \in U$ then $f^{-1}(U) = X$, and if $b \notin U$ then $f^{-1}(U) = \emptyset$. As both X and $\emptyset$ are open in X, we see that $f^{-1}(U)$ is always open. This proves that f is continuous.

Proposition 3.23. *Let f be as above. Then f is continuous iff for every closed subset $G \subseteq Y$, the preimage $f^{-1}(G)$ is closed in X.*

Proof. Suppose that f is continuous. Let $G \subseteq Y$ be closed; we must show that $f^{-1}(G)$ is closed in X. As G is closed in Y, we know that the complement $Y \setminus G$ must be open in Y, so $f^{-1}(Y \setminus G)$ is open in X (by the definition of continuity). However, $f^{-1}(Y \setminus G)$ is the same as $X \setminus f^{-1}(G)$. As this is open, we see that $f^{-1}(G)$ must be closed, as required.

The converse can be proved in essentially the same way. □

Proposition 3.24. *Suppose that X, Y and Z are topological spaces, and we have continuous maps $f\colon X \to Y$ and $g\colon Y \to Z$. Then the composite map $g \circ f\colon X \to Z$ is also continuous. Moreover, the identity map $\mathrm{id}\colon X \to X$ is also continuous.*

Proof. Consider an open set $W \subseteq Z$; we must show that the preimage $(g \circ f)^{-1}(W)$ is open in X. As g is continuous, we see that $g^{-1}(W)$ is an open subset of Y. As f is continuous, it follows in turn that $f^{-1}(g^{-1}(W))$ is open in X. However, we have $x \in f^{-1}(g^{-1}(W))$ iff $f(x) \in g^{-1}(W)$ iff $g(f(x)) \in W$ iff $x \in (g\circ f)^{-1}(W)$, so $f^{-1}(g^{-1}(W))$ is the same as $(g \circ f)^{-1}(W)$. We have therefore proved that $(g \circ f)^{-1}(W)$ is open in X, as required.

For the second claim, suppose that U is an open subset of X. Then $\mathrm{id}^{-1}(U)$ is just the same as U, and so is open. This proves that id is continuous. □

Definition 3.25. Let X be a topological space, and let Y be a subset of X. We declare that a subset $V \subseteq Y$ is open in Y if there exists an open set U of X such that $V = U \cap Y$.

Proposition 3.26. *The above definition gives a topology on Y (which we call the* subspace topology*).*

Video

Proof.

(a) The empty set can be written as the intersection of Y with the open set $\emptyset$ of X, so the empty set is open in Y. The full set Y can be written as the intersection of Y with the open set X of X, so Y is also open in Y.

(b) Suppose we have a collection of sets $V_i \subseteq Y$ that are open in Y; we must show that the union $V^* = \bigcup_i V_i$ is also open in Y. As each V_i is open in Y, we can find open sets $U_i \subseteq X$ such that $V_i = U_i \cap Y$. We are assuming that the axioms for a topology are satisfied by the open sets in X, so the union $U^* = \bigcup_i U_i$ is again open in X. The set V^* can be written as $U^* \cap Y$, so it is open as required.

(c) Now suppose instead that we have a finite collection of sets $V_1, \dots, V_n$ that are open in Y; we must show that the intersection $V^\# = V_1 \cap \cdots \cap V_n$ is also open in Y. As in (b), we can choose open subsets $U_1, \dots, U_n$ with $V_i = U_i \cap Y$. We are assuming that the axioms for a topology are satisfied by the open sets in X, so the intersection $U^\# = \bigcap_i U_i$ is again open in X. The set $V^\#$ can be written as $U^\# \cap Y$, so it is open as required.

□

Video (Lemma 3.27 to Proposition 3.29)

One very basic point about the subspace topology is as follows. We can define $i\colon Y \to X$ by $i(y) = y$ for all $y \in Y$. We call this the *inclusion map.*

Lemma 3.27. *If we give Y the subspace topology, then the inclusion map $i\colon Y \to X$ is continuous.*

Proof. Consider an open set $U \subseteq X$; we must show that $i^{-1}(U)$ is open with respect to the subspace topology on Y. But $i^{-1}(U)$ is just the same as $U \cap Y$, which is open by the definition of the subspace topology. □

We next deal with a slightly technical point. Suppose, for example, we want to discuss whether the function $f(x) = \sin(x)$ is continuous. We might want to regard this as a function from $\mathbb{R}$ to $\mathbb{R}$, or as a function from $\mathbb{R}$ to $[-1, 1]$. We also might want to restrict the range of values of x, and consider f as a function defined on $[0, 2\pi]$ or $[-\pi, \pi]$ or $[0, \infty)$ instead of all of $\mathbb{R}$. This leads us to worry about the following possibility: perhaps some of these versions of $f(x)$ are continuous, and some of them are not. It would then be a nightmare to keep track of everything. Fortunately, however, this nightmare does not arise: if we can check that f is continuous as a map $\mathbb{R} \to [-1, 1]$, then all other versions of f are automatically continuous. This is the message of the following proposition.

Proposition 3.28. *Let X and Y_1 be topological spaces. Let X_0 be a subset of X, and let Y be a subset of Y_1, with inclusion maps $i\colon X_0 \to X$ and $j\colon Y \to Y_1$. Let f be a continuous function from X to Y. Let $\overline{f} = j \circ f \circ i$ be the corresponding map $X_0 \to Y_1$, obtained by restricting the domain to X_0 and enlarging the codomain to Y_1, so we have a commutative diagram as follows:*

$$\begin{tikzcd} X_0 \arrow[r, "\overline{f}"] \arrow[d, "i"'] & Y_1 \\ X \arrow[r, "f"'] & Y \arrow[u, "j"'] \end{tikzcd}$$

Then $\overline{f}$ is also continuous.

Proof. This is just because i and j are continuous by Lemma 3.27, so $j \circ f \circ i$ is continuous by Proposition 3.24. □

There is also a partial converse for the above result.

Proposition 3.29. *Let X and Y_1 be topological spaces, and let $j\colon Y \to Y_1$ be the inclusion of a subset with the subspace topology. Let f be a function from X to Y, and suppose that $j \circ f$ is continuous (or equivalently: f is continuous when regarded as a function $X \to Y_1$). Then f is continuous (as a function $X \to Y$).*

Proof. Let V be a subset of Y that is open with respect to the subspace topology. We must show that $f^{-1}(V)$ is open in X. By the definition of the subspace topology, there must exist an open set $V_1 \subseteq Y_1$ such that $V = V_1 \cap Y$. This can also be written as $V = j^{-1}(V_1)$, so $f^{-1}(V) = f^{-1}(j^{-1}(V_1)) = (j \circ f)^{-1}(V_1)$. We are given that $j \circ f\colon X \to Y_1$ is continuous and V_1 is open in Y_1 so $(j \circ f)^{-1}(V_1)$ is open in X. In other words, $f^{-1}(V)$ is open in X as required. □

Video (Lemmas 3.30 and 3.32)

Lemma 3.30. *Let Y be an open subset of X, and let V be a subset of Y. Then V is open with respect to the subspace topology on Y iff V is open with respect to the original topology on X.*

Proof. First suppose that V is open with respect to the subspace topology. By definition, this means that $V = U \cap Y$ for some subset $U \subseteq X$ that is open in X. Now both U and Y are open in X, so the intersection $V = U \cap Y$ is also open in X, as required.

Suppose instead that we start from the assumption that V is open with respect to the original topology on X. If we just take $U = V$, we see that U is open in X and $V = U \cap Y$, so V is open with respect to the subspace topology. □

Lemma 3.31. *Let Y be an arbitrary subset of X, and let G be a subset of Y. Then G is closed (with respect to the subspace topology on Y) iff there exists a closed set $F \subseteq X$ such that $G = F \cap Y$.*

Proof. If G is closed in Y, then the complement $V = Y \setminus G$ must be open in Y. This means by definition that there exists an open set $U \subseteq X$ such that $V = U \cap Y$. We can now put $F = X \setminus U$, which is a closed subset of X. We find that

$$F \cap Y = (X \setminus U) \cap Y = Y \setminus (U \cap Y) = Y \setminus V = G,$$

so G has the form $F \cap Y$, as required. We leave the converse to the reader. $\square$

Lemma 3.32. *Let Y be a closed subset of X, and let G be a subset of Y. Then G is closed with respect to the subspace topology on Y iff G is closed with respect to the original topology on X.*

Proof. Now that we have Lemma 3.31, this can be proved in essentially the same way as Lemma 3.30. $\square$

Lemma 3.33. *Let X be a topological space, and let U be a subset of X. Suppose that for each $x \in U$ we can find an open set V such that $x \in V$ and $V \subseteq U$. Then U itself is open.*

Proof. For each point $x \in U$, choose a set V_x as described, so V_x is open and $x \in V_x$ and $V_x \subseteq U$. Put $V^* = \bigcup_x V_x$. This is the union of a family of open sets, so it is open by the axioms for a topology. If we can prove that V^* is the same as U, then we will be done. Each set V_x is contained in U, and it follows that the union V^* is also contained in U. On the other hand, for each $x \in U$ we have $x \in V_x \subseteq V^*$, so $U \subseteq V^*$. As $U \subseteq V^*$ and $V^* \subseteq U$ we have $U = V^*$ as required. $\square$

Video (Propositions 3.34 and 3.35)

Proposition 3.34 (Open patching)**.** *Let $f : X \to Y$ be a function between topological spaces. Suppose we have subsets $U_1, \dots, U_n \subseteq X$ such that*

(a) *Each set U_i is open.*
(b) $X = U_1 \cup \dots \cup U_n$
(c) *For each i, the restricted map $f_i : U_i \to Y$ is continuous (with respect to the subspace topology on U_i).*

Then f is continuous.

Proof. Let V be an open subset of Y; we must show that $f^{-1}(V)$ is open in X. Let $f_i : U_i \to Y$ be the restriction of f. As f_i is continuous by assumption, we know that $f_i^{-1}(V)$ is open in U_i. By Lemma 3.30, this is the same as being open in X. Moreover, we see from the definitions that $f_i^{-1}(V) = f^{-1}(V) \cap U_i$. As $\bigcup_i U_i = X$, we see that

$$f^{-1}(V) = \bigcup_i (f^{-1}(V) \cap U_i) = \bigcup_i f_i^{-1}(V).$$

We have seen that each set $f_i^{-1}(V)$ is open, so the union of these sets is open, so $f^{-1}(V)$ is open as required. $\square$

Proposition 3.35 (Closed patching)**.** *Again let $f : X \to Y$ be a function between topological spaces. Suppose we have subsets $F_1, \dots, F_n \subseteq X$ such that*

(a) *Each set F_i is closed.*
(b) $X = F_1 \cup \dots \cup F_n$
(c) *For each i, the restricted map $f_i : F_i \to Y$ is continuous (with respect to the subspace topology on F_i).*

Then f is continuous.

Proof. By Proposition 3.23, it will be enough to show that $f^{-1}(G)$ is closed in X whenever G is closed in Y. After this preliminary step, the rest of the proof is essentially the same as for the previous proposition. $\square$

4. Homeomorphism

Video (Definition 4.1 to Example 4.8)

Definition 4.1. Let X and Y be topological spaces, and let $f\colon X \to Y$ be a function. We say that f is a *homeomorphism* if

(a) f is a bijection, so there is an inverse map $f^{-1}\colon Y \to X$, satisfying $f^{-1}(f(x)) = x$ for all $x \in X$ and $f(f^{-1}(y)) = y$ for all $y \in Y$.
(b) Both f and f^{-1} are continuous.

We say that X and Y are *homeomorphic* if there exists a homeomorphism from X to Y. We will write $X \simeq Y$ to indicate that X and Y are homeomorphic.

Remark 4.2. Later we will introduce a different notion called *homotopy equivalence*, and write $X \cong Y$ (with an extra bar) to indicate that X and Y are homotopy equivalent. It is important to distinguish between these two ideas. However, you should be aware that there is no consistency in the literature about the notation used. It is safest to say "homeomorphic" or "homotopy equivalent" in words, to avoid confusion.

Remark 4.3. There are very few indirect techniques for proving that two spaces are homeomorphic. Instead, we just have to find a specific homeomorphism. Many examples will be given below.

Proposition 4.4. *Suppose that $a, b, c, d \in \mathbb{R}$ with $a < b$ and $c < d$. Then $[a, b] \simeq [c, d]$ and $(a, b) \simeq (c, d)$ and $[a, b) \simeq [c, d)$ and $(a, b] \simeq (c, d]$.*

Proof. We can define maps $[a, b] \xrightarrow{f} [c, d] \xrightarrow{g} [a, b]$ by

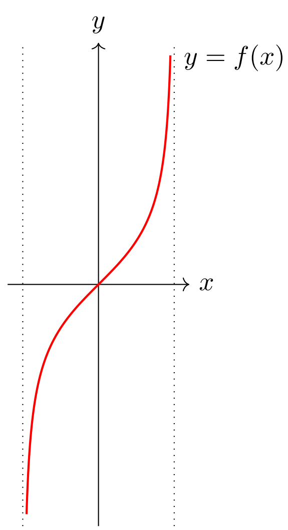
$$f(t) = c + \frac{d-c}{b-a}(t-a) \qquad\qquad g(t) = a + \frac{b-a}{d-c}(t-c).$$

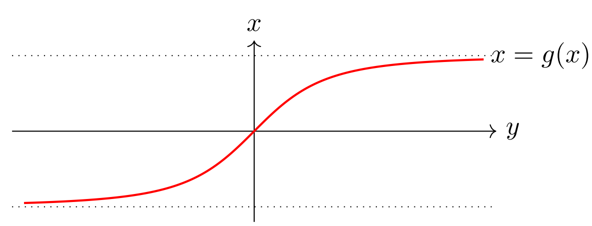
We find that $f(g(t)) = t$ and $g(f(t)) = t$, so g is inverse to f. It is clear that f and g are continuous, so f is a homeomorphism. The same formulae also give homeomorphisms $(a, b) \to (c, d)$ and so on. $\square$

Proposition 4.5. *The formulae*

$$f(x) = \frac{x}{\sqrt{1-x^2}} \qquad\qquad g(y) = \frac{y}{\sqrt{1+y^2}}$$

give continuous maps $(-1, 1) \xrightarrow{f} \mathbb{R} \xrightarrow{g} (-1, 1)$ which are inverse to each other. Thus, they are both homeomorphisms, and $(-1, 1) \simeq \mathbb{R}$. Moreover, the same maps restrict to give homeomorphisms between $[0, 1)$ and $[0, \infty)$.





Proof. First, note that when $x \in (-1, 1)$ we have $0 \leq x^2 < 1$ so $0 < 1 - x^2 \leq 1$, so dividing by $\sqrt{1-x^2}$ does not cause any problems. Thus, f is a continuous function from $(-1, 1)$ to $\mathbb{R}$. Similarly, for all y we have $1 + y^2 \geq 1 > 0$ so $\sqrt{1+y^2} \geq 1 > 0$, so dividing by $\sqrt{1+y^2}$ does not cause a problem, and g is a continuous map from $\mathbb{R}$ to $\mathbb{R}$. Note also that

$$g(y)^2 = y^2/(1+y^2) = 1 - 1/(1+y^2) < 1,$$

so $g(y) \in (-1, 1)$, so g can in fact be regarded as a continuous map $g\colon \mathbb{R} \to (-1, 1)$.

Note that if $y = f(x) = x/\sqrt{1-x^2}$ then

$$1 + y^2 = 1 + \frac{x^2}{1-x^2} = \frac{1}{1-x^2}$$

so $\sqrt{1+y^2} = (1-x^2)^{-1/2}$, so $g(y) = y/\sqrt{1+y^2} = x$. This shows that $g(f(x)) = x$.

Conversely, suppose we start with y and put $x = g(y) = y/\sqrt{1+y^2}$. We then have

$$1 - x^2 = 1 - \frac{y^2}{1+y^2} = \frac{1}{1+y^2},$$

so $\sqrt{1-x^2} = (1+y^2)^{-1/2}$, so $f(x) = x/\sqrt{1-x^2} = y$. This shows that $f(g(y)) = y$, so $f\colon (-1, 1) \to \mathbb{R}$ and $g\colon \mathbb{R} \to (-1, 1)$ are inverses of each other, as required. $\square$

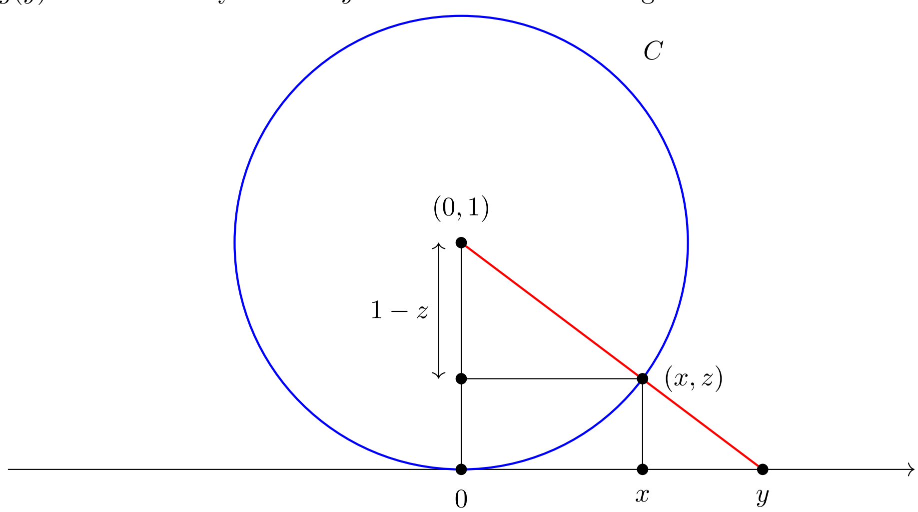
Remark 4.6. Suppose we are trying to define a map $p\colon U \to V$, and we give a formula for $p(u)$. For this to be valid, we must check two things.

(a) The formula must be meaningful for all elements $u \in U$. It must never involve division by zero, or square roots of negative numbers in a context where a real result is required, for example.

(b) The resulting value $p(u)$ must lie in V. For example, V might be a subspace of $\mathbb{R}^3$ defined by various equations or inequalities. Typically it will only be obvious from the formula that $p(u)$ lies in $\mathbb{R}^3$, so we need to check that the equations or inequalities are satisfied as an extra step.

You should observe how these checks were done in the proof of Proposition 4.5. It is sadly common for them to be omitted in homework or exam answers submitted by students. Do not let that be you.

Remark 4.7. There is some interesting geometry behind the maps f and g in Proposition 4.5: we have $f(x) = y$ and $g(y) = x$ if and only if x and y are related as in the diagram below.



In more detail, let C be the circle of radius one centred at $(0, 1)$. Given a point $x \in (-1, 1)$, let (x, z) be the point on the lower half of C lying directly above $(x, 0)$. Then draw a line from $(0, 1)$ through (x, z), and let y be the point where it meets the axis. I claim that $y = x/\sqrt{1-x^2} = f(x)$. Indeed, as the point (x, z) lies on C, the distance from (x, z) to $(0, 1)$ must be one, so $x^2 + (1-z)^2 = 1$, so $1 - z = \sqrt{1-x^2}$. On the other hand, the diagram contains a triangle of base y and height 1, and a nested triangle of base x and height $1 - z$. These triangles have the same angles, so we must have $y/1 = x/(1-z)$, so $y = x/\sqrt{1-x^2}$ as claimed.

Example 4.8. The previous example can be generalised as follows. Put $OB^n = \{x \in \mathbb{R}^n \mid \|x\| < 1\}$ (so in particular $OB^1 = (-1,1) \subset \mathbb{R}$). We can define $f\colon OB^n \to \mathbb{R}^n$ and $g\colon \mathbb{R}^n \to OB^n$ by

$$f(x) = x/\sqrt{1-\|x\|^2}$$
$$g(y) = y/\sqrt{1+\|y\|^2}.$$

Using essentially the same argument as given above, we see that f and g are mutually inverse homeomorphisms, so OB^n is homeomorphic to $\mathbb{R}^n$.

Example 4.9. The map $f(x) = (\|x\|, x/\|x\|)$ gives a homeomorphism $\mathbb{R}^n \setminus \{0\} \to (0,\infty) \times S^{n-1}$, with inverse $f^{-1}(t,y) = ty$.

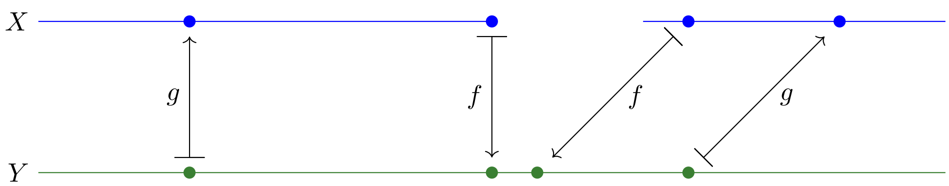
Interactive demo

Example 4.10. Video

Note that Definition 4.1 specifies that both f and f^{-1} must be continuous. It is necessary to make the definition this way, because it can easily happen that we have a bijection $f\colon X \to Y$ where f is continuous but f^{-1} is not, and we do not want to count maps like that as homeomorphisms. We give such an example here.

Put $X = (-\infty, 0] \cup (1,\infty) = \mathbb{R} \setminus (0,1]$ and $Y = \mathbb{R}$. Define maps $X \xrightarrow{f} Y \xrightarrow{g} X$ by

$$f(x) = \begin{cases} x & \text{if } x \leq 0 \\ x-1 & \text{if } x > 0. \end{cases} \qquad g(x) = \begin{cases} x & \text{if } x \leq 0 \\ x+1 & \text{if } x > 0. \end{cases}$$



It is easy to see that f and g are inverse to each other. We claim that f is continuous but that g is not. To see this, we introduce the subsets

$$U = (-\infty, 0] = X \cap (-\infty, \tfrac{1}{2})$$
$$V = (1,\infty) = X \cap (\tfrac{1}{2}, \infty).$$

The sets $(-\infty, \frac{1}{2})$ and $(\frac{1}{2}, \infty)$ are open in $\mathbb{R}$, so U and V are open in the subspace topology on X. The map f is given by $f(x) = x$ on U and by $f(x) = x-1$ on V, so the restrictions to U and V are both continuous, so f is continuous on X by open patching (Proposition 3.34). On the other hand, we have $g^{-1}(U) = (-\infty, 0]$, which is not an open subset of the space $Y = \mathbb{R}$; this proves that g is not continuous. We therefore have a continuous bijection whose inverse is not continuous, so it does not count as a homeomorphism.

Example 4.11. Video

Define $f\colon S^1 \times \{1,-1\} \to O_2$ by

$$f(x,y,z) = \begin{bmatrix} x & -yz \\ y & xz \end{bmatrix}$$

so

$$f(\cos(\theta), \sin(\theta), +1) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \text{ rotation matrix}$$

and

$$f(\cos(\theta), \sin(\theta), -1) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix} = \text{ reflection matrix}$$

It is easy to see that this is a homeomorphism, with inverse

$$f^{-1}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = (a, c, ad - bc).$$

We can also define $g\colon \mathbb{R}^3 \times S^1 \times \{1, -1\} \to GL_2(\mathbb{R})$ by

$$g(u, v, w, x, y, z) = \begin{bmatrix} e^u & 0 \\ 0 & e^v \end{bmatrix} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x & -yz \\ y & xz \end{bmatrix}.$$

This can also be shown to be a homeomorphism. The inverse is

$$g^{-1}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = (u, v, w, x, y, z),$$

where

$$\begin{aligned} \Delta &= ad - bc & w &= (ac + bd)/|\Delta| \\ u &= \ln|\Delta| - \tfrac{1}{2}\ln(c^2 + d^2) & x &= \operatorname{sgn}(\Delta)d/\sqrt{c^2 + d^2} \\ v &= \tfrac{1}{2}\ln(c^2 + d^2) & y &= c/\sqrt{c^2 + d^2} \\ & & z &= \operatorname{sgn}(\Delta). \end{aligned}$$

Remark 4.12. We could try to define a map $g\colon S^1 \to S^1$ by the rule

(A) $$g((\cos(\theta), \sin(\theta)) = (\cos(\theta/2), \sin(\theta/2)).$$

There could be two different things wrong with this definition, depending on how we interpret it.

(a) Any point in S^1 can be expressed as $(\cos(\theta), \sin(\theta))$ for infinitely many different values of θ, differing by multiples of 2π. For example, the point $(-1, 0)$ can be represented as $(\cos(\pi), \sin(\pi))$ or as $(\cos(-\pi), \sin(-\pi))$. The first representation gives $g((1, 0)) = (\cos(\pi/2), \sin(\pi/2)) = (0, 1)$, and the second representation gives $g((-1, 0)) = (\cos(-\pi/2), \sin(-\pi/2)) = (0, -1)$. Thus, equation (A) does not give us a well-defined function $S^1 \to S^1$.

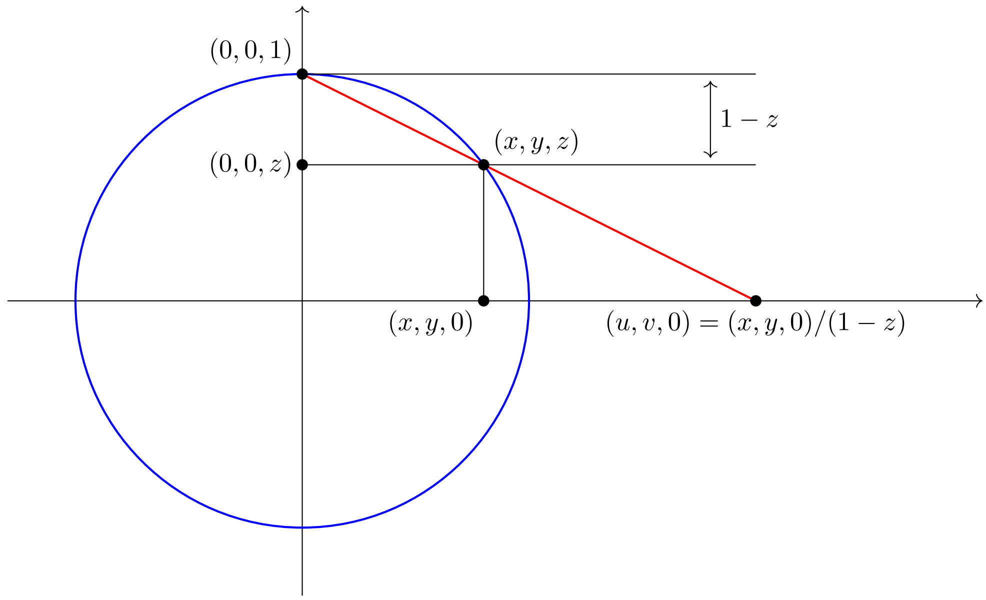
(b) We could instead note that any point in S^1 can be expressed as $(\cos(\theta), \sin(\theta))$ for a unique choice of θ satisfying the auxiliary condition $-\pi < \theta \leq \pi$, and we could define g by saying that formula (A) holds for this choice of θ. This gives a well-defined function, which satisfies $g((-1, 0)) = (0, 1)$. However, for points $(x, y) \in S^1$ lying just below $(-1, 0)$, we find that $\theta \approx -\pi$ and so $g(x, y)$ is close to $(0, -1)$. This shows that g is discontinuous, despite the apparent continuity of all ingredients in equation (A).

For this kind of reason, it is generally a bad idea to define functions in terms of θ. This is why our initial definition of $f\colon S^1 \times \{1, -1\} \to O_2$ in Example 4.11 was given directly in terms of the coordinates x, y and z.

Example 4.13. Let N be the "north pole" of the sphere S^2, in other words the point $(0, 0, 1)$. There are mutually inverse continuous maps $S^2 \setminus \{N\} \xrightarrow{f} \mathbb{R}^2 \xrightarrow{g} S^2 \setminus \{N\}$ given by

$$f(x, y, z) = \frac{(x,\ y)}{1 - z} \qquad\qquad g(u, v) = \frac{(2u,\ 2v,\ u^2 + v^2 - 1)}{u^2 + v^2 + 1},$$

so $S^2 \setminus \{N\}$ is homeomorphic to $\mathbb{R}^2$. This is called *stereographic projection*. Geometrically, $f(x, y, z)$ is the unique point where the line joining N to (x, y, z) meets the plane $z = 0$, and $g(u, v)$ is the unique point where the line joining N to $(u, v, 0)$ meets S^2:



Interactive demo

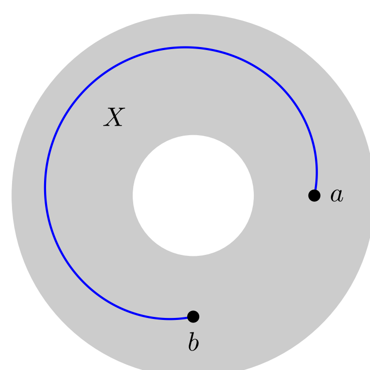
We will leave it to the reader to check directly from the formulae that $f(g(u,v)) = (u,v)$ and $g(f(x,y,z)) = (x,y,z)$. Essentially the same formulae can be used to prove that $S^n \setminus \{P\} \simeq \mathbb{R}^n$ for any $n \geq 1$ and any $P \in S^n$.

5. Paths

Video (Definition 5.1 to Definition 5.10)

Definition 5.1. Let X be a topological space. A *path* in X is a continuous function $u\colon [0,1] \to X$. If $u(0) = a$ and $u(1) = b$ then we say that u is a path from a to b and write $u\colon a \rightsquigarrow b$.

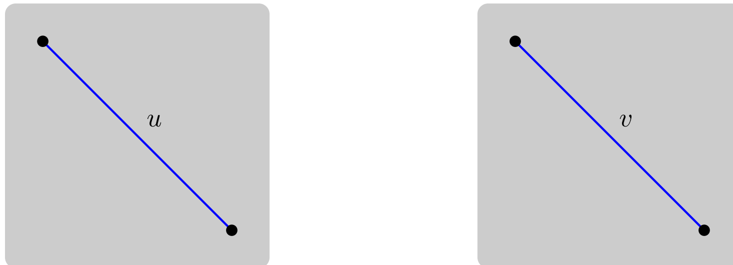
Example 5.2. This picture shows a path u from the point $a = (2,0)$ to the point $b = (0,-2)$ in the space $X = \{(x,y) \in \mathbb{R}^2 \mid 1 \leq \|(x,y)\| \leq 3\}$:



Explicitly, the formula is

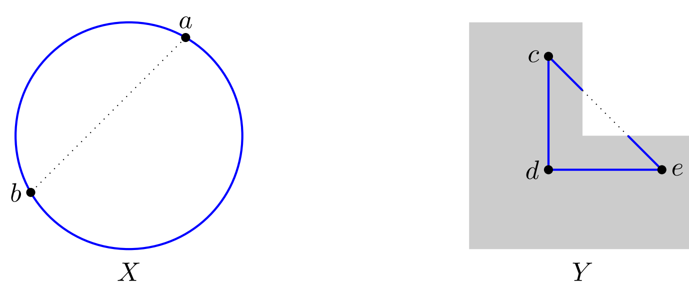
$$u(t) = 2(1 + t - t^2)\ (\cos(3\pi t/2),\ \sin(3\pi t/2))\,.$$

Example 5.3. Note, however, that pictures like the one above can be a little misleading, because they only show the points traversed by the path, not the time at which those points are reached. Consider the maps $u, v\colon [0,1] \to \mathbb{R}^2$ given by $u(t) = (t, 1-t)$ and $v(t) = (t^2, 1-t^2)$. The resulting tracks are just the same:



However, if we add markers for $t = 0.1, 0.2, \dots, 0.9$ then we can see the difference:

Example 5.4. Suppose that X is a subset of $\mathbb{R}^n$, with the subspace topology. For any points $a, b \in X$, we can define $u\colon [0,1] \to \mathbb{R}^n$ by $u(t) = (1-t)a + tb$. This gives a path from a to b in $\mathbb{R}^n$, which we call a *straight line path*. However, this path might or might not lie in X; in any case where we want to use straight line paths, we need to check this. For example, if X is a circle then the straight line path from a to b is not contained in X, except in the trivial case where $a = b$. In the space Y shown on the right below, the straight line paths from c to d and from d to e are contained in Y, but the straight line path from c to e is not contained in Y.



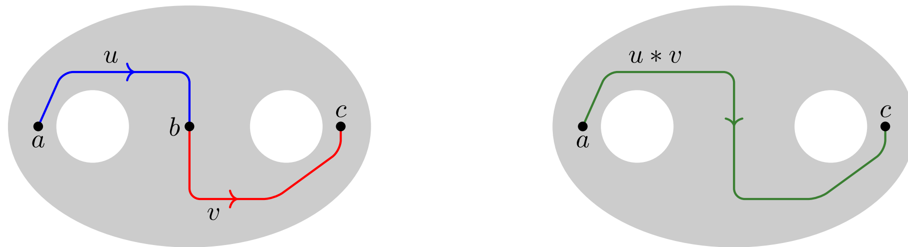
Remark 5.5. Later, we will want to consider paths as continuous maps $\Delta_1 \to X$ rather than continuous maps $[0,1] \to X$. We will always identify the point $t \in [0,1]$ with the point $(1-t, t) \in \Delta_1$. This ensures that the point $0 \in [0,1]$ gets identified with $e_0 = (1,0) \in \Delta_1$, and the point $1 \in [0,1]$ gets identified with $e_1 = (0,1) \in \Delta_1$.

Definition 5.6. Let X be a topological space.

(a) For any $a \in X$, the *constant path* $c_a\colon [0,1] \to X$ is just given by $c_a(t) = a$ for all t. This is a path $a \rightsquigarrow a$.
(b) Now suppose we have a path $u\colon a \rightsquigarrow b$. We define a path $\overline{u}\colon b \rightsquigarrow a$ by $\overline{u}(t) = u(1-t)$, and we call this the *reverse* of u.
(c) Now suppose we also have a path $v\colon b \rightsquigarrow c$. We define a path $u * v\colon a \rightsquigarrow c$ by

$$(u * v)(t) = \begin{cases} u(2t) & \text{if } 0 \leq t \leq \frac{1}{2} \\ v(2t-1) & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

(Roughly speaking, in the first half-second we go from a to b by following u at double speed, then in the next half-second we go from b to c by following v at double speed.)



Remark 5.7. There are a number of things that we need to check in order to validate the above definitions, especially part (c). Firstly, c_a is continuous by Example 3.22, and $\overline{u}$ is continuous by Proposition 3.24, because it is the composite of u with the continuous function $t \mapsto 1-t$.

Now consider $u * v$. Firstly, if $t = \frac{1}{2}$ then both clauses in the definition of $(u * v)(t)$ are applicable. This would be a problem if the two clauses gave different answers for the value of $(u * v)(\frac{1}{2})$. However, the first clause gives the answer $u(1)$, and the second clause gives the answer $v(0)$. As $u\colon a \rightsquigarrow b$ and $v\colon b \rightsquigarrow c$ we have $u(1) = b = v(0)$ so the two answers are the same.

Next, we need to show that the map $u * v\colon [0,1] \to X$ is continuous. We can write $[0,1]$ as $F_1 \cup F_2$, where $F_1 = [0, \frac{1}{2}]$ and $F_2 = [\frac{1}{2}, 0]$. By closed patching (Proposition 3.35), it will be enough to show that $u * v$ is continuous on F_1 and also continuous on F_2. On F_1, we have $(u * v)(t) = u(2t)$. This is the composite of u with the map $t \mapsto 2t$, and both of these maps are continuous, so $u * v$ is continuous by Proposition 3.24. A similar argument shows that $u * v$ is continuous on F_2, so it is continuous on $[0,1]$ as required. We also have $(u * v)(0) = u(2 \times 0) = u(0) = a$ and $(u * v)(1) = v(2 \times 1 - 1) = v(1) = c$, so $u * v\colon a \rightsquigarrow c$.

Definition 5.8. We introduce a relation on X by declaring that $a \sim b$ iff there exists a path $u\colon a \rightsquigarrow b$ in X.

Proposition 5.9. *This relation is an equivalence relation.*

(If you need to review the basic ideas about equivalence relations, you can look forward to Definition 7.13.)

Proof. We must show that the relation is reflexive, symmetric and transitive. In more detail, the conditions are as follows:

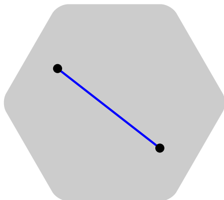
(a) For all $a \in X$, we must have $a \sim a$.
(b) For all $a, b \in X$ with $a \sim b$, we must have $b \sim a$.
(c) For all $a, b, c \in X$ with $a \sim b$ and $b \sim a$ we must have $a \sim c$.

For condition (a), we always have $c_a\colon a \rightsquigarrow a$, and this shows that $a \sim a$. For condition (b), suppose that $a \sim b$. This means that there exists a path $u\colon a \rightsquigarrow b$. It follows that the reverse path $\overline{u}$ has $\overline{u}\colon b \rightsquigarrow a$, and thus that $b \sim a$. Finally, suppose that $a \sim b$ and $b \sim c$. This means that there exist paths $u\colon a \rightsquigarrow b$ and $v\colon b \rightsquigarrow c$. The joined path $u * v$ then goes from a to c, proving that $a \sim c$ as required. ☐

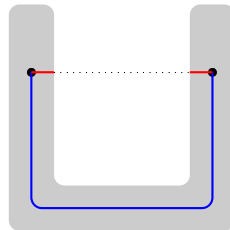
Definition 5.10. The equivalence classes for $\sim$ are called the *path components* of X. We write $[a]$ for the path component containing a, so that $[a] = [b]$ iff $a \sim b$. We write $\pi_0(X)$ for the quotient set $X/\sim$, or equivalently, the set of path components. We say that X is *path connected* if it has precisely one path component. This means that $X \neq \emptyset$, and $a \sim b$ for all $a, b \in X$.

Video (Example 5.11 to Proposition 5.14)

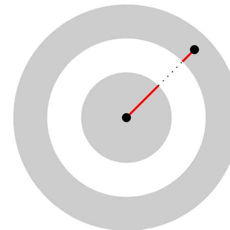
Example 5.11. Let X be a subset of $\mathbb{R}^n$. We say that X is *convex* if for all $a, b \in X$ and $t \in [0,1]$ we have $(1-t)a + tb \in X$. Equivalently, this means that any straight line path with endpoints in X is contained wholly in X, and so counts as a path in X between those endpoints. It therefore follows that X is path connected (provided that it is not empty).



convex



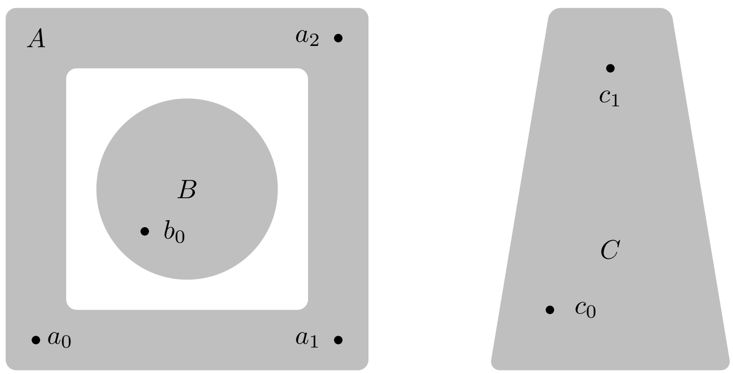
path connected, but not convex



not path connected

In particular, the ball B^n, the cube $[0,1]^n$ and the simplex Δ_n are all nonempty and convex, so they are path connected.

Example 5.12. Consider again the following space Y, which we discussed in Section 2:

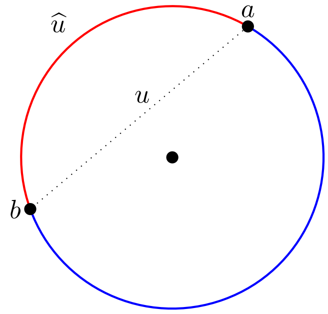


This is the disjoint union of the three subsets A, B and C. The set A is the set of all points that can be connected to a_0 by a continuous path in Y, or in other words $A = [a_0] \in \pi_0(Y)$. Because a_0, a_1 and a_2 all lie in A they can all be connected to each other, which means that $a_0 \sim a_1 \sim a_2$ and $[a_0] = [a_1] = [a_2] = A$. In the same way, we have $B = [b_0] \in \pi_0(Y)$ and $C = [c_0] = [c_1] \in \pi_0(Y)$. From this we see that $\pi_0(Y) = \{A, B, C\}$ and so $|\pi_0(Y)| = 3$.

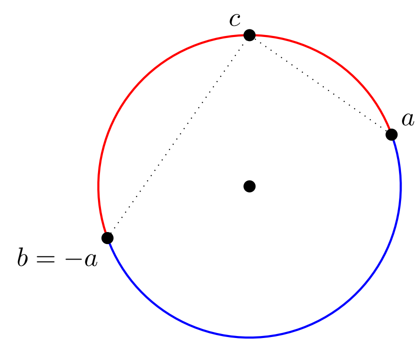
Remark 5.13. You should not be confused by the fact that A itself is an infinite set. The whole set A taken as a single object is an element of $\pi_0(Y)$, the whole set B taken as a single object is another element, and the whole set C is the third element.

Proposition 5.14. *For $n > 0$, the sphere S^n is path connected.*

Proof. Suppose we have points $a, b \in S^n$. Suppose for the moment that they are not opposite points, so $b \neq -a$. Consider the linear path $u\colon [0,1] \to \mathbb{R}^{n+1}$ given by $u(t) = (1-t)a + tb$. Because a and b are not opposite, we see that the straight line from a to b does not pass through the origin, so $u(t)$ is never zero. It is therefore legitimate to define $\widehat{u}(t) = u(t)/\|u(t)\|$. This gives a continuous map $\widehat{u}\colon [0,1] \to S^n$ with $\widehat{u}(0) = a/\|a\| = a$ and $\widehat{u}(1) = b/\|b\| = b$, so $a \sim b$ in S^n. Now consider the exceptional case where $b = -a$. If we allowed the case $n = 0$ then S^n would just consist of two points, but we have specified that $n > 0$, so S^n is infinite, so we can choose a point c that is different from a and $b = -a$. The ordinary case now tells us that $a \sim c$ and $b \sim c$, and $\sim$ is an equivalence relation so $a \sim b$ as required.



ordinary case



exceptional case

$\square$

Video (Proposition 5.15 to Proposition 5.18)

Proposition 5.15. *Let a and b be points in a topological space X. Suppose that there is a continuous function $f\colon X \to \mathbb{R}$ such that*

(a) *$f(x) \neq 0$ for all $x \in X$.*
(b) *$f(a) < 0 < f(b)$.*

Then $a \not\sim b$.

Proof. Suppose (for a contradiction) that there is a path u from a to b in X. We then have a continuous map $g = f \circ u\colon [0,1] \to \mathbb{R}$ with $g(0) = f(a) < 0$ and $g(1) = f(b) > 0$. By the Intermediate Value Theorem, there exists $t_0 \in [0,1]$ with $g(t_0) = 0$. This means that the point $x = u(t_0) \in X$ satisfies $f(x) = 0$, contradicting assumption (a). We therefore conclude that no such path u can exist, so $a \not\sim b$. $\square$

Example 5.16. Consider the space $\mathbb{Z}$. If $a, b \in \mathbb{Z}$ with $a < b$, we can define $f\colon \mathbb{Z} \to \mathbb{R}$ by $f(x) = x - b + \frac{1}{2}$. This is nonzero for all $x \in \mathbb{Z}$, and satisfies $f(a) < 0 < f(b)$, so $a \not\sim b$. It follows that the connected components are just the singleton sets $[a] = \{a\}$ for all $a \in \mathbb{Z}$, so $\pi_0(\mathbb{Z})$ is essentially the same as $\mathbb{Z}$.

Example 5.17. Consider the space $GL_2(\mathbb{R})$ of invertible 2×2 matrices, and the elements $I, J \in GL_2(\mathbb{R})$, where

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad\qquad J = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

We can define a continuous map $f\colon GL_2(\mathbb{R}) \to \mathbb{R}$ by $f(A) = \det(A)$. It is a standard fact of linear algebra that invertible matrices have nonzero determinant, so f is nonzero everywhere on $GL_2(\mathbb{R})$. We also have $f(J) = -1$ and $f(I) = 1$. Proposition 5.15 therefore tells us that $J \not\sim I$. More generally, we could take any $n > 0$ and put

$$U = \{A \in GL_n(\mathbb{R}) \mid \det(A) > 0\}$$
$$V = \{B \in GL_n(\mathbb{R}) \mid \det(B) < 0\}.$$

The same line of argument shows that if $B \in V$ and $A \in U$ then $B \not\sim A$. However, it can also be shown that the subsets U and V are both path connected (we will not give the proof here). Assuming this, we see that U and V are precisely the path components of $GL_n(\mathbb{R})$, so $\pi_0(GL_n(\mathbb{R})) = \{U, V\}$ and $|\pi_0(GL_n(\mathbb{R}))| = 2$.

Proposition 5.18. *Suppose that X can be written as $X = U \cup V$, where U and V are open subsets of X with $U \cap V = \emptyset$. Then for $a \in U$ and $b \in V$ we have $a \not\sim b$.*

Proof. We define $f\colon X \to \mathbb{R}$ by $f(x) = -1$ for all $x \in U$ and $f(x) = 1$ for all $x \in V$. (This defines $f(x)$ for all x, because $X = U \cup V$. There is no clash between the two clauses, because $U \cap V = \emptyset$.) We claim that f is continuous. (Assuming this, the main claim follows by Proposition 5.15.) Consider an open subset $A \subseteq \mathbb{R}$;

we must show that $f^{-1}(A)$ is open in X. We have

$$f^{-1}(A) = \begin{cases} \emptyset & \text{if } -1 \notin A \text{ and } 1 \notin A \\ U & \text{if } -1 \in A \text{ and } 1 \notin A \\ V & \text{if } -1 \notin A \text{ and } 1 \in A \\ X & \text{if } -1 \in A \text{ and } 1 \in A. \end{cases}$$

In all cases, we see that $f^{-1}(A)$ is open, as required. □

Video (Lemma 5.19 to Proposition 5.20)

Lemma 5.19. *Let $f\colon X \to Y$ be a continuous map between topological spaces. Suppose that $a, b \in X$ with $a \sim b$, so that $[a] = [b]$ in $\pi_0(X)$. Then we also have $f(a) \sim f(b)$, and so $[f(a)] = [f(b)]$ in $\pi_0(Y)$.*

Proof. We are assuming that $a \sim b$, which means that there exists a continuous function $u\colon [0,1] \to X$ with $u(0) = a$ and $u(1) = b$. Put $v = f \circ u$, so v is a continuous function from $[0,1]$ to Y. It has $v(0) = f(u(0)) = f(a)$ and $v(1) = f(u(1)) = f(b)$, so $v\colon f(a) \rightsquigarrow f(b)$. As there exists a path in Y from $f(a)$ to $f(b)$, we have $f(a) \sim f(b)$ as claimed. □

Proposition 5.20. *Let $f\colon X \to Y$ be a continuous map between topological spaces. Then there is a well-defined map $f_*\colon \pi_0(X) \to \pi_0(Y)$, given by $f_*[a] = [f(a)]$ for all $a \in X$. Moreover:*

- *For the identity map* $\mathrm{id}\colon X \to X$, *we have* $\mathrm{id}_* = \mathrm{id}\colon \pi_0(X) \to \pi_0(X)$.
- *For any pair of continuous maps $f\colon X \to Y$ and $g\colon Y \to Z$, we have $(g \circ f)_* = g_* \circ f_*\colon \pi_0(X) \to \pi_0(Z)$.*

Proof. Let u be an element of $\pi_0(X)$; we need to define $f_*(u)$. We can choose a point $a \in X$ such that $u = [a]$, and we want to define $f_*(u) = [f(a)]$. The only problem with this is that it seems to depend on the choice of a. If we chose a different element b with $u = [b]$, then we would also want to define $f_*(u) = [f(b)]$, and that would be inconsistent if $[f(a)]$ was different from $[f(b)]$. However, Lemma 5.19 tells us that $[f(a)] = [f(b)]$, so this problem does not arise, and we have a well-defined function as claimed. We also have $\mathrm{id}_*[a] = [\mathrm{id}(a)] = [a]$, so id_* is the identity map. Similarly, if $g\colon Y \to Z$ is another continuous map, we have

$$g_*(f_*([a])) = g_*[f(a)] = [g(f(a))] = [(g \circ f)(a)] = (g \circ f)_*[a],$$

so $g_* \circ f_* = (g \circ f)_*$. □

6. Interlude on categories and functors

Video (Definition 6.1 to Example 6.6)

Definition 6.1. A *category* $\mathcal{C}$ consists of

(a) A class $\mathrm{obj}(\mathcal{C})$ of mathematical objects (such as groups, rings or metric spaces).
(b) For each pair of objects $A, B \in \mathrm{obj}(\mathcal{C})$, a set $\mathcal{C}(A, B)$ of *morphisms* from A to B. We will write $f\colon A \to B$ or $A \xrightarrow{f} B$ to indicate that $f \in \mathcal{C}(A, B)$.
(c) For each object $A \in \mathrm{obj}(\mathcal{C})$, a morphism $\mathrm{id}_A \in \mathcal{C}(A, A)$ (called the *identity morphism*).
(d) A composition rule for morphisms. This should define, for every pair of morphisms $A \xrightarrow{f} B \xrightarrow{g} C$, a new morphism $g \circ f\colon A \to C$.

These must satisfy the following properties:

(e) For every morphism $f\colon A \to B$, we have $f \circ \mathrm{id}_A = f = \mathrm{id}_B \circ f$.
(f) For every triple of morphisms $A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D$, we have $h \circ (g \circ f) = (h \circ g) \circ f$.

Example 6.2. There is a category Group of groups. The objects are groups, and the morphisms are group homomorphisms. The identity morphism id_G is just the identity function $G \to G$, which is a group homomorphism by a trivial argument. The composition rule is just ordinary composition of functions. For this to be valid, we need to check that the composite of any two group homomorphisms is another group homomorphism, but this is easy. Properties (e) and (f) are also straightforward.

Remark 6.3. (a) Most of our other examples will have the same nature: the objects will be sets with some kind of added structure, and the morphisms will be functions that preserve that structure in some sense. The only point of any real content will be to check that if we compose two structure-preserving maps, then the composite also preserves structure in the same sense.

(b) Two examples considered later will be a bit different: the category of topological spaces and homotopy classes of continuous functions, and the category of chain complexes and chain-homotopy classes of chain maps. In these examples we start with a category $\mathcal{C}$ as in (a), and define an equivalence relation on each morphism set $\mathcal{C}(A, B)$, and put $\overline{\mathcal{C}}(A, B) = \mathcal{C}(A, B)/\sim$. We then want to say that we have a new category with the same objects as $\mathcal{C}$ and morphism sets $\overline{\mathcal{C}}(A, B)$. For this to work, we need to check that the equivalence relations are compatible with composition in an appropriate sense. See Definition 9.9 for an example of this.

Example 6.4. Some other algebraic categories:

(a) The category Ab of abelian groups and group homomorphisms. In abelian groups, we will always write the group operation as addition, the identity element as 0 and the inverse of a as $-a$. Thus, the morphisms from A to B are functions $\alpha\colon A \to B$ satisfying $\alpha(a_0 + a_1) = \alpha(a_0) + \alpha(a_1)$ for all $a_0, a_1 \in A$. (This implies $\alpha(0) = 0$ and $\alpha(-a) = -\alpha(a)$, by standard arguments.)
(b) The category Ring of rings and ring homomorphisms.
(c) The categories $\mathrm{Vect}_{\mathbb{Q}}$, $\mathrm{Vect}_{\mathbb{R}}$ and $\mathrm{Vect}_{\mathbb{C}}$ of vector spaces over $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ respectively. In each case, the morphisms are maps that are linear over the relevant field.
(d) Underlying all of these, we have the category Set: the objects are just sets, and the morphisms are just functions.
(e) We can also consider subcategories defined by various finiteness conditions. For example, we have the category FAb: the objects are finite abelian groups, and the morphisms are just group homomorphisms between finite abelian groups.

In some sense, the main project of algebraic topology is to compare these algebraic categories with various topological categories.

Example 6.5. We have a category Metric: the objects are metric spaces, and the morphisms are continuous maps. To validate this, we need to know that composites of continuous maps are continuous, which is Proposition 3.24 above, or [MS, page 54]. It is sometimes also useful to consider a slightly different category Metric_1: the objects are again metric spaces, but the morphisms are

$$\mathrm{Metric}_1(X, Y) = \{f\colon X \to Y \mid d(f(x), f(x')) \leq d(x, x') \text{ for all } x, x' \in X\}.$$

It is straightforward to check that if $f \in \mathrm{Metric}_1(X, Y)$ and $g \in \mathrm{Metric}_1(Y, Z)$ then $g \circ f \in \mathrm{Metric}_1(X, Z)$, so this definition does indeed give a category. Lemma 3.5 tells us that $\mathrm{Metric}_1(X, Y) \subseteq \mathrm{Metric}(X, Y)$.

Example 6.6. Similarly, there is a category Top whose objects are topological spaces, and whose morphisms are continuous maps.

We next need to discuss how to compare different categories. The key concept here is as follows:

Video (Definition 6.7 to Example 6.11)

Definition 6.7. Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A *functor* F from $\mathcal{C}$ to $\mathcal{D}$ consists of

(a) A rule giving an object $FA \in \mathrm{obj}(\mathcal{D})$ for every object $A \in \mathrm{obj}(\mathcal{C})$; and
(b) A rule giving a morphism $Ff \in \mathcal{D}(FA, FA')$ for each morphism $f \in \mathcal{C}(A, A')$

such that

(c) For each $A \in \mathrm{obj}(\mathcal{C})$, we have $F(\mathrm{id}_A) = \mathrm{id}_{FA}$
(d) For every pair of morphisms $A \xrightarrow{f} A' \xrightarrow{f'} A''$ in $\mathcal{C}$, we have $F(f' \circ f) = F(f') \circ F(f) \in \mathcal{D}(FA, FA'')$. In other words, the following diagram should commute:

$$\begin{tikzcd} FA \arrow[r, "Ff"] \arrow[dr, "F(f'\circ f)"'] & FA' \arrow[d, "Ff'"] \\ & FA'' \end{tikzcd}$$

We will sometimes write f_* (or $f_\#$ or $f_\bullet$ or some similar notation) instead of Ff.

Example 6.8. For every topological space $X \in \text{obj}(\text{Top})$, we have a set $\pi_0(X) \in \text{obj}(\text{Set})$, as in Definition 5.10. For every continuous map $f \in \text{Top}(X,Y)$, we have a function $f_* = \pi_0(f) \in \text{Set}(\pi_0(X), \pi_0(Y))$, as in Proposition 5.20. The same proposition proves conditions (c) and (d) in Definition 6.7, so we have a functor $\pi_0 \colon \text{Top} \to \text{Set}$.

Example 6.9. For any metric space X, we let TX denote the same set regarded as a topological space using the metric topology. If $f \in \text{Metric}(X,Y)$ then f is just a continuous map from X to Y and so can also be regarded as an element of $\text{Top}(TX, TY)$. We define $Tf = f$. This gives a functor $T \colon \text{Metric} \to \text{Top}$.

Example 6.10. For any set X, we have another set $PX = X \times \mathbb{Z}$. We would like to make this construction into a functor $P \colon \text{Set} \to \text{Set}$. Given a morphism $f \in \text{Set}(X,Y)$ (i.e. a function $f \colon X \to Y$), we need to define a corresponding function $Pf \colon PX \to PY$, or in other words $Pf \colon X \times \mathbb{Z} \to Y \times \mathbb{Z}$. Thus, given a function $f \colon X \to Y$, a point $x \in X$ and a number $n \in \mathbb{Z}$, we need to define a point $(Pf)(x,n) \in Y \times \mathbb{Z}$. This must have the form $(Pf)(x,n) = (y,m)$ for some $y \in Y$ and $m \in \mathbb{Z}$. The only element of Y that we can produce from these ingredients is $f(x)$, so we need to take $y = f(x)$. We must have $P(\text{id}_X) = \text{id} \colon X \times \mathbb{Z} \to X \times \mathbb{Z}$, so in the case $f = \text{id}_X$ we need to take $m = n$, so the simplest thing is to take $m = n$ in all cases. We therefore arrive at the definition $(Pf)(x,n) = (f(x), n)$, or in other words $Pf = f \times \text{id}_{\mathbb{Z}} \colon X \times \mathbb{Z} \to Y \times \mathbb{Z}$. This gives

$$P(\text{id}_X)(x,n) = (\text{id}_X(x), n) = (x,n) = \text{id}_{PX}(x,n).$$

Also, if we have functions $X \xrightarrow{f} Y \xrightarrow{g} Z$, then

$$P(g)(P(f)(x,n)) = P(g)(f(x),n) = (g(f(x)),n) = ((g \circ f)(x), n) = P(g \circ f)(x,n).$$

Thus, we have $P(\text{id}_X) = \text{id}_{PX}$ and $P(g \circ f) = P(g) \circ P(f)$. We have therefore succeeded in defining a functor.

Example 6.11. For any abelian group A we note that the subset $DA = \{2a \mid a \in A\}$ is a subgroup of A. If $\alpha \colon A \to B$ is a homomorphism, then $\alpha(2a) = 2\alpha(a)$ for all $a \in A$, so $\alpha(DA) \leq DB$, so we have a function $D\alpha = \alpha|_{DA} \colon DA \to DB$, which is again a homomorphism. It is easy to see that $D(\text{id}_A) = \text{id}_{DA}$ and $D(\beta \circ \alpha) = (D\beta) \circ (D\alpha)$, so we have defined a functor $D \colon \text{Ab} \to \text{Ab}$. This satisfies $D(\mathbb{Z}/2) = 0$ and $D(\mathbb{Z}/3) = \mathbb{Z}/3$, for example. We can also define $QA = A/DA$. Given a coset $a + DA \in QA$, we would like to define $(Q\alpha)(a + DA) = \alpha(a) + DB \in QB$. This is well-defined, because if $a + DA = a' + DA$ then $a' - a \in DA$, so $\alpha(a') - \alpha(a) = \alpha(a' - a) \in \alpha(DA) \leq DB$, so $\alpha(a') + DB = \alpha(a) + DB$. One can check that this gives a homomorphism $Q\alpha \colon QA \to QB$, and that $Q(\text{id}_A) = \text{id}_{QA}$ and $Q(\beta \circ \alpha) = (Q\beta) \circ (Q\alpha)$. Thus, we have defined another functor $Q \colon \text{Ab} \to \text{Ab}$. This satisfies $Q(\mathbb{Z}/2) \simeq \mathbb{Z}/2$ and $Q(\mathbb{Z}/3) = 0$, for example.

Video (Definition 6.12 to Corollary 6.18)

Definition 6.12. Let $\mathcal{C}$ be a category, and let $f \colon X \to Y$ be a morphism in $\mathcal{C}$. An *inverse* for f is a morphism $g \colon Y \to X$ such that $g \circ f = \text{id}_X$ and $f \circ g = \text{id}_Y$. We say that f is an *isomorphism* if it has an inverse. We say that X and Y are *isomorphic* if there exists an isomorphism from X to Y. We will usually write $X \simeq Y$ to indicate that X and Y are isomorphic.

Lemma 6.13. *If f has an inverse, then it is unique.*

Proof. Let g_1 and g_2 be inverses for f. Then

$$g_1 = g_1 \circ \text{id}_Y = g_1 \circ (f \circ g_2) = (g_1 \circ f) \circ g_2 = \text{id}_X \circ g_2 = g_2.$$

∎

Because of the lemma, we can write f^{-1} for the inverse of f without creating any ambiguity.

Example 6.14. In the category Group, the isomorphisms are just group isomorphisms as usually defined in abstract algebra. In the category Set, the isomorphisms are just bijections. In the categories Metric and Top, the isomorphisms are homeomorphisms.

Proposition 6.15.

(a) *For every object $X \in \mathcal{C}$, the identity morphism* id_X *is an isomorphism.*
(b) *If $f\colon X \to Y$ is an isomorphism, then so is $f^{-1}\colon Y \to X$.*
(c) *If $f\colon X \to Y$ and $g\colon Y \to Z$ are isomorphisms, then so is $g \circ f\colon X \to Z$.*

Proof.

(a) id_X is an inverse for itself.
(b) f is an inverse for f^{-1}.
(c) $f^{-1} \circ g^{-1}$ is an inverse for $g \circ f$.

□

Corollary 6.16. *Let X, Y and Z be objects in a category $\mathcal{C}$.*

(a) $X \simeq X$
(b) *If $X \simeq Y$ then $Y \simeq X$*
(c) *If $X \simeq Y$ and $Y \simeq Z$ then $X \simeq Z$.*

Proof. Immediate from the proposition. □

Proposition 6.17. *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor, and let $f\colon X \to Y$ be an isomorphism in $\mathcal{C}$. Then the morphism $Ff\colon FX \to FY$ is an isomorphism in $\mathcal{D}$, with inverse $F(f^{-1})$.*

Proof. By the definition of an inverse, we have $f^{-1} \circ f = \mathrm{id}_X$ and $f \circ f^{-1} = \mathrm{id}_Y$. Using the functor axioms we obtain

$$F(f^{-1}) \circ Ff = F(f^{-1} \circ f) = F(\mathrm{id}_X) = \mathrm{id}_{FX}$$
$$Ff \circ F(f^{-1}) = F(f \circ f^{-1}) = F(\mathrm{id}_Y) = \mathrm{id}_{FY}\,.$$

These prove that $F(f^{-1})$ is an inverse for Ff, as required. □

Corollary 6.18. *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor, and let X and Y be objects of $\mathcal{C}$. If $X \simeq Y$ in $\mathcal{C}$, then $FX \simeq FY$ in $\mathcal{D}$.*

Proof. Immediate from the proposition. □

Video (Definition 6.19 to Proposition 6.26)

We will also need a weaker concept, which is only half as good as being isomorphic.

Definition 6.19. Let X and Y be objects in a category $\mathcal{C}$. We say that X is a *retract* of Y if there exist morphisms $X \xrightarrow{f} Y \xrightarrow{g} X$ with $g \circ f = \mathrm{id}_X$. (We make no assumption about $f \circ g$.) Any pair (f, g) with this property will be called a *retraction pair* for (X, Y).

Example 6.20. Let G and H be groups. We can define homomorphisms

$$G \xrightarrow{j} G \times H \xrightarrow{q} G$$

by $j(g) = (g, 1)$ and $q(g, h) = g$. These satisfy $q \circ j = \mathrm{id}_G$, so G is a retract of $G \times H$ in Group.

Example 6.21. We can define continuous maps $S^2 \xrightarrow{f} \mathbb{R}^3 \setminus \{0\} \xrightarrow{g} S^2$ by $f(u) = u$ and $g(v) = v/\|v\|$. These satisfy $g \circ f = \mathrm{id}_{S^2}$, so S^2 is a retract of $\mathbb{R}^3 \setminus \{0\}$ in Top.

Example 6.22. Let X and Y be nonempty finite sets with $|X| \leq |Y|$. We claim that X is a retract of Y in the category Set. Indeed, we can list the elements as $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_m\}$ with $1 \leq n \leq m$. We can then define $f\colon X \to Y$ by $f(x_i) = y_i$. In the opposite direction, we define $g\colon Y \to X$ by

$$g(y_i) = \begin{cases} x_i & \text{if } 1 \leq i \leq n \\ x_n & \text{if } n < i \leq m. \end{cases}$$

We then have $g \circ f = \mathrm{id}_X$, as required.

We will mostly be interested in cases where we can prove that X is *not* a retract of Y. The main tool for this is as follows:

Proposition 6.23. *Let $\mathcal{C}$ be a category in which the objects are sets with extra structure, and the morphisms are the functions that preserve that structure. (This covers all the examples that we have discussed so far.) Let (f, g) be a retraction pair in $\mathcal{C}$. Then f is injective, and g is surjective.*

Proof. We can ignore the structure-preserving properties of f and g; we just need to know that we have functions $f\colon X \to Y$ and $g\colon Y \to X$ satisfying $g(f(x)) = x$ for all $x \in X$. In particular, this shows that x can be written as $g(y)$ for some y (namely $y = f(x)$), so g is surjective. Now suppose we have $x, x' \in X$ with $f(x) = f(x')$. By applying g to both sides we obtain $g(f(x)) = g(f(x'))$, but $g(f(x)) = x$ and $g(f(x')) = x'$ so we get $x = x'$. This proves that f is injective. □

Corollary 6.24. *Let G and H be groups.*

(a) *If G is nonabelian and H is abelian, then G is not a retract of H.*
(b) *If G is infinite and H is finite, then G is not a retract of H.*
(c) *If $G \simeq \mathbb{Z}/2$ and $H \simeq \mathbb{Z}$, then G is not a retract of H.*

Proof. Suppose that $G \xrightarrow{j} H \xrightarrow{q} G$ is a retraction pair. Then j is injective, so G is isomorphic to $j(G)$, which is a subgroup of H. By the contrapositive, if G is not isomorphic to any subgroup of H, then G cannot be a retract of H. In case (a), every subgroup of H is abelian, so G cannot be isomorphic to any subgroup of H. In case (b), every subgroup of H is finite, so G cannot be isomorphic to any subgroup of H. In case (c), the group G contains an element of order precisely two, but all elements of $\mathbb{Z}$ have order 1 or ∞, so again G cannot be isomorphic to any subgroup of H. □

Proposition 6.25. *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor, and let X and Y be objects of $\mathcal{C}$. If X is a retract of Y in $\mathcal{C}$, then FX is a retract of FY in $\mathcal{D}$. Thus, by the contrapositive, if FX is not a retract of FY, then X cannot be a retract of Y.*

Proof. If X is a retract of Y, then we can choose a retraction pair $X \xrightarrow{f} Y \xrightarrow{g} X$ with $g \circ f = \mathrm{id}_X$. This gives maps $FX \xrightarrow{Ff} FY \xrightarrow{Fg} FX$ with $Fg \circ Ff = F(g \circ f) = F(\mathrm{id}_X) = \mathrm{id}_{FX}$, proving that FX is a retract of FY. □

As a basic example of how this can be used, we have the following:

Proposition 6.26. *Let X and Y be topological spaces such that $|\pi_0(Y)|$ is finite and $|\pi_0(X)| > |\pi_0(Y)|$. Then X is not a retract of Y.*

Proof. If X was a retract of Y, then $\pi_0(X)$ would be a retract of $\pi_0(Y)$, so in particular, we would have an injective function $f_*\colon \pi_0(X) \to \pi_0(Y)$. This is impossible because $|\pi_0(X)| > |\pi_0(Y)|$. □

7. Constructing new spaces

We next need to discuss several ways of constructing new topological spaces from spaces that we already know about.

Video (Definition 7.1 to Remark 7.4)

Definition 7.1. Let Y and Z be disjoint sets, and put $X = Y \cup Z$. Suppose we are given topologies on Y and Z, which we use to regard them as topological spaces. We then declare that a subset $U \subseteq X$ is open iff $U \cap Y$ is open in the given topology on Y, and $U \cap Z$ is open in the given topology on Z. This is easily seen to give a topology on X, which we call the *coproduct topology*.

Remark 7.2. We now have a topology on X, and Y is a subset of X, so we can use Definition 3.25 to define a subspace topology on Y. It is easy to see that this is just the the same as the original topology on Y. Similarly, if we regard Z as a subspace of X, then the subspace topology is just the same as the original topology. In particular, the inclusion maps $Y \xrightarrow{i} X \xleftarrow{j} Z$ are continuous.

The category-theoretic viewpoint encourages us to ask the following kind of question: whenever we construct a new object X in a category $\mathcal{C}$, we should try to prove theorems describing the set $\mathcal{C}(T,X)$ of morphisms into X, and/or the set $\mathcal{C}(X,T)$ of morphisms out of X, for an arbitrary object $T \in \mathcal{C}$.

Proposition 7.3. *Let X, Y and Z be as above, and let T be another topological space. To describe a function $f\colon X \to T$, it is enough to specify the restricted functions $g = f|_Y\colon Y \to T$ and $h = f|_Z\colon Z \to T$. Moreover, the combined map f is continuous (with respect to the coproduct topology) iff g and h are continuous (with respect to the originally given topologies on Y and Z). Thus, we have a bijection between $\mathrm{Top}(X,T)$ and $\mathrm{Top}(Y,T) \times \mathrm{Top}(Z,T)$.*

Proof. Consider an open set $A \subseteq T$. We then find that $f^{-1}(A) \cap Y = g^{-1}(A)$ and $f^{-1}(A) \cap Z = h^{-1}(A)$. Thus, $f^{-1}(A)$ is open in X iff $g^{-1}(A)$ is open in Y and $h^{-1}(A)$ is open in Z. Thus, f is continuous iff this condition holds for all A iff g and h are continuous. Thus, to give a continuous map $f \in \mathrm{Top}(X,T)$ is the same as to give a pair of continuous maps $g \in \mathrm{Top}(Y,T)$ and $h \in \mathrm{Top}(Z,T)$, or in other words a pair $(g,h) \in \mathrm{Top}(Y,T) \times \mathrm{Top}(Z,T)$. □

Remark 7.4. There is a general notion of coproduct objects in category theory. The above proposition can be interpreted as saying that X (equipped with the coproduct topology) is the coproduct of Y and Z in this general categorical sense.

Proposition 7.5. *Let X, Y, Z and T be as above, and consider a function $p\colon T \to X$. Then p is continuous iff*

(a) *The sets $A = p^{-1}(Y)$ and $B = p^{-1}(Z)$ are open in T.*
(b) *The restricted map $q = p|_A\colon A \to Y$ is continuous with respect to the subspace topology on $A \subseteq T$.*
(c) *The restricted map $r = p|_B\colon B \to Z$ is continuous with respect to the subspace topology on $B \subseteq T$.*

Proof. Left as an exercise. □

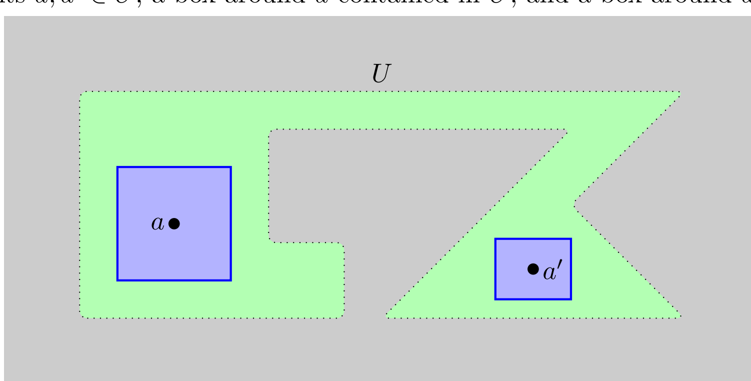
Video (Definition 7.6 to Remark 7.12)

Definition 7.6. Let Y and Z be topological spaces, and consider the product

$$X = Y \times Z = \{(y,z) \mid y \in Y \text{ and } z \in Z\}.$$

For a point $a = (b,c) \in Y \times Z$, a *box around a* means a set of the form $V \times W$, where $b \in V$ and $c \in W$ and V is open in Y and W is open in Z. We declare that a set $U \subseteq X$ is open iff for all $a \in U$, there is a box around a that is contained in U.

Example 7.7. The picture shows the case where $Y = [0,2]$ and $Z = [0,1]$. We have indicated a subset $U \subset Y \times Z$, two points $a, a' \in U$, a box around a contained in U, and a box around a' contained in U.



$X = Y \times Z$

Lemma 7.8. *Suppose that V is an open subset of Y and W is an open subset of Z. Then $V \times W$ is open in $Y \times Z$.*

Proof. Consider a point $a = (b,c) \in V \times W$, so $b \in V$ and $c \in W$. We need to show that there is a box B around a such that $B \subseteq V \times W$. We can just take $B = V \times W$. □

Proposition 7.9. *Definition 7.6 specifies a topology on $Y \times Z$ (which we call the* product topology*).*

Proof. First suppose we have a family of open sets $U_i \subseteq Y \times Z$; we must show that the union $U^* = \bigcup_i U_i$ is open. Consider a point $a \in U^*$. By the definition of the union, this means that $a \in U_i$ for some i. As U_i is open, we can choose a box $B = V \times W$ around a such that $B \subseteq U_i$. Now $U_i \subseteq U^*$, so we also have $B \subseteq U^*$. Thus, every point in U^* has a box that is contained in U^*, so U^* is open as required.

Now suppose we have a finite list of open sets $U_1, \dots, U_n$; we must show that the intersection $U^\# = U_1 \cap \dots \cap U_n$ is open. Consider a point $a = (b, c) \in U^\#$. By the definition of the intersection, we have $a \in U_i$ for all i. As U_i is open, we can choose a box $B_i = V_i \times W_i$ around a such that $B_i \subseteq U_i$. Put $V^\# = V_1 \cap \dots \cap V_n$ and $W^\# = W_1 \cap \dots \cap W_n$ and $B^\# = B_1 \cap \dots \cap B_n = V^\# \times W^\#$. For all i we have $b \in V_i$ and $c \in W_i$, so $b \in V^\#$ and $c \in W^\#$. By the topology axioms for Y, the set $V^\#$ is open in Y. By the topology axioms for Z, the set $W^\#$ is open in Z. It follows that $B^\#$ is a box around a that is contained in $U^\#$, as required.

We also see that $\emptyset$ is open (because there are no points $a \in \emptyset$ to check, so the definition is vacuously satisfied). Also, the set $X = Y \times Z$ itself is a box around each of its points, so X is open. This completes the proof that we have defined a topology. □

Again, the categorical viewpoint encourages us to try to analyse the continuous maps to or from $Y \times Z$. For maps out of $Y \times Z$ there is no simple answer, but maps into $Y \times Z$ are quite easy.

Lemma 7.10. *The projection maps $Y \xleftarrow{p} Y \times Z \xrightarrow{q} Z$ (given by $p(y, z) = y$ and $q(y, z) = z$) are continuous.*

Proof. Let $V \subseteq Y$ be an open set. We must show that $p^{-1}(V)$ is open in $Y \times Z$, but $p^{-1}(V)$ is just the same as $V \times Z$, which is open by Lemma 7.8. The proof for q is essentially the same. □

Now consider a topological space T and a function $f \colon T \to Y \times Z$. This must have the form $f(t) = (g(t), h(t))$ for some functions $g \colon T \to Y$ and $h \colon T \to Z$. We denote this by $f = \langle g, h \rangle$. We can also express the component functions g and h as $g = p \circ f$ and $h = q \circ f$.

Proposition 7.11. *The combined map $f = \langle g, h \rangle \colon T \to Y \times Z$ is continuous (with respect to the product topology) iff the component functions $g = p \circ f \colon T \to Y$ and $h = q \circ f \colon T \to Z$ are continuous. Thus, there is a bijection between the sets $\mathrm{Top}(T, Y \times Z)$ and $\mathrm{Top}(T, Y) \times \mathrm{Top}(T, Z)$, in which the element $f \in \mathrm{Top}(T, Y \times Z)$ corresponds to the pair $(g, h) \in \mathrm{Top}(T, Y) \times \mathrm{Top}(T, Z)$.*

Proof. If f is continuous, then the composites $g = p \circ f$ and $h = q \circ f$ are continuous by Lemma 7.10 and Proposition 3.24.

Suppose instead that we start from the assumption that g and h are both continuous. Consider an open set $U \subseteq Y \times Z$; we must show that $f^{-1}(U)$ is open. The simplest case is where U is a box, say $U = V \times W$, where V is open in Y and W is open in Z. We then have

$$t \in f^{-1}(V \times W) \iff (g(t), h(t)) \in V \times W \iff (g(t) \in V \text{ and } h(t) \in W) \iff t \in g^{-1}(V) \cap h^{-1}(W).$$

This shows that $f^{-1}(V \times W) = g^{-1}(V) \cap h^{-1}(W)$, and this set is open because g and h are continuous.

Now return to the general case of an arbitrary open set $U \subseteq Y \times Z$, which need not be a box. For each point $x \in f^{-1}(U)$, we have $f(x) \in U$. As U is open with respect to the product topology, there is a box B_x with $f(x) \in B_x \subseteq U$. We put $A_x = f^{-1}(B_x)$, so $x \in A_x \subseteq f^{-1}(U)$. The previous paragraph shows that A_x is open. Moreover, we find that $f^{-1}(U)$ is the union of all the sets A_x. The union of any family of open sets is open, so $f^{-1}(U)$ is open as required. □

Remark 7.12. There is a general notion of product objects in category theory. The above proposition can be interpreted as saying that $Y \times Z$ (equipped with the product topology) is the product of Y and Z in this general categorical sense.

We now turn our attention to quotient constructions. We recall the basic definitions.

Video (Definition 7.13 to Corollary 7.16)

Definition 7.13. Let X be a set, and let E be a relation on X (so for each pair $(x, y) \in X \times X$ we have a statement xEy, which may or may not be satisfied).

(a) We say that the relation is *reflexive* if xEx for all $x \in X$.

(b) We say that the relation is *symmetric* if whenever xEy, we also have yEx.
(c) We say that the relation is *transitive* if whenever xEy and yEz, we also have xEz.
(d) We say that the relation is an *equivalence relation* if it is reflexive, symmetric and transitive.
(e) If we have an equivalence relation, we write $[x] = \{y \mid xEy\} \subseteq X$, and call this the *equivalence class* of x. We note that
   (i) The relation xEy holds iff $[x] = [y]$
   (ii) For each $x \in X$ we have $x \in [x]$
   (iii) If $x \not\!E y$, then the subsets $[x], [y] \subseteq X$ are disjoint, i.e. $[x] \cap [y] = \emptyset$.

   Although the equivalence classes are defined as subsets of X, we will often deemphasise that fact, and just treat them as abstract symbols satisfying property (i).
(f) We write X/E for the set of all equivalence classes, so $X/E = \{[x] \mid x \in X\}$. We define a map $\pi\colon X \to X/E$ by $\pi(x) = [x]$, so π is surjective and $\pi(x) = \pi(y)$ iff xEy. Informally, we say that X/E is obtained from X by identifying points x and y whenever xEy. We may also say "gluing together" instead of "identifying".

Definition 7.14. Suppose we have sets X and Y, and an equivalence relation E on X. We say that a function $f\colon X \to Y$ is *E-saturated* if whenever xEx' in X, we have $f(x) = f(x')$ in Y.

Proposition 7.15.
(a) *The quotient map $\pi\colon X \to X/E$ is E-saturated.*
(b) *For any function $\overline{f}\colon X/E \to Y$, the composite $\overline{f} \circ \pi$ is E-saturated.*
(c) *For any E-saturated function $f\colon X \to Y$, there is a well-defined function $\overline{f}\colon X/E \to Y$ given by $\overline{f}([x]) = f(x)$. This is the unique $\overline{f}\colon X/E \to Y$ such that $\overline{f} \circ \pi = f$. (We call it the* function induced by *f.)*

Proof. Claim (a) follows from Definition 7.13(e)(i), and claim (b) follows immediately from (a). For claim (c), we need to define $\overline{f}(u) \in Y$ for each equivalence class $u \in X/E$. To do this, we pick any x with $u = [x]$, and take $\overline{f}(u) = f(x)$. There could in principle be a problem with this. If we choose a different x' such that u is also equal to $[x']$, then we should also have $\overline{f}(u) = f(x')$, and this would be inconsistent if $f(x')$ was different from $f(x)$. However, in this situation we have $[x] = [x']$ so xEx' so $f(x) = f(x')$ by the E-saturation condition. Thus, no inconsistency can arise, and we have a well-defined function as claimed. The equation $\overline{f} \circ \pi = f$ is just another way of writing the condition $\overline{f}([x]) = f(x)$, so it is clear that there is a unique function $\overline{f}$ with this property. □

Corollary 7.16. *There is a one-to-one correspondence between functions $X/E \to Y$, and E-saturated functions $X \to Y$.* □

Video (Definition 7.17 to Remark 7.21)

Definition 7.17. Now suppose we have a topological space X, together with an equivalence relation E on X. We declare that a subset $V \subseteq X/E$ is open iff the preimage $\pi^{-1}(V) \subseteq X$ is open in X.

Proposition 7.18. *The above definition gives a topology on X/E (which we call the* quotient topology*). Moreover, the quotient map $\pi\colon X \to X/E$ is continuous with respect to this topology.*

Proof. Suppose we have a family of subsets $V_i \subseteq X/E$ that are open with respect to the quotient topology. We must show that the union $V^* = \bigcup_i V_i$ is also open. By the definition of the quotient topology, we see that the sets $U_i = \pi^{-1}(V_i)$ are open in X, and we must show that $\pi^{-1}(V^*)$ is also open in X. As the sets U_i are all open, the set $U^* = \bigcup_i U_i$ is also open, by the topology axioms for X. Moreover, we have $\pi^{-1}(V^*) = U^*$ by Lemma 3.11(a), so $\pi^{-1}(V^*)$ is open as required. The proof for finite intersections is similar. We also have $\pi^{-1}(X/E) = X$, and X is open in X by the topology axioms for X, so X/E is open in X/E by our definition of open sets in X/E. Similarly, we have $\pi^{-1}(\emptyset) = \emptyset$, which is open in X, so $\emptyset$ is open in X/E. This completes the proof that we have a topology on X/E.

We also want to prove that $\pi\colon X \to X/E$ is continuous. In other words, for every set $V \subseteq X/E$ that is open with respect to the quotient topology, we must show that $\pi^{-1}(V)$ is open in X. But this is true by the very definition of the quotient topology. □

Proposition 7.19. *Let X and E be as above, and let $\overline{f}$ be a function from X/E to another topological space Y. Then $\overline{f}$ is continuous (with respect to the quotient topology on X/E) iff the composite $\overline{f} \circ \pi \colon X \to Y$ is continuous.*

Proof. We have seen that π is continuous and that composites of continuous functions are continuous. Thus, if $\overline{f}$ is continuous, then the composite $\overline{f} \circ \pi$ is also continuous as claimed.

Suppose instead that we start from the assumption that $\overline{f} \circ \pi \colon X \to Y$ is continuous; we must show that $\overline{f} \colon X/E \to Y$ is continuous. Let $W \subseteq Y$ be open; we must show that the preimage $V = \overline{f}^{-1}(W)$ is open in X/E. By the definition of the quotient topology, it is equivalent to prove that the set $\pi^{-1}(V)$ is open in X. However, we have $\pi^{-1}(V) = \pi^{-1}(\overline{f}^{-1}(W)) = (\overline{f} \circ \pi)^{-1}(W)$, and this is open as required because $\overline{f} \circ \pi$ is assumed to be continuous. $\square$

Corollary 7.20. *Let $f \colon X \to Y$ be an E-saturated function. Then the induced function $\overline{f} \colon X/E \to Y$ is continuous iff f is continuous. Thus, we have a one-to-one correspondence between continuous functions $X/E \to Y$, and E-saturated continuous functions $X \to Y$.*

Proof. As $\overline{f} \circ \pi = f$, this is just a restatement of the proposition. $\square$

Remark 7.21. This corollary could be put in a common framework with the first isomorphism theorem for groups, if we took the time to develop the relevant categorical notion of coequalisers.

Example 7.22. We can define an equivalence relation E on S^n by xEy iff $y = \pm x$. The real projective space $\mathbb{R}P^n$ is defined to be the quotient space S^n/E. Similarly, we can regard S^{2n+1} as the unit sphere in the space $\mathbb{C}^{n+1} = \mathbb{R}^{2n+2}$. We can then define another equivalence relation F on S^{2n+1} by declaring that xFy iff $y = e^{i\theta}x$ for some $\theta \in \mathbb{R}$. The complex projective space $\mathbb{C}P^n$ is defined to be S^{2n+1}/F.

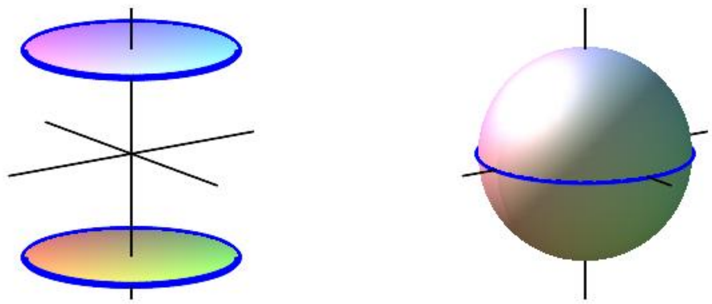
Example 7.23. We will prove that the real projective space $\mathbb{R}P^1$ is homeomorphic to S^1. Recall that $\mathbb{R}P^1 = S^1/E$, where uEv iff $v = \pm u$. We will identify S^1 with $\{z \in \mathbb{C} \mid |z| = 1\}$ by the usual Argand corrspondence $(x, y) \leftrightarrow x + iy$. We define $f \colon S^1 \to S^1$ by $f(z) = z^2$. This clearly satisfies $f(-z) = f(z)$, so it is E-saturated and induces a map $\overline{f} \colon \mathbb{R}P^1 \to S^1$ with $\overline{f}(\pi(z)) = z^2$ for all z. Now suppose we have a point $w = e^{i\theta} \in S^1$. Then the set $g(w) = \{e^{i\theta/2}, -e^{i\theta/2}\}$ is an E-equivalence class, or in other words a point of the space $\mathbb{R}P^1$, so we have a map $g \colon S^1 \to \mathbb{R}P^1$. It is easy to see that this is inverse to $\overline{f}$, so $\overline{f}$ is a continuous bijection. We just need to check that g is also continuous. Suppose that $U \subseteq \mathbb{R}P^1$ is open, and contains $g(w)$ for some point $w = e^{i\theta} \in S^1$. This means that $e^{i\theta/2} \in \pi^{-1}(U)$, and $\pi^{-1}(U)$ is open in S^1 by the definition of the quotient topology, so there is some $\epsilon > 0$ such that $e^{i\phi} \in \pi^{-1}(U)$ whenever $|\phi - \theta/2| < \epsilon$. It follows that $e^{i\theta'} \in g^{-1}(U)$ whenever $|\theta' - \theta| < 2\epsilon$. From this we can deduce that $g^{-1}(U)$ is open as required.

Interactive demo

Example 7.24. Put

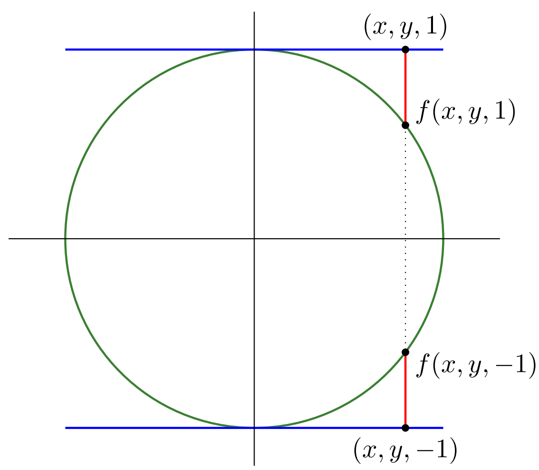
$$X = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 \leq 1 \text{ and } z = \pm 1\}.$$

This consists of two closed discs, as shown on the left below.



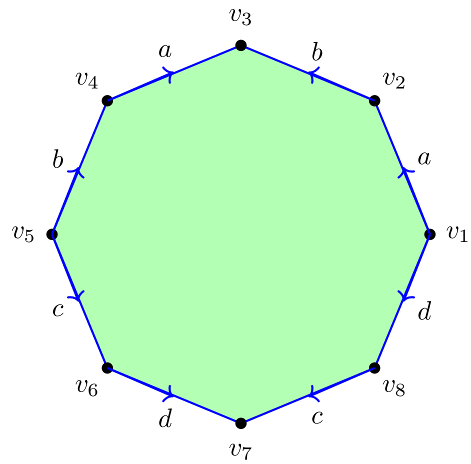
Interactive demo

If we glue together the boundary circles of the two discs, we get a sphere, as shown on the right. We can make this more formal and rigorous as follows. We introduce an equivalence relation E on X by declaring that $(x, y, z)E(x', y', z')$ if $(x', y', z') = (x, y, z)$, or if $(x', y', z') = (x, y, -z)$ with $x^2 + y^2 = 1$. We define $f\colon X \to S^2$ by $f(x, y, z) = (x, y, z\sqrt{1 - x^2 - y^2})$, as illustrated below.



This is clearly continuous. Also, if $x^2 + y^2 = 1$ then we have $f(x, y, 1) = f(x, y, -1) = (x, y, 0)$, so f is E-saturated. We therefore have an induced map $\overline{f}\colon X/E \to S^2$, and Corollary 7.20 tells us that this is continuous. It is not hard to check that $\overline{f}$ is a bijection. However, we saw in Example 4.10 that not every continuous bijection has a continuous inverse, so we cannot immediately conclude that $\overline{f}$ is a homeomorphism. It would be possible but fiddly to prove this directly. However, we will see an efficient general method for this in Section 8, so we will defer further comment until then.

Example 7.25. From the Knots and Surfaces course you should remember pictures like this:



We take a filled octagon X, with edges marked a, b, c or d as shown. There are two edges marked a, and we glue them together in the direction indicated by the arrows. We do the same for the labels b, c and d. To make this more formal and rigorous, we introduce a relation E as follows: we declare that xEy if

- $x = y$; or
- for some $t \in [0,1]$ we have $x = (1-t)v_1 + tv_2$ and $y = (1-t)v_4 + tv_3$; or
- for some $t \in [0,1]$ we have $x = (1-t)v_2 + tv_3$ and $y = (1-t)v_4 + tv_3$; or
- for some $t \in [0,1]$ we have $x = (1-t)v_5 + tv_6$ and $y = (1-t)v_8 + tv_7$; or
- for some $t \in [0,1]$ we have $x = (1-t)v_6 + tv_7$ and $y = (1-t)v_1 + tv_8$; or
- for some i, j we have $x = v_i$ and $y = v_j$.

One can check that this is an equivalence relation, so we can form the quotient space X/E. One can then show that X/E is homeomorphic to the usual kind of double torus embedded in $\mathbb{R}^3$.

Example 7.26. Consider the spaces $X = \mathbb{R}^2$ and

$$Y = S^1 \times S^1 = \{(w, x, y, z) \in \mathbb{R}^4 \mid w^2 + x^2 = y^2 + z^2 = 1\}$$

(so Y is a version of the torus). We introduce an equivalence relation on X by declaring that uEv iff $u - v \in \mathbb{Z}^2$. We then define $f\colon X \to Y$ by

$$f(x, y) = (\cos(2\pi x),\ \sin(2\pi x),\ \cos(2\pi y),\ \sin(2\pi y)).$$

It is easy to see that this is continuous and E-saturated, so it induces a continuous map $\overline{f}\colon X/E \to Y$. It is also easy to see that $\overline{f}$ is a bijection. It is again true that $\overline{f}^{-1}$ is also continuous, so that $\overline{f}$ is a homeomorphism, but we will again defer the proof.

Example 7.27. Here is a more exotic example. We start with the space $X = \mathbb{R} \times \{1, -1\}$, and we declare that $(x', y')E(x, y)$ if either $(x', y') = (x, y)$, or $(x', y') = (x, -y)$ with $x \neq 0$. In other words, we start with the two lines where $y = 1$ and $y = -1$ and then we glue them together, except that we leave the points $(0, 1)$ and $(0, -1)$ unglued. The quotient space X/E is called the *line with doubled origin*. It has various unpleasant properties, and we mostly choose not to study spaces with those properties; that is the point of the Hausdorff condition to be introduced in Section 8.

Remark 7.28. There is one more construction that we would like to treat, but will not. Let X be any metric space. From the Metric Spaces course, you should remember that the set $C(X, \mathbb{R}) = \mathrm{Top}(X, \mathbb{R})$ (of continuous functions from X to $\mathbb{R}$) has a metric of its own, which is useful for many purposes. It would also be useful to define a similar topology on $\mathrm{Top}(X, Y)$ for arbitrary topological spaces X and Y. This seems natural, because it is easy to imagine what we mean by saying that two functions are close to each other, which is the basic idea that we need when defining a topology. Unfortunately this leads to a host of technical difficulties and subtle distinctions. A huge detour is necessary before one can set up a clear and coherent theory. Thus, we will motivate various constructions using the idea of treating $\mathrm{Top}(X, Y)$ as a topological space, but we will not use this idea in our formal definitions.

Proposition 7.29. *Let Y and Z be disjoint spaces, and take $X = Y \cup Z$, with the coproduct topology. Then $\pi_0(X)$ is the disjoint union of $\pi_0(Y)$ and $\pi_0(Z)$.*

Proof. We can define $f\colon X \to \{1, -1\} \subset \mathbb{R}$ by $f(y) = 1$ for $y \in Y$, and $f(z) = -1$ for $z \in Z$. This is clearly continuous when restricted to Y or Z, so it is continuous on all of X by Proposition 7.3. If $u\colon [0,1] \to X$ is a continuous path, then $f \circ u\colon [0,1] \to \{-1, 1\}$ is also continuous, so it clearly must be constant. (Formally, the proof uses the Intermediate Value Theorem.) Thus, the path lies wholly in Y or wholly in Z. Given this, the claim is clear from the definitions. $\square$

Proposition 7.30. *Let Y and Z be topological spaces. Then $\pi_0(Y \times Z)$ can be identified with $\pi_0(Y) \times \pi_0(Z)$. More precisely, let $Y \xleftarrow{p} Y \times Z \xrightarrow{q} Z$ be the projections, which give rise to maps $\pi_0(Y) \xleftarrow{p_*} \pi_0(Y \times Z) \xrightarrow{q_*} \pi_0(Z)$ as in Proposition 5.20. We can therefore define a map*

$$\phi\colon \pi_0(Y \times Z) \to \pi_0(Y) \times \pi_0(Z)$$

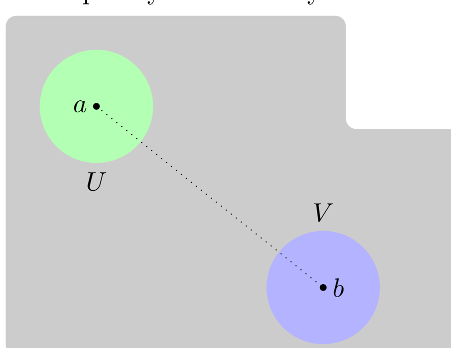
by $\phi(u) = (p_*(u), q_*(u))$, and this is a bijection.

Proof. We would like to define $\psi\colon \pi_0(Y) \times \pi_0(Z) \to \pi_0(Y \times Z)$ as follows: an element of $\pi_0(Y) \times \pi_0(Z)$ can be written as $([a], [b])$ for some $a \in Y$ and $b \in Z$, and these give a point $(a, b) \in Y \times Z$ and a path component $[(a, b)] \in \pi_0(Y, Z)$, and we want to define $\psi([a], [b]) = [(a, b)]$. We must check that this is well-defined. Suppose that $([a], [b]) = ([a'], [b'])$ in $\pi_0(Y) \times \pi_0(Z)$. This means that $[a] = [a']$ in $\pi_0(Y)$, so there is a continuous path $u\colon a \rightsquigarrow a'$ in Y. It also means that $[b] = [b']$ in $\pi_0(Z)$, so there is a continuous path $v\colon b \rightsquigarrow b'$ in Z. We define $w\colon [0,1] \to Y \times Z$ by $w(t) = (u(t), v(t))$ (which is continuous by Proposition 7.11). This gives a path $(a, b) \rightsquigarrow (a', b')$ in $Y \times Z$, proving that $[(a, b)] = [(a', b')]$, as required. Thus, our definition of ψ is valid. We can describe ϕ in similar terms by $\phi([a, b]) = ([a], [b])$. (We do not need to check separately that this is well-defined, because that was done already in Proposition 5.20 which we used to define p_* and q_*.) It is now clear that ϕ and ψ are inverse to each other, so ϕ is a bijection as claimed. $\square$

8. The Hausdorff property, and compactness

Video (Definition 8.1 to Proposition 8.8)

Let X be a metric space, and let a and b be points in X such that $a \neq b$. By the first axiom of metric spaces, this means that $d(a, b) > 0$. We can choose r with $0 < r \leq d(a, b)/2$ and put $U = OB(a, r)$ and $V = OB(b, r)$. Then U and V are open, and $a \in U$ and $b \in V$, and we have $U \cap V = \emptyset$. This is an obvious and natural construction that occurs frequently in the theory of metric spaces.



Now suppose that X is a general topological space, and we again have points $a, b \in X$ with $a \neq b$. We might again want to have open sets U and V with properties as above, but there is no longer any obvious way to produce them, and in fact, they need not exist in all cases. This leads us to introduce the following definition.

Definition 8.1. Let X be a topological space. If $a, b \in X$ with $a \neq b$, then a *Hausdorff separation* for a and b is a pair of open sets U and V such that $a \in U$ and $b \in V$ and $U \cap V = \emptyset$. We say that X is a *Hausdorff space* if every pair of distinct points has a Hausdorff separation.

Example 8.2. As in our previous discussion, if a and b are distinct points in a metric space X, then the open balls $U = OB(a, d(a,b)/2)$ and $V = OB(b, d(a,b)/2)$ give a Hausdorff separation of a and b. Thus, all metric spaces are Hausdorff spaces.

Example 8.3. Let X be a space with the indiscrete topology, as in Example 3.15, so the only open sets are $\emptyset$ and X. Suppose also that $|X| \geq 2$, so we can choose two distinct points $a, b \in X$ with $a \neq b$. It is then clear that there can be no Hausdorff separation of a and b (because both sets U and V would have to be equal to X). Thus, X is not Hausdorff.

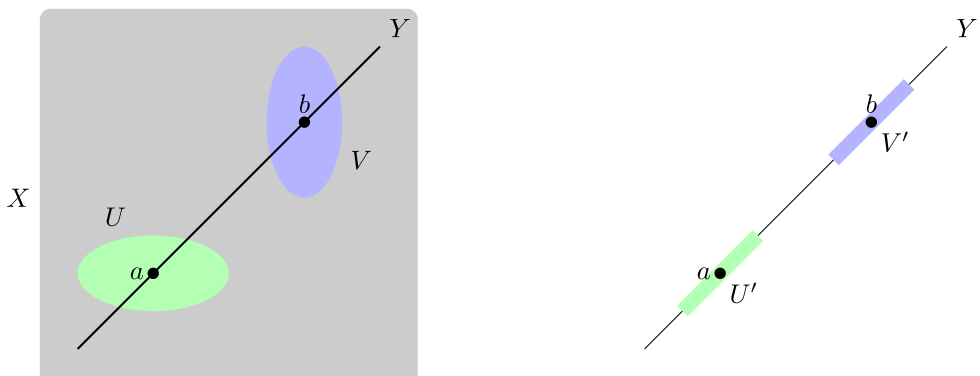
Example 8.4. The Sierpiński space X is defined as follows: the underlying set is $\{0,1\}$, and the sets $\emptyset, \{1\}$ and $\{0,1\}$ are declared to be open, but the set $\{0\}$ is not open. One can check that this is indeed a topology, and that there is no Hausdorff separation for 0 and 1, so we do not have a Hausdorff space.

Example 8.5. Consider the line with doubled origin X/E as in Example 7.27. This contains two distinct points $a = \pi(0,1)$ and $b = \pi(0,-1)$. One can check that there is no Hausdorff separation for these points, so X/E is not a Hausdorff space. In particular, this illustrates the fact that a quotient of a Hausdorff space need not be Hausdorff.

Although non-Hausdorff spaces are important in various different branches of mathematics, we will mostly restrict attention to Hausdorff spaces.

Proposition 8.6. *Let X be a Hausdorff space, and let Y be a subset of X, with the subspace topology. Then Y is also Hausdorff.*

Proof. Let a and b be distinct points in Y. As X is Hausdorff, we can choose sets U, V that are open in X with $a \in U$ and $b \in V$ and $U \cap V = \emptyset$. Put $U' = U \cap Y$ and $V' = V \cap Y$. By the definition of the subspace topology, these are open in Y. It is also clear that $a \in U'$ and $b \in V'$ and $U' \cap V' = U \cap V \cap Y = \emptyset$, so U' and V' give a Hausdorff separation of a and b in Y.



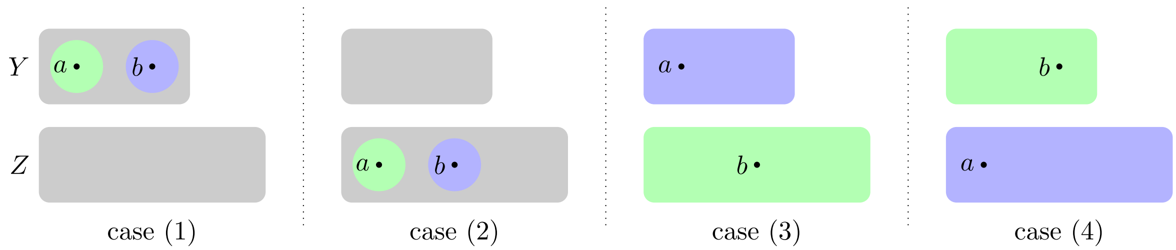
□

Proposition 8.7. *Let Y and Z be disjoint topological spaces, and put $X = Y \cup Z$, and give X the coproduct topology as in Definition 7.1. Suppose that Y and Z are both Hausdorff; then X is also Hausdorff.*

Proof. Consider a pair of distinct points $a, b \in X$. There are four possible cases:

(1) Both a and b lie in Y
(2) Both a and b lie in Z
(3) a is in Y and b is in Z
(4) a is in Z and b is in Y.

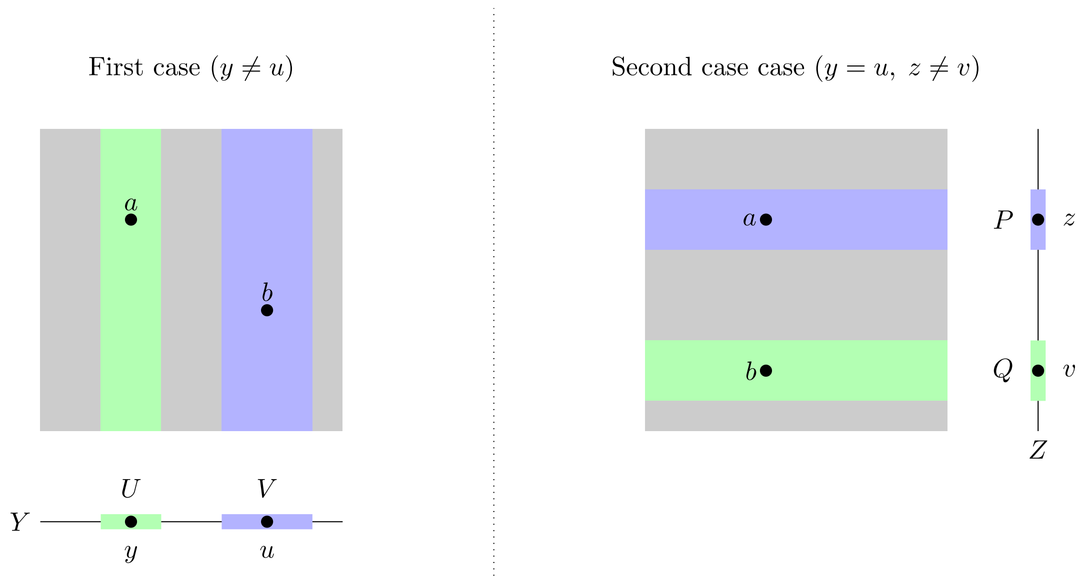
In case (a), we use the fact that Y is Hausdorff. We can thus find sets U and V that are open in Y with $a \in U$ and $b \in V$ and $U \cap V = \emptyset$. From the definition of the coproduct topology we see that U and V are still open when considered as subsets of X, so they provide the required Hausdorff separation of a and b. Case (2) is essentially the same. In case (3) the pair (Y, Z) is the required Hausdorff separation, and in case (4) we use (Z, Y) instead.



∎

Proposition 8.8. *Let Y and Z be topological spaces, and consider the space $Y \times Z$ with the product topology. Suppose that Y and Z are both Hausdorff; then $Y \times Z$ is also Hausdorff.*

Proof. Consider a pair of distinct points $a = (y, z) \in Y \times Z$ and $b = (u, v) \in Y \times Z$. As $a \neq b$, we must either have $y \neq u$ or $z \neq v$. First suppose that $y \neq u$ in Y. As Y is Hausdorff, we can choose a Hausdorff separation (U, V) for y and u in Y. It is then not hard to see that the pair $(U \times Z, V \times Z)$ is a Hausdorff separation for a and b in $Y \times Z$. Similarly, if $z \neq v$ then we can choose a Hausdorff separation (P, Q) for z and v in Z, and we find that the pair $(Y \times P, Y \times Q)$ is a Hausdorff separation for a and b in $Y \times Z$.



∎

We now turn to the notion of compactness. Compactness for metric spaces was covered in some detail in [MS, Section 7]. In particular, it was shown as [MS, Theorem 7.8] that a subset $X \subseteq \mathbb{R}^n$ is compact iff it is bounded and closed. For our purposes, you should think of this as a prototypical example: a space is compact if it is similar to a bounded, closed subset of $\mathbb{R}^n$. It was also proved in [MS, Theorem 7.20] that a metric space is compact iff it has the Heine-Borel property. In the context of general topological spaces, we take the Heine-Borel property as the *definition* of compactness. We now recall the relevant details.

Video (Definition 8.9 to Example 8.16)

Definition 8.9. Let X be a topological space. By an *open cover* of X we mean a family of open subsets $(U_i)_{i \in I}$ such that $X = \bigcup_{i \in I} U_i$. Equivalently, for every point $x \in X$ there must exist a set U_i in the family such that $x \in U_i$. By a *finite subcover* we mean a subcollection $U_{i_1}, \dots, U_{i_r}$, containing only finitely many of the subsets U_i, such that we still have $X = U_{i_1} \cup \cdots \cup U_{i_r}$. We say that X is *compact* if every open cover of X has a finite subcover.

Example 8.10. Consider the space $\mathbb{R}$. The sets $U_n = (n-1, n+1)$ (for $n \in \mathbb{Z}$) form an open cover of $\mathbb{R}$. However, we claim that there is no finite subcover. In other words, if we take any finite collection of these sets, say $U_{n_1}, \dots, U_{n_r}$, we claim that the union $U^* = U_{n_1} \cup \dots \cup U_{n_r}$ is not all of $\mathbb{R}$. This is clear: if $p = \min(n_1, \dots, n_r)$ and $q = \max(n_1, \dots, n_r)$ then $U^* \subseteq (p-1, q+1)$ and so $q+2 \notin U^*$, for example. From this it follows that $\mathbb{R}$ is not compact. (We could also have deduced this from the results in [MS, Section 7].)

Example 8.11. Now consider the space $X = (0, 1)$ and the subspaces $U_n = (1/n, 1 - 1/n)$ (for $n > 1$). These form an open cover of X but there is no finite subcover, so X is again not compact.

Definition 8.12. Let X be a subset of $\mathbb{R}^n$. We say that X is *bounded* if there is a constant $R \geq 0$ such that $\|x\| \leq R$ for all $x \in X$.

Lemma 8.13. *Let X be a space with only finitely many points; then X is compact.*

Proof. Supose that $X = \{x_1, \dots, x_n\}$. Let $(U_i)_{i \in I}$ be an open cover of X. For each t, the element x_t must lie in some set of the cover, say $x_t \in U_{i_t}$. This means that $X \subseteq U_{i_1} \cup \dots \cup U_{i_n}$, so we have a finite subcover as required. □

Proposition 8.14. *Let X be a subset of $\mathbb{R}^n$, with the subspace topology. Then X is compact iff it is bounded, and closed in $\mathbb{R}^n$.*

Proof. This follows from [MS, Theorems 7.8 and 7.20]. (It is apparently identical to Theorem 7.8, but that is slightly misleading because [MS] uses a different definition of compactness; we need Theorem 7.20 as well to show that the two versions of compactness are the same.) □

Example 8.15. The spaces S^{n-1}, Δ^{n-1} and B^n are bounded and closed in $\mathbb{R}^n$, so they are all compact.

Example 8.16. Recall that $O_n = \{A \in M_n(\mathbb{R}) \mid A^T A = I\}$. We will show that this is closed and bounded in $M_n(\mathbb{R}) \simeq \mathbb{R}^{n^2}$, so it is compact. We can define $f\colon M_n(\mathbb{R}) \to M_n(\mathbb{R})$ by $f(A) = A^T A - I$. This is continuous, and $\{0\}$ is closed in $M_n(\mathbb{R})$, and $O_n = f^{-1}\{0\}$, so O_n is closed in $M_n(\mathbb{R}) \simeq \mathbb{R}^{n^2}$. We have also explained that the standard metric on $M_n(\mathbb{R})$ can be expressed in the form

$$d(A, B) = \sqrt{\operatorname{trace}((A-B)^T (A-B))}.$$

For the identity matrix $I \in M_n(\mathbb{R})$ we have $\operatorname{trace}(I) = n$. From this we see that $d(A, 0) = \sqrt{n}$ for all $A \in O_n$, so O_n is bounded.

Video (Proposition 8.17 to Corollary 8.19)

Proposition 8.17. *Let X be a compact space, and let Y be a closed subset of X, equipped with the subspace topology. Then Y is also compact.*

Proof. Let $(V_i)_{i \in I}$ be an open cover of Y. By the definition of the subspace topology, we can find sets U_i that are open in X such that $V_i = U_i \cap Y$. As $Y = \bigcup_i V_i$, we see that $Y \subseteq \bigcup_i U_i$. Note that the set $Y^c = X \setminus Y$ is also open in X (because Y is assumed to be closed). The set Y^c together with the sets U_i then form an open cover of X. As X is compact, we can choose a finite subcover, which will consist of some finite collection of sets $U_{i_1}, \dots, U_{i_r}$, possibly together with Y^c. These sets cover all of X, so in particular they cover Y. It is clear that Y^c cannot contribute to covering Y, so we must have $Y \subseteq U_{i_1} \cup \dots \cup U_{i_r}$. This means that $Y = V_{i_1} \cup \dots \cup V_{i_r}$, so we have a finite subcover of the original cover, as required. □

Proposition 8.18. *Let X be a Hausdorff space, and let Y be a subset that is compact with respect to the subspace topology. Then Y is closed in X.*

Proof. We must show that the set $Y^c = X \setminus Y$ is open. Choose a point $x \in Y^c$. For each point $y \in Y$ we have $x \neq y$, so we can choose a Hausdorff separation, say (U_y, V_y). This means that U_y and V_y are open in X and $x \in U_y$ and $y \in V_y$ and $U_y \cap V_y = \emptyset$. Now put $W_y = V_y \cap Y$, which is open for the subspace topology on Y. As $y \in W_y$, we see that the sets W_y cover Y. By the compactness property, we can choose a finite subcollection $W_{y_1}, \dots, W_{y_r}$ that still covers Y. This means that $Y \subseteq V_{y_1} \cup \dots \cup V_{y_r}$. Now put $U^\# = U_{y_1} \cap \dots \cap U_{y_r}$. This is a finite intersection of open sets, so it is still open in X. Each of the sets U_{y_i}

contains x, so we have $x \in U^{\#}$. We also claim that $U^{\#} \subseteq Y^c$, or equivalently $U^{\#} \cap Y = \emptyset$. To see this, suppose (for a contradiction) that we have $y \in U^{\#} \cap Y$. This means that $y \in Y$, and $Y \subseteq V_{y_1} \cup \dots \cup V_{y_r}$, so we have $y \in V_{y_t}$ for some t. On the other hand, we have $y \in U^{\#} = U_{y_1} \cap \dots \cap U_{i_r}$, so in particular $y \in U_{y_t}$. We now see that $y \in U_{y_t} \cap V_{y_t}$, which is impossible because $U_z \cap V_z = \emptyset$ for all z. This contradiction shows that we in fact have $U^{\#} \subseteq Y^c$. We started with an arbitrary element $x \in Y^c$, and we produced an open set $U^{\#}$ that contains x and is contained in Y^c. By Lemma 3.33, this is enough to prove that Y^c is open. □

Corollary 8.19. *Let X be a compact Hausdorff space, and let Y be a subset of X. Then Y is closed in X iff it is compact (in the subspace topology).*

Proof. This follows easily by combining Propositions 8.17 and 8.18. □

Video (Proposition 8.20 to Proposition 8.23)

Proposition 8.20. *Let X and Y be topological spaces. Suppose that X is compact, and that we have a surjective, continuous map $f\colon X \to Y$. Then Y is also compact.*

Proof. Consider an open cover $(V_i)_{i\in I}$ of Y. Put $U_i = f^{-1}(V_i) \subseteq X$, so that U_i is open by the definition of continuity. If x is any point in X, then $f(x) \in Y$, and the sets V_i cover Y, so we can choose i with $f(x) \in V_i$, which means that $x \in U_i$. This shows that the sets U_i form an open cover of X. As X is assumed to be compact, there must be a finite subcover, say $U_{i_1}, \dots, U_{i_r}$. We claim that the corresponding sets $V_{i_1}, \dots, V_{i_r}$ cover Y. To see this, consider an arbitrary point $y \in Y$. As f is assumed to be surjective, we can choose $x \in X$ with $f(x) = y$. As the sets $U_{i_1}, \dots, U_{i_r}$ cover X, we can choose t such that $x \in U_{i_t}$. Here U_{i_t} was defined to be $f^{-1}(V_{i_t})$, so we must have $f(x) \in V_{i_t}$, or in other words $y \in V_{i_t}$ as required. Thus, the list $V_{i_1}, \dots, V_{i_r}$ is a finite subcover of our original cover. □

Corollary 8.21. *Let X be a compact space, and let E be an equivalence relation on X. Then the quotient space X/E is also compact.*

Proof. The quotient map $\pi\colon X \to X/E$ is surjective and continuous, so this follows from Proposition 8.20. □

Example 8.22. We can define an equivalence relation E on S^n by xEy iff $y = \pm x$. The real projective space $\mathbb{R}P^n$ is defined to be the quotient space S^n/E. This is compact, because S^n is compact. Similarly, we can regard S^{2n+1} as the unit sphere in the space $\mathbb{C}^{n+1} = \mathbb{R}^{2n+2}$. We can then define another equivalence relation F on S^{2n+1} by declaring that xFy iff $y = e^{i\theta}x$ for some $\theta \in \mathbb{R}$. The complex projective space $\mathbb{C}P^n$ is defined to be S^{2n+1}/F; this is again compact.

Proposition 8.23. *Let Y and Z be disjoint spaces, and take $X = Y \cup Z$, with the coproduct topology. If Y and Z are both compact, then so is X.*

Proof. Let $(U_i)_{i\in I}$ be an open cover of X. Then each set U_i must have the form $V_i \cup W_i$, where V_i is an open subset of Y, and W_i is an open subset of Z. It is easy to see that the sets V_i form an open cover of Y, so there must be a finite subcover, say $Y_{i_1}, \dots, Y_{i_r}$. Similarly, the sets W_i form an open cover of Z, so there must be a finite subcover, say $W_{j_1}, \dots, W_{j_s}$. We then find that the list $U_{i_1}, \dots, U_{i_r}, U_{j_1}, \dots, U_{j_s}$ forms a finite subcover of our original cover. □

Video (Theorem 8.24 to Lemma 8.26)

Theorem 8.24. *Let Y and Z be compact spaces. Then the product $Y \times Z$ is also compact (under the product topology).*

We will give the proof after some preliminary definitions and results.

Definition 8.25. Let Y and Z be as above, and let $(U_i)_{i\in I}$ be an open cover of $Y \times Z$. We will say that a subset $K \subseteq Y \times Z$ is *finitely covered* if there is a finite subcollection $U_{i_1}, \dots, U_{i_r}$ such that $K \subseteq U_{i_1} \cup \dots \cup U_{i_r}$.

We need to prove that the whole space $Y \times Z$ is finitely covered, but we will build up to this by proving that certain other sets are finitely covered first.

Lemma 8.26. *In the above context, let y be a point in Y. Then there is an open set $S \subseteq Y$ such that $y \in S$ and $S \times Z$ is finitely covered.*

Proof. For each $z \in Z$, choose an index i_z such that $(y, z) \in U_{i_z}$. As U_{i_z} is open in the product topology, we can then choose a box $B_z = V_z \times W_z$ such that $y \in V_z$ and $z \in W_z$ and $V_z \times W_z \subseteq U_{i_z}$. As $z \in W_z$ for all z, we see that the sets $(W_z)_{z \in Z}$ form an open cover of Z. As Z is compact, we can choose a finite subcover, say $Z = W_{z_1} \cup \ldots \cup W_{z_r}$. Now put $S = V_{z_1} \cap \cdots \cap V_{z_r}$. This is the intersection of a finite list of open sets, each of which contains y; so S itself is an open set containing y. We next claim that $S \times Z \subseteq B_{z_1} \cup \cdots \cup B_{z_r}$. Indeed, suppose we have a point $(s, z) \in S \times Z$. As $Z = W_{z_1} \cup \ldots \cup W_{z_r}$, we can choose k such that $z \in W_{z_k}$. Now $s \in S = V_{z_1} \cap \cdots \cap V_{z_r}$, so we also have $s \in V_{z_k}$. This proves that $(s, z) \in V_{z_k} \times W_{z_k} = B_{z_k}$, as required. We also have $B_{z_k} \subseteq U_{i_{z_k}}$, so

$$S \times Z \subseteq U_{i_{z_1}} \cup \cdots \cup U_{i_{z_r}}.$$

This shows that $S \times Z$ is finitely covered. □

Proof of Theorem 8.24. For each $y \in Y$, the lemma tells us that we can choose an open set S_y containing y such that $S_y \times Z$ is finitely covered. The family $(S_y)_{y \in Y}$ is then an open cover of Y. As Y is assumed to be compact we can choose a finite subcover, say $Y = S_{y_1} \cup \cdots \cup S_{y_r}$. It follows that $Y \times Z = \bigcup_{i=1}^{r} (S_{y_i} \times Z)$. Here each of the subsets $S_{y_i} \times Z$ is finitely covered, and there are only finitely many of them, so the union is also finitely covered. This means that $Y \times Z$ is finitely covered, or in other words that our original cover of $Y \times Z$ has a finite subcover, as required. □

Video (Proposition 8.27 and Example 8.28)

Proposition 8.27. *Let X be a compact space, and let Y be a Hausdorff space. Let $f \colon X \to Y$ be a continuous bijection. Then the inverse map $f^{-1} \colon Y \to X$ is also continuous, so f is a homeomorphism.*

Proof. Let $g \colon Y \to X$ be the inverse of f, so the claim is that g is continuous. By Proposition 3.23, it will be enough to check that for every closed subset $F \subseteq X$, the preimage $g^{-1}(F)$ is closed in Y. Here F is a closed subset of a compact space, so it is compact by Proposition 8.17. As g is inverse to f, we have $g^{-1}(F) = f(F)$. We can regard f as a continuous surjective map from F to $f(F)$ (where F and $f(F)$ are given the respective subspace topologies). It follows by Proposition 8.20 that $f(F)$ is also compact. Now $f(F)$ is a compact subspace of the Hausdorff space Y, so it is closed by Proposition 8.18. In other words, the preimage $g^{-1}(F)$ is closed, as required. □

Example 8.28. In Example 7.24 we considered a space X consisting of two disjoint discs, and the space $Y = S^2$. We introduced an equivalence relation E on X (corresponding to the idea of gluing the boundary circles of the two discs) and defined a continuous bijection $\overline{f} \colon X/E \to Y$. Now X/E is a quotient of a closed bounded subset of $\mathbb{R}^3$, so it is compact, and Y is a metric space, so it is Hausdorff. It follows from Proposition 8.27 that $\overline{f}$ is actually a homeomorphism.

Example 8.29. In Example 7.26 we constructed a continuous bijection $\overline{f} \colon \mathbb{R}^2/E \to T$, where T is the torus and the equivalence relation E is given by xEy iff $x - y \in \mathbb{Z}^2$. Here T is a metric space and therefore Hausdorff. It is easy to see that the composite

$$[0,1]^2 \xrightarrow{\text{inc}} \mathbb{R}^2 \xrightarrow{\pi} \mathbb{R}^2/E$$

is surjective, and $[0,1]^2$ is compact, so $\mathbb{R}^2/E$ is compact by Proposition 8.20.

We conclude with some slightly different results about open covers, that are only relevant for metric spaces. Their use will become apparent later.

Video (Definition 8.30 and Proposition 8.31)

Definition 8.30. Let X be a metric space, and let $(U_i)_{i \in I}$ be an open cover. A *Lebesgue number* for this cover is a number $\epsilon > 0$ with the following property: for every point $x \in X$, there is an index i such that $OB(x, \epsilon) \subseteq U_i$.

You should think of a Lebesgue number as something like the minimum size of overlaps between adjacent sets in the cover.

Proposition 8.31. *Let X be a compact metric space. Then every open cover of X has a Lebesgue number.*

Proof. Let $(U_i)_{i\in I}$ be an open cover of X. For each point $x \in X$, we can choose an index i_x such that $x \in U_{i_x}$. As U_{i_x} is open, we can then choose $r_x > 0$ such that $OB(x, r_x) \subseteq U_{i_x}$. Put $V_x = OB(x, r_x/2)$, so V_x is an open subset of X containing x. The sets V_x then form an open cover of X, so we can choose a finite subcover, say $X = V_{x_1} \cup \cdots \cup V_{x_n}$. Put $\epsilon = \min(r_{x_1}, \ldots, r_{x_n})/2$. We claim that this is a Lebesgue number. To see this, let x be an arbitrary point in X. As the sets $V_{x_1}, \ldots, V_{x_n}$ form a cover, we can choose k such that $x \in V_{x_k}$, or in other words $d(x, x_k) < r_{x_k}/2$. If $u \in OB(x, \epsilon)$ then $d(u, x) < \epsilon \leq r_{x_k}/2$ and $d(x, x_k) < r_{x_k}/2$ so $d(u, x_k) < r_{x_k}$. This proves that $OB(x, \epsilon) \subseteq OB(x_k, r_{r_k}) \subseteq U_{i_{x_k}}$, so $OB(x, \epsilon)$ is contained in one of the sets U_i, as required. □

9. Homotopy

Video (Definition 9.1 to Example 9.6)

We would like to make the following definition:

Definition 9.1 (Bogus)**.** Let X and Y be topological spaces, and let $f, g \colon X \to Y$ be two continuous maps from X to Y, or in other words, two elements of $\mathrm{Top}(X, Y)$. We say that f and g are *homotopic* if there is a path $u \colon [0, 1] \to \mathrm{Top}(X, Y)$ from f to g, so that f and g lie in the same path component of $\mathrm{Top}(X, Y)$. This is an equivalence relation, by Proposition 5.9; the equivalence classes are called *homotopy classes* of maps. We define $[X, Y] = \pi_0 \mathrm{Top}(X, Y)$, so $[X, Y]$ is the set of all homotopy classes of maps from X to Y.

This is bogus, because we have not introduced a topology on the set $\mathrm{Top}(X, Y)$, so it is not meaningful to talk about continuous paths. As we mentioned in Remark 7.28, it is possible to introduce an appropriate topology on $\mathrm{Top}(X, Y)$, but this involves many subtle technicalities. We therefore reformulate the definition in a different way.

Definition 9.2. Let X and Y be topological spaces, and let $f, g \colon X \to Y$ be two continuous maps from X to Y. A *homotopy* from f to g is a continuous map $h \colon [0, 1] \times X \to Y$ such that $h(0, x) = f(x)$ and $h(1, x) = g(x)$ for all $x \in X$. We say that f and g are *homotopic* if there is a homotopy between them, and we write $f \equiv g$ in this case.

Proposition 9.3. *The relation of being homotopic is an equivalence relation.*

Proof. This is closely analogous to Proposition 5.9.

(a) We can define a homotopy h from f to f by $h(t, x) = f(x)$ for all t. This proves that $f \equiv f$, so the relation is reflexive.
(b) Suppose that $f \equiv g$, so we can choose a homotopy h from f to g. The function $\overline{h}(t, x) = h(1 - t, x)$ then gives a homotopy from g to f, proving that $g \equiv f$. Thus, the relation is symmetric.
(c) Now suppose that $e \equiv f$ and $f \equiv g$, so we can choose a homotopy k from e to f, and another homotopy h from f to g. We then define $k * h \colon [0, 1] \times X \to Y$ by

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

The two clauses are consistent, because when $t = \frac{1}{2}$ we have

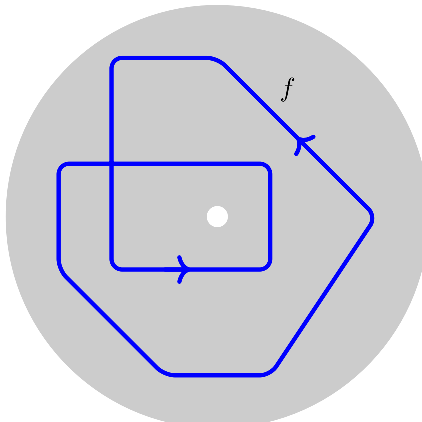
$$k(2t, x) = k(1, x) = f(x) = h(0, x) = h(2t - 1, x).$$

The combined map $k * h$ is easily seen to be continuous on when restricted to the sets $[0, \frac{1}{2}] \times X$ and $[\frac{1}{2}, 1] \times X$. These sets are closed (in the product topology), and their union is all of $[0, 1] \times X$, so the full map $k * h$ is continuous by Proposition 3.35. It gives a homotopy from e to g, proving that the homotopy relation is transitive.

□

Definition 9.4. The equivalence class $[f]$ of a continuous map $f\colon X \to Y$ is called the *homotopy class* of f. We define $[X, Y] = \mathrm{Top}(X, Y)/\equiv$, so this is the set of all homotopy classes.

Example 9.5. Consider a continuous map $f\colon S^1 \to \mathbb{C} \setminus \{0\}$. Then f represents a closed loop in the complex plane, which cannot pass through the origin, but which can wind around the origin. Let $n(f)$ be the total number of times the curve winds around the origin, with anticlockwise turns counting positively, and clockwise turns counting negatively. We call this the *winding number* of f. (We will formulate the definition more carefully at a later stage.)



It can be shown that two maps from S^1 to $\mathbb{C} \setminus \{0\}$ are homotopic iff they have the same winding number. It follows that the map $n\colon \mathrm{Top}(S^1, \mathbb{C} \setminus \{0\}) \to \mathbb{Z}$ induces a bijection $[S^1, \mathbb{C} \setminus \{0\}] \to \mathbb{Z}$.

Example 9.6. Let X be an arbitrary topological space, and let Y be a subset of $\mathbb{R}^n$, with the subspace topology. Let $f, g\colon X \to Y$ be continuous maps. We can then define a map $h\colon [0,1] \times X \to \mathbb{R}^n$ by

$$h(t, x) = (1 - t)\, f(x) + t\, g(x).$$

In general, we have no right to expect that this will land in the subspace $Y \subseteq \mathbb{R}^n$, so it will probably not provide a homotopy between f and g. However, in some special cases we may be able to prove that $h([0,1] \times X) \subseteq Y$, and in that case we do get a homotopy from f to g, which we will call a *straight line homotopy*.

For the most extreme example of this, suppose that Y is convex (as in Example 5.11), so that for all $a, b \in Y$, the line segment from a to b is contained wholly in Y. Let f and g be continuous maps from X to Y, and put $h(t, x) = (1 - t)\, f(x) + t\, g(x)$ as before. Then $h(t, x)$ lies on the line segment from $f(x) \in Y$ to $g(x) \in Y$, so $h(t, x) \in Y$ for all t and x. Thus, h gives a homotopy from f to g, proving that $[f] = [g]$ in $[X, Y]$. As f and g were arbitrary, it follows that $|[X, Y]| = 1$.

We next check that the notion of homotopy is compatible with composition.

Video (Proposition 9.7 to Definition 9.9)

Proposition 9.7. *Suppose we have continuous maps*

$$X \underset{f_1}{\overset{f_0}{\rightrightarrows}} Y \underset{g_1}{\overset{g_0}{\rightrightarrows}} Z$$

Suppose that f_0 is homotopic to f_1, and that g_0 is homotopic to g_1. Then $g_1 \circ f_1$ is homotopic to $g_0 \circ f_0$.

Proof. We are assuming that f_0 is homotopic to f_1, which means that there is a homotopy $h\colon [0,1] \times X \to Y$ with $h(0, x) = f_0(x)$ and $h(1, x) = f_1(x)$ for all $x \in X$. We are also assuming that g_0 is homotopic to g_1, which means that there is a homotopy $k\colon [0,1] \times Y \to Z$ with $k(0, y) = g_0(y)$ and $k(1, y) = g_1(y)$ for all $y \in Y$. We can therefore define $m\colon [0,1] \times X \to Z$ by

$$m(t, x) = k(t,\ h(t, x)).$$

This satisfies $m(0, x) = k(0, h(0, x)) = k(0, f_0(x)) = g_0(f_0(x))$ and $m(1, x) = k(1, h(1, x)) = k(1, f_1(x)) = g_1(f_1(x))$. Thus, m gives the required homotopy, provided that we can check that it is continuous. The tidiest proof of continuity is as follows: we let $p\colon [0,1] \times X \to [0,1]$ be the projection, and note that m can be written as the composite

$$[0,1] \times X \xrightarrow{\langle p,h\rangle} [0,1] \times Y \xrightarrow{k} Z.$$

Here p is continuous by Lemma 7.10, so $\langle p, h\rangle$ is continuous by Proposition 7.11, so m is continuous by Proposition 3.24. □

Corollary 9.8. *For $u \in [X, Y]$ and $v \in [Y, Z]$, there is a well-defined composite $v \circ u \in [X, Z]$, given by $v \circ u = [g \circ f]$ for any choice of maps f, g with $u = [f]$ and $v = [g]$.*

Proof. Immediate from the proposition. □

Definition 9.9. We can now define a new category hTop, called the *homotopy category*. The objects are topological spaces (just as for the category Top), but the morphisms are now homotopy classes of maps, so

$$\mathrm{hTop}(X, Y) = \mathrm{Top}(X, Y)/\equiv = [X, Y].$$

The composition rule is given by Corollary 9.8. The identity morphism in hTop for an object X is just the homotopy class $[\mathrm{id}_X]$ corresponding to the identity function.

Video (Definition 9.10 to Proposition 9.20)

Definition 9.10. A continuous map $f\colon X \to Y$ is a *homotopy equivalence* if the corresponding homotopy class $[f]$ is an isomorphism in the homotopy category. Explicitly, this means that there must exist a continuous map $g\colon Y \to X$ such that $g \circ f \equiv \mathrm{id}_X$ and $f \circ g \equiv \mathrm{id}_Y$. Any such map g is called a *homotopy inverse* for f. If there exists a homotopy equivalence from X to Y, we will say that X and Y are *homotopy equivalent.* We will sometimes use the notation $X \cong Y$ to indicate this.

Lemma 9.11. *Every homeomorphism is a homotopy equivalence.*

Proof. Let $f\colon X \to Y$ be a homeomorphism, with inverse $g\colon Y \to X$. This means that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. As $g \circ f$ is equal to id_X, it is certainly homotopic to id_X. As $f \circ g$ is equal to id_Y, it is certainly homotopic to id_Y. Thus, g is also a homotopy inverse for f, so f is a homotopy equivalence. □

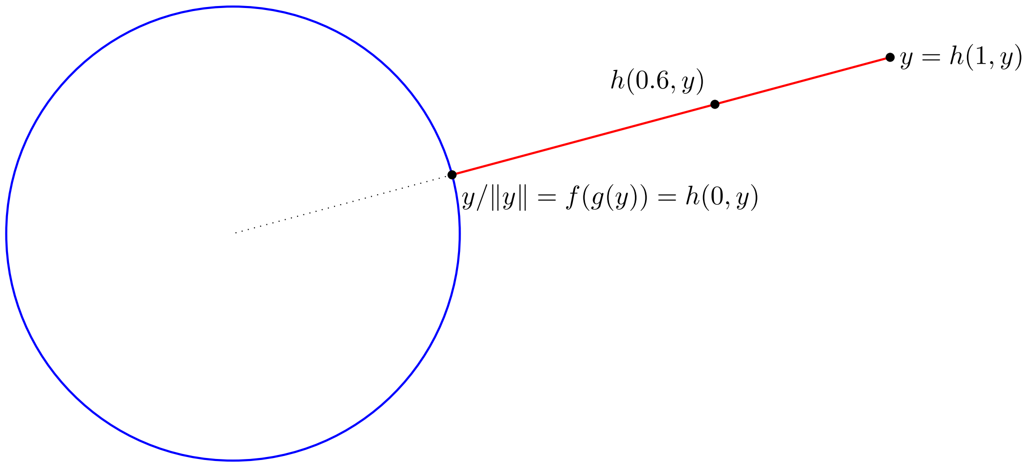
A basic example is as follows:

Proposition 9.12. *For $n > 0$, the sphere $X = S^{n-1}$ is homotopy equivalent to the space $Y = \mathbb{R}^n \setminus \{0\}$.*

Proof. Recall that $X = S^{n-1} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$. We define $f\colon X \to Y$ to be the inclusion function, so we just have $f(x) = x$. We define $g\colon Y \to X$ by $g(y) = y/\|y\|$. (It is important here that Y is the space of *nonzero* vectors in $\mathbb{R}^n$, so $\|y\| > 0$ and it is valid to divide by $\|y\|$.) Note that for $x \in X$ we have $\|x\| = 1$ so $g(f(x)) = x/\|x\| = x$, so $g \circ f = \mathrm{id}_X$. As $g \circ f$ is equal to id_X, it is certainly homotopic to id_X. In the opposite direction, however, the composite $g \circ f$ is not equal to the identity. Nonetheless, we can define $h\colon [0,1] \times Y \to \mathbb{R}^n$ by

$$h(t, y) = (1-t)\frac{y}{\|y\|} + ty = \left((1-t)\|y\|^{-1} + t\right) y.$$

Here $(1-t)\|y\|^{-1} + t$ is always strictly positive (for $0 \leq t \leq 1$), so $h(t, y)$ can never be zero, so we can regard h as a map $[0,1] \times Y \to Y$. Now $h(0, y) = f(g(y))$ and $h(1, y) = y$, so h is a homotopy from $g \circ f$ to id_Y. This proves that g is a homotopy inverse for f, so that f is a homotopy equivalence.



□

Definition 9.13. Let X be a topological space, and let a be a point in X. A *contraction* of X to a is a continuous map $h\colon [0,1] \times X \to X$ such that $h(0, x) = a$ for all x, and $h(1, x) = x$ for all x. We say that X is *contractible* if it has a contraction (to some point $a \in X$).

Example 9.14. Let $X \subseteq \mathbb{R}^n$ be a convex space, and let a be a point in X. We can then define a linear contraction of X to a by $h(t, x) = (1 - t)a + tx$, so X is contractible. In particular, the spaces B^n, Δ_n and $[0,1]^n$ are all contractible.

Example 9.15. The sphere S^2 is not contractible, as you will probably agree if you try to imagine a contraction. However, at the moment we do not have any way of proving this.

Definition 9.16. We write 1 for the set $\{0\}$ (or any other set with precisely one element). We regard this as a topological space using the discrete topology (which is in fact the only possible topology in this case).

Proposition 9.17. *A space X is contractible iff it is homotopy equivalent to* 1.

Proof. Suppose that X is homotopy equivalent to 1. This means that we have maps $f\colon X \to 1$ and $g\colon 1 \to X$, and a homotopy h from $g \circ f$ to id_X, and a homotopy k from $f \circ g$ to id_1. As $1 = \{0\}$ there is no choice about f: we must have $f(x) = 0$ for all x. Similarly, there is no choice about k: we must have $k(t, 0) = 0$ for all t. However, there is some choice about g and h. Put $a = g(0) \in X$. We then have $g(f(x)) = g(0) = a$ for all x. Moreover, h is a homotopy from $g \circ f$ to id_X, so we have $h(0, x) = g(f(x)) = a$ and $h(1, x) = x$ for all x. Thus, h is a contraction t a, showing that X is contractible.

Conversely, suppose that we start from the assumption that X is contractible. All the above steps can be reversed in a straightforward way to prove that X is homotopy equivalent to 1. □

Remark 9.18. The map $[0,1] \to 1$ is the most basic example of a homotopy equivalence that is not a homeomorphism.

Proposition 9.19. *The relation of being homotopy equivalent is an equivalence relation.*

Proof. The identity function from X to itself is clearly a homotopy equivalence, so the relation is reflexive. Suppose that X is homotopy equivalent to Y, so we can choose a homotopy equivalence $f\colon X \to Y$ and a homotopy inverse $g\colon Y \to X$, so $fg\colon Y \to Y$ and $gf\colon X \to X$ are homotopic to the respective identity maps. This means that g is a homotopy equivalence with homotopy inverse f, so Y is homotopy equivalent to X. This proves that our relation is symmetric. Finally, suppose that X is homotopy equivalent to Y and Y is homotopy equivalent to Z. Let d and e be homotopy inverses for f and g, so $df \simeq 1_X$ and $fd \simeq 1_Y \simeq eg$ and $ge \simeq 1_Z$. Using Proposition 9.7, we deduce that

$$degf = d(eg)f \simeq d1_Y f = df \simeq 1_X$$
$$gfde = g(fd)e \simeq g1_Y e = ge \simeq 1_Z.$$

This proves that $gf\colon X \to Z$ is a homotopy equivalence, with homotopy inverse $de\colon Z \to X$. This in turn shows that our relation is transitive, as required. □

Proposition 9.20. *Suppose that X_0 is homotopy equivalent to Y_0 and X_1 is homotopy equivalent to Y_1. Then $X_0 \times X_1$ is homotopy equivalent to $Y_0 \times Y_1$.*

Proof. For $i = 0, 1$ we let $f_i : X_i \to Y_i$ be a homotopy equivalence, with homotopy inverse $g_i : Y_i \to X_i$. This means that we have homotopies h_i from $g_i \circ f_i$ to id_{X_i}, and homotopies k_i from $f_i \circ g_i$ to id_{Y_i}. Now make the following definitions:

$$\begin{aligned}
&f : X_0 \times X_1 \to Y_0 \times Y_1 && f(x_0, x_1) = (f_0(x_0), f_1(x_1)) \\
&g : Y_0 \times Y_1 \to X_0 \times X_1 && g(y_0, y_1) = (g_0(y_0), g_1(y_1)) \\
&h : [0,1] \times X_0 \times X_1 \to X_0 \times X_1 && h(t, x_0, x_1) = (h_0(t, x_0),\ h_1(t, x_1)) \\
&k : [0,1] \times Y_0 \times Y_1 \to Y_0 \times Y_1 && k(t, y_0, y_1) = (k_0(t, y_0),\ k_1(t, y_1)).
\end{aligned}$$

We find that h gives a homotopy from $g \circ f$ to the identity, and k gives a homotopy from $f \circ g$ to the identity, so we have a homotopy equivalence between $X_0 \times X_1$ and $Y_0 \times Y_1$, as required. $\square$

Example 9.21. The solid torus, the Möbius band, and $\mathbb{C} \setminus \{0\}$ are all homotopy equivalent to S^1. To explain this in more detail, let D be the vertical disc in the xz plane of radius 1 centred at $(2, 0, 0)$. The "solid torus" is the space obtained by revolving D around the z-axis; this is easily seen to be homeomorphic to $S^1 \times D^2$. Now D^2 is convex, so it is homotopy equivalent to 1. It follows that $S^1 \times D^2$ is homotopy equivalent to $S^1 \times 1 = S^1$.

Next, for $\theta \in [0, 2\pi]$, let P_θ be the vertical plane through the z-axis in $\mathbb{R}^3$ that has angle θ with the xz-plane. Let D_θ be the intersection of P_θ with the solid torus, which is a vertical disc of radius 1 centred at $(2\cos(\theta), 2\sin(\theta), 0)$. Let I_θ be the diameter of D_θ that makes an angle of $\theta/4$ to the vertical, and let M be the union of all the sets D_θ. This is a version of the Möbius band. It is homeomorphic to the space

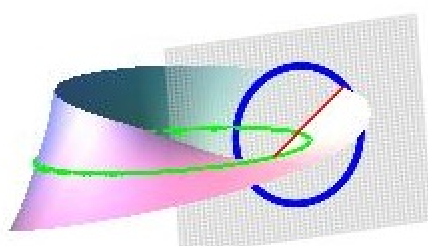
$$M' = \{(z, w) \in S^1 \times B^2 \mid w^2/z \text{ is real and nonnegative } \}.$$

Interactive demo

Define $f : S^1 \to M'$ and $g : M' \to S^1$ and $h : I \times M' \to M'$ by

$$\begin{aligned}
f(z) &= (z, 0) \\
g(z, w) &= z \\
h(t, (z, w)) &= (z, tw).
\end{aligned}$$

Then $gf = 1$ and h is a homotopy from fg to 1, so g is a homotopy inverse for f, so M' is homotopy equivalent to S^1 as claimed.



Interactive demo

Finally, Proposition 9.12 tells us that $\mathbb{R}^2 \setminus \{0\}$ is homotopy equivalent to S^1, and $\mathbb{C}$ can be identified with $\mathbb{R}^2$, so $\mathbb{C} \setminus \{0\}$ is also homotopy equivalent to S^1.

Interactive demo

Definition 9.22. Let X and Y be topological spaces. We say that X is a *homotopy retract* of Y if there exist continuous maps $X \xrightarrow{f} Y \xrightarrow{g} X$ such that $g \circ f$ is homotopic to id_X. (We make no assumption about $f \circ g$.) Any pair (f, g) with this property will be called a *homotopy retraction pair* for (X, Y).

Example 9.23. Put

$$X = \mathbb{R}^2 \setminus \{0\}$$
$$Y = \text{ figure eight } = \{(x,y) \in \mathbb{C} \mid (x-1)^2 + y^2 = 1 \text{ or } (x+1)^2 + y^2 = 1\}.$$

Note that X is two-dimensional whereas Y is one-dimensional so there is no injective continuous map from X to Y, so X is not an actual retract of Y. However, it is a homotopy retract of Y. To see this, define maps $X \xrightarrow{f} Y \xrightarrow{g} X$ by

$$f(x,y) = (1,0) + (x,y)/\|(x,y)\|$$
$$g(x,y) = (x-1,y).$$

The composite $g \circ f : X \to X$ is just $(g \circ f)(x,y) = (x,y)/\|(x,y)\|$. The straight line joining (x,y) to $(x,y)/\|(x,y)\|$ does not pass through the origin, so $g \circ f$ is homotopic to the identity as required. (The map $f \circ g : Y \to Y$ is not homotopic to the identity, so we do not have a homotopy equivalence, but that is not important.)

Proposition 9.24. *Suppose that X is a homotopy retract of Y and that Y is contractible. Then X is also contractible.*

Proof. By hypothesis, we have continuous maps

$$\begin{tikzcd} X \arrow[r, shift left, "f"] & Y \arrow[l, shift left, "g"] \arrow[r, shift left, "p"] & 1 \arrow[l, shift left, "q"] \end{tikzcd}$$

such that gf, qp and pq are homotopic to the respective identity maps. Now put $m = pf : X \to 1$ and $n = gq : 1 \to X$. We then have $nm = gqpf \simeq gf \simeq \mathrm{id} : X \to X$. Also, mn is a map from 1 to 1, and the only map from 1 to 1 is the identity, so $mn = \mathrm{id}$. Thus, m and n give a homotopy equivalence from X to 1, proving that X is contractible. □

Proposition 9.25. *Suppose we have two continuous maps $f, g : X \to Y$, giving maps $f_*, g_* : \pi_0(X) \to \pi_0(Y)$ as in Proposition 5.20. If f is homotopic to g, then $f_* = g_*$.*

Proof. Let $h : [0,1] \times X \to Y$ be a homotopy between f and g, so $h(0,x) = f(x)$ and $h(1,x) = g(x)$ for all x. For any $a \in X$ we have $f_*[a] = [f(a)]$ and $g_*[a] = [g(a)]$. We need to prove that these are the same. Equivalently, we need to find a path u from $f(a)$ to $g(a)$ in Y. We can just take $u(t) = h(t,a)$. □

10. Homology

Video (Definition 10.1 to Example 10.9)

Definition 10.1. Let X be a space. A *singular k-simplex* in X is a continuous map $u : \Delta_k \to X$. We write $S_k(X)$ for the set of singular k-simplices in X.

Example 10.2. Recall that Δ_0 is the set $\{e_0\}$ with just one point. To give a function $u : \Delta_0 \to X$ is the same as to give a point $u(e_0) \in X$; this, we can identify $S_0(X)$ with X.

Example 10.3. As usual, we identify Δ_1 with $[0,1]$, with the point $(1-t,t) \in \Delta_1$ corresponding to the point $t \in [0,1]$. Thus, a singular 1-simplex in X is the same as a continuous map $u : [0,1] \to X$, or in other words a path in X. This means that $S_1(X)$ is the set of all possible paths in X.

Example 10.4. Suppose that $a_0, \ldots, a_k \in \mathbb{R}^N$. We can then define a map

$$\langle a \rangle = \langle a_0, \ldots, a_k \rangle : \Delta_k \to \mathbb{R}^N$$

(or in other words an element $\langle a \rangle \in S_k\mathbb{R}^N$) by

$$\langle a \rangle(t_0, \ldots, t_n) = t_0 a_0 + \cdots + t_k a_k.$$

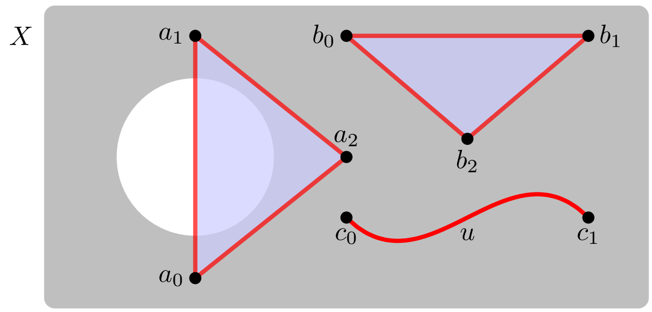
We call maps of this type *linear simplices.*

In the case $k = 0$ we have $S_0(X) = X$ and the map $\langle a_0 \rangle$ just corresponds to the point a_0. In the case $k = 1$, the map $\langle a_0, a_1 \rangle$ corresponds to the straight line path from a_0 to a_1. In the case $k = 2$, the image of the map $\langle a_0, a_1, a_2 \rangle : \Delta_2 \to \mathbb{R}^N$ is the triangle with vertices a_0, a_1 and a_2.

Now suppose that $X \subseteq \mathbb{R}^N$ and that $a_0, \dots, a_k \in X$. It may or may not happen that the image of the map $\langle a \rangle \colon \Delta_k \to \mathbb{R}^N$ actually lies in X; this must be checked carefully in any context where we want to use this construction. If so, we can regard $\langle a \rangle$ as an element of $S_k(X)$.

Example 10.5. This picture shows a space $X \subset \mathbb{R}^2$, together with:

- A linear 2-simplex $\langle a_0, a_1, a_2 \rangle \in S_2\mathbb{R}^2$, which is not an element of $S_2(X)$.
- Another linear 2-simplex $\langle b_0, b_1, b_2 \rangle \in S_2(X) \subset S_2(\mathbb{R}^2)$.
- A nonlinear 1-simplex $u \in S_1(X)$.



Definition 10.6. Let P be a set. We write $\mathbb{Z}\{P\}$ for the set of formal $\mathbb{Z}$-linear combinations of elements of P. Thus, if $p, q, r \in P$ then $5p - 9q + 7r \in \mathbb{Z}\{P\}$, for example. We call $\mathbb{Z}\{P\}$ the *free abelian group generated by* P. (It is clearly an abelian group under addition.)

Remark 10.7. Suppose that P is finite, say $P = \{p_1 \dots, p_n\}$. We then have an isomorphism $\phi \colon \mathbb{Z}^n \to \mathbb{Z}\{P\}$ given by

$$\phi(a_1, \dots, a_n) = a_1 p_1 + \cdots + a_n p_n.$$

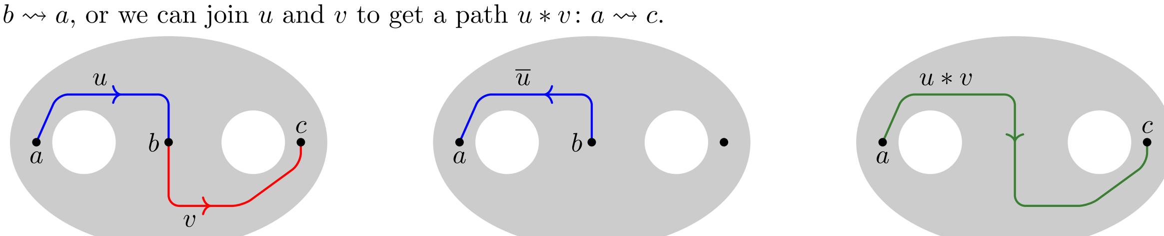
However, we will most often be considering cases where P is infinite.

Definition 10.8. A *singular k-chain* in X is a formal $\mathbb{Z}$-linear combination of singular k-simplices, or in other words, an element of $\mathbb{Z}\{S_k(X)\}$. We write $C_k(X) = \mathbb{Z}\{S_k(X)\}$ for the group of singular k-chains. For convenience, we also define $C_k(X) = 0$ for $k < 0$.

Example 10.9. Consider again the picture in Example 10.5:

- The expression $6a_1 - 4b_2 + 7c_1 \in C_0(X)$ is a singular 0-chain.
- The expression $3\langle a_0, a_2 \rangle - \langle b_0, b_1 \rangle + u \in C_1(X)$ is a singular 1-chain.
- The expression $\langle b_0, b_1, b_2 \rangle \in S_2(X) \subset C_2(X)$ is a singular 2-chain.
- No expression involving $\langle a_0, a_1 \rangle$ gives a singular chain in X, because the straight line from a_0 to a_1 is not contained in X.

Remark 10.10. Suppose we have paths $u \colon a \rightsquigarrow b$ and $v \colon b \rightsquigarrow c$ in X. We can reverse u to get a path $\overline{u} \colon b \rightsquigarrow a$, or we can join u and v to get a path $u * v \colon a \rightsquigarrow c$.



We can regard $u * v$ and $u + v$ as elements of $C_1(X)$, but they are not the same. Similarly, we can regard $\overline{u}$ and $-u$ as elements of $C_1(X)$, but they are not the same. There is clearly an important relationship between $u * v$ and $u + v$, and between $\overline{u}$ and $-u$, but it will take a little work to formulate this mathematically.

Video (Predefinition 10.11 to Example 10.15)

We next need to define the algebraic boundary $\partial u \in C_{k-1}(X)$ for a k-chain $u \in C_k(X)$. We start by considering the cases $k = 0$, $k = 1$ and $k = 2$.

Predefinition 10.11.

- For $u \in C_0(X) = \mathbb{Z}\{X\}$ we just define $\partial u = 0$.
- Now consider a singular 1-simplex $u\colon \Delta_1 \to X$. This is a path with endpoints $u(e_0)$ and $u(e_1)$. These endpoints are elements of the set X, which we identify with $S_0(X)$, so the difference $u(e_1) - u(e_0)$ can be regarded as an element of $C_0(X)$. We define $\partial(u) = u(e_1) - u(e_0)$. More generally, suppose we have a 1-chain $u = n_1 u_1 + \cdots + n_r u_r$, with $u_i\colon \Delta_1 \to X$ and $n_i \in \mathbb{Z}$. We then put

$$\partial(u) = n_1\partial(u_1) + \cdots + n_r\partial(u_r) = \sum_{i=1}^{r} n_i(u_i(e_1) - u_i(e_0)).$$

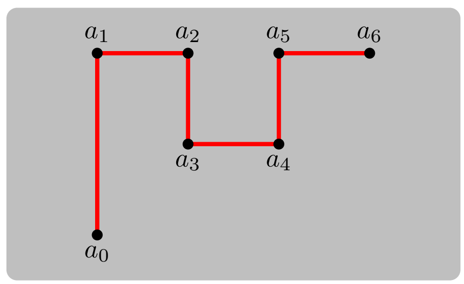
  This defines a homomorphism $\partial\colon C_1(X) \to C_0(X)$.

  Note that for a linear 1-simplex $\langle a_0, a_1\rangle$, we just have $\partial(\langle a_0, a_1\rangle) = a_1 - a_0$. Thus, in the picture below we have a 1-chain

$$u = \langle a_0, a_1\rangle + \langle a_1, a_2\rangle + \langle a_2, a_3\rangle + \langle a_3, a_4\rangle + \langle a_4, a_5\rangle + \langle a_5, a_6\rangle$$

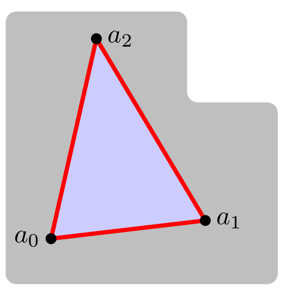
  with

$$\partial(u) = (a_1 - a_0) + (a_2 - a_1) + (a_3 - a_2) + (a_4 - a_3) + (a_5 - a_4) + (a_6 - a_5) = a_6 - a_0.$$

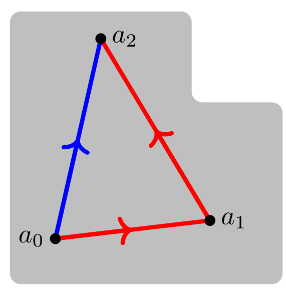


- We now consider 2-chains. For the simplest case, suppose that $X \subseteq \mathbb{R}^N$ and $u = \langle a_0, a_1, a_2\rangle$ is a linear 2-simplex. In this case, we define

$$\partial(\langle a_0, a_1, a_2\rangle) = \langle a_1, a_2\rangle - \langle a_0, a_2\rangle + \langle a_0, a_1\rangle.$$



$u = \langle a_0, a_1, a_2\rangle$



$\partial(u) = \langle a_1, a_2\rangle - \langle a_0, a_2\rangle + \langle a_0, a_1\rangle$

  The rule for nonlinear singular 2-simplices is essentially a straightforward adaptation of the linear case, but it will rely on some auxiliary definitions given below. Once we have defined $\partial(u)$ for all $u \in S_2(X)$, we will then define $\partial(u)$ for all $u \in C_2(X)$ by the rule

$$\partial(n_1 u_1 + \cdots + n_r u_r) = n_1\partial(u_1) + \cdots + n_r\partial(u_r),$$

  just as we did for singular 1-chains. This gives a homomorphism $\partial\colon C_2(X) \to C_1(X)$.
- For a linear 3-simplex $u = \langle a_0, a_1, a_2, a_3\rangle$, we will have

$$\partial(u) = \langle a_1, a_2, a_3\rangle - \langle a_0, a_2, a_3\rangle + \langle a_0, a_1, a_3\rangle - \langle a_0, a_1, a_2\rangle.$$

  For a general linear k-simplex $u = \langle a\rangle = \langle a_0, \cdots, a_k\rangle$, we will have

$$\partial(u) = \sum_{i=0}^{k} (-1)^i(\langle a\rangle \text{ with } a_i \text{ omitted }).$$

Definition 10.12. For $0 \leq i \leq n$ with $n > 0$ we define $\delta_i \colon \Delta_{n-1} \to \Delta_n$ by

$$\delta_i(t_0, \ldots, t_{n-1}) = (t_0, \ldots, t_{i-1}, 0, t_i, \ldots, t_{n-1}).$$

Equivalently, we have

$$\delta_i(t)_j = \begin{cases} t_j & \text{if } j < i \\ 0 & \text{if } j = i \\ t_{j-1} & \text{if } j > i. \end{cases}$$

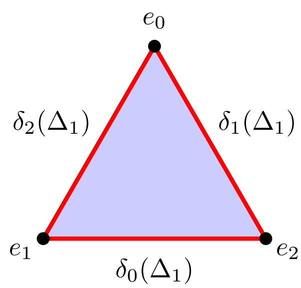
Thus, the coordinates of $\delta_i(t)$ are the same as the coordinates of t, except that we insert a zero in position i.

Example 10.13. In the case $n = 1$ we have maps $\delta_0, \delta_1 \colon \Delta_0 = \{e_0\} = \{1\} \to \Delta_1$. These are given by $\delta_0(e_0) = \delta_0(1) = (0, 1) = e_1$ and $\delta_1(e_0) = \delta_1(1) = (1, 0) = e_0$.

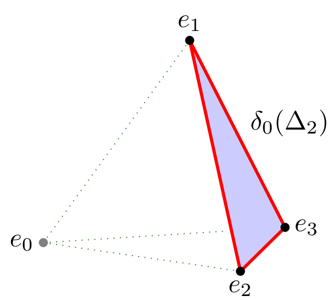
In the case $n = 2$, we have

$$\delta_0(t_0, t_1) = (0, t_0, t_1) \qquad \delta_1(t_0, t_1) = (t_0, 0, t_1) \qquad \delta_2(t_0, t_1) = (t_0, t_1, 0).$$

Thus, the image of $\delta_i \colon \Delta_1 \to \Delta_2$ is the edge of Δ_2 opposite the vertex e_i.



Similarly, in the case $n = 3$, we have a map δ_i from the triangle Δ_2 to the tetrahedron Δ_3, and the image $\delta_i(\Delta_2)$ is the face of the tetrahedron that is opposite the vertex δ_i. The case $i = 0$ is shown below.



Interactive demo

Even more generally, we see that the map $\delta_i \colon \Delta_{n-1} \to \Delta_n$ gives a homeomorphism from Δ_{n-1} to $\{t \in \Delta_n \mid t_i = 0\}$.

Definition 10.14. Consider an element $u \in S_k(X)$ (with $k > 0$), or equivalently a continuous map $u \colon \Delta_k \to X$. For each i with $0 \leq i \leq k$ we have a map $\delta_i \colon \Delta_{k-1} \to \Delta_k$ and we can compose this with u to get a map $u \circ \delta_i \colon \Delta_{k-1} \to X$, or in other words an element $u \circ \delta_i \in S_{k-1}(X) \subset C_{k-1}(X)$. We put

$$\partial(u) = \sum_{i=0}^{k} (-1)^i (u \circ \delta_i) \in C_{k-1}(X).$$

More generally, given an element $u = \sum_{p=1}^{r} a_p u_p \in C_k(X)$, we define $\partial(u) = \sum_{p=1}^{r} a_p \partial(u_p) \in C_{k-1}(X)$.

Example 10.15.

- For a singular 1-simplex $u \colon \Delta_1 \to X$ we have $\partial(u) = (u \circ \delta_0) - (u \circ \delta_1)$. Here δ_0 sends the unique point of Δ_0 to e_1, so the map $u \circ \delta_0 \colon \Delta_0 \to X$ corresponds to the point $u(e_1) \in X$. Similarly, δ_1

sends the unique point of Δ_0 to e_0, so the map $u \circ \delta_1 \colon \Delta_0 \to X$ corresponds to the point $u(e_0) \in X$. We therefore have $\partial(u) = u(e_1) - u(e_0)$, just as in Predefinition 10.11.

- Now consider a linear 2-simplex $u = \langle a_0, a_1, a_2 \rangle$, so
$$u(t_0, t_1, t_2) = t_0 a_0 + t_1 a_1 + t_2 a_2.$$
We find that
$$\begin{aligned}
(u \circ \delta_0)(t_0, t_1) &= u(0, t_0, t_1) = t_0 a_1 + t_1 a_2 = \langle a_1, a_2 \rangle (t_0, t_1) \\
(u \circ \delta_1)(t_0, t_1) &= u(t_0, 0, t_1) = t_0 a_0 + t_1 a_2 = \langle a_0, a_2 \rangle (t_0, t_1) \\
(u \circ \delta_2)(t_0, t_1) &= u(t_0, t_1, 0) = t_0 a_0 + t_1 a_1 = \langle a_0, a_1 \rangle (t_0, t_1),
\end{aligned}$$
so $u \circ \delta_0 = \langle a_1, a_2 \rangle$ and $u \circ \delta_1 = \langle a_0, a_2 \rangle$ and $u \circ \delta_2 = \langle a_0, a_1 \rangle$. This gives
$$\partial(u) = \langle a_1, a_2 \rangle - \langle a_0, a_2 \rangle + \langle a_0, a_1 \rangle,$$
just as in Predefinition 10.11. It should be clear that the same pattern works for all k, giving
$$\partial(\langle a_0, \ldots, a_k \rangle) = \sum_{i=0}^{k} (-1)^i (\langle a_0, \ldots, a_k \rangle \text{ with } a_i \text{ omitted}).$$

The following result is crucial for the development of homology theory.

Video (Proposition 10.16 to Definition 10.21)

Proposition 10.16. *For all $u \in C_k(X)$, we have $\partial^2(u) = \partial(\partial(u)) = 0$ in $C_{k-2}(X)$. Thus, the composite*
$$C_k(X) \xrightarrow{\partial} C_{k-1}(X) \xrightarrow{\partial} C_{k-2}(X)$$
is zero.

Example 10.17. Recall that we defined $C_j(X) = 0$ for $j < 0$, and any homomorphism to the zero group is automatically the zero homomorphism. Thus, the proposition has no content for $k < 2$. For the first nontrivial case, suppose that $X \subseteq \mathbb{R}^N$, and consider a linear 2-simplex $u = \langle a_0, a_1, a_2 \rangle$. We then have
$$\begin{aligned}
\partial(u) &= \langle a_1, a_2 \rangle - \langle a_0, a_2 \rangle + \langle a_0, a_1 \rangle \\
\partial^2(u) &= \partial(\langle a_1, a_2 \rangle) - \partial(\langle a_0, a_2 \rangle) + \partial(\langle a_0, a_1 \rangle) \\
&= (a_2 - a_1) - (a_2 - a_0) + (a_1 - a_0) = 0.
\end{aligned}$$
We will often use abbreviated notation for this kind of calculation, writing 012 for $\langle a_0, a_1, a_2 \rangle$ and 02 for $\langle a_0, a_2 \rangle$, for example. With this notation, the above calculation becomes
$$\partial^2(012) = \partial(12) - \partial(02) + \partial(01) = (2 - 1) - (2 - 0) + (1 - 0) = 0.$$
We now discuss $\partial^2(u)$ where $u = \langle a_0, a_1, a_2, a_3, a_4 \rangle \in C_4(X)$, using the same kind of notation. First, we have
$$\partial(u) = 1234 - 0234 + 0134 - 0124 + 0123.$$
We can write the terms of $\partial^2(u)$ in a square array, with $\partial(1234)$ in the first column, $\partial(-0234)$ in the second column, and so on. The result is as follows:
$$\begin{array}{ccccc}
+234 & -234 & +134 & -124 & +123 \\
-134 & +034 & -034 & +024 & -023 \\
+124 & -024 & +014 & -014 & +013 \\
-123 & +023 & -013 & +012 & -012
\end{array}$$
We find that the terms above the wavy line cancel in the indicated groups with the terms below the wavy line, leaving $\partial^2(u) = 0$ as claimed.

Lemma 10.18. *If* $0 \leq i < j \leq k$ *then* $\delta_j\delta_i = \delta_i\delta_{j-1} \colon \Delta_{k-2} \to \Delta_k$.

Proof. Consider a point $t = (t_0, \dots, t_{k-2}) \in \Delta_{k-2}$. To form $\delta_i(t)$, we insert a zero in position i. To form $\delta_j(\delta_i(t))$, we insert another zero in position j. Because $j > i$, inserting this second zero does not move the first zero, so we end up with zeros in positions i and j.

Similarly, to form $\delta_{j-1}(t)$, we insert a zero in position $j-1$. To form $\delta_i(\delta_{j-1}(t))$, we insert another zero in position i. As $j - 1 \geq i$ we see that the first zero is to the right of the point where we insert the second zero, so the first zero gets moved over by one space into position j. Thus, we again end up with zeros in positions i and j. In the remaining positions, we have the numbers $t_0, \dots, t_{k-2}$ in order. Thus, we have $\delta_j(\delta_i(t)) = \delta_i(\delta_{j-1}(t))$ as claimed. $\square$

Example 10.19. In the case where $(i, j, k) = (2, 4, 6)$ the claim is that $\delta_4\delta_2 = \delta_2\delta_3 \colon \Delta_4 \to \Delta_6$. Explicitly, for $t = (t_0, \dots, t_4) \in \Delta_4$ we have

$$\begin{array}{rccccccc}
 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\
\delta_2(t) = (& t_0, & t_1, & 0, & t_2, & t_3, & t_4) & \\
\delta_4(\delta_2(t)) = (& t_0, & t_1, & 0, & t_2, & 0, & t_3, & t_4) \\
\delta_3(t) = (& t_0, & t_1, & t_2, & 0, & t_3, & t_4) & \\
\delta_2(\delta_3(t)) = (& t_0, & t_1, & 0, & t_2, & 0, & t_3, & t_4)
\end{array}$$

Example 10.20. We will now prove Proposition 10.16 in the case $k = 4$. Consider a continuous map $u \colon \Delta_4 \to X$, or equivalently an element $u \in S_4(X) \subset C_4(X)$. We have

$$\partial(u) = u\delta_0 - u\delta_1 + u\delta_2 - u\delta_3 + u\delta_4.$$

We can write the terms of $\partial^2(u)$ in a square array, with $\partial(u\delta_0)$ in the first column, $\partial(-u\delta_1)$ in the second column, and so on. The result is as follows:

$$\begin{array}{ccccc}
+u\delta_0\delta_0 & -u\delta_1\delta_0 & +u\delta_2\delta_0 & -u\delta_3\delta_0 & +u\delta_4\delta_0 \\
-u\delta_0\delta_1 & +u\delta_1\delta_1 & -u\delta_2\delta_1 & +u\delta_3\delta_1 & -u\delta_4\delta_1 \\
+u\delta_0\delta_2 & -u\delta_1\delta_2 & +u\delta_2\delta_2 & -u\delta_3\delta_2 & +u\delta_4\delta_2 \\
-u\delta_0\delta_3 & +u\delta_1\delta_3 & -u\delta_2\delta_3 & +u\delta_3\delta_3 & -u\delta_4\delta_3
\end{array}$$

Lemma 10.18 gives us the following identities:

$$\begin{array}{llll}
\delta_1\delta_0 = \delta_0\delta_0 & \delta_2\delta_0 = \delta_0\delta_1 & \delta_3\delta_0 = \delta_0\delta_2 & \delta_4\delta_0 = \delta_0\delta_3 \\
 & \delta_2\delta_1 = \delta_1\delta_1 & \delta_3\delta_1 = \delta_1\delta_2 & \delta_4\delta_1 = \delta_1\delta_3 \\
 & & \delta_3\delta_2 = \delta_2\delta_2 & \delta_4\delta_2 = \delta_2\delta_3 \\
 & & & \delta_4\delta_3 = \delta_3\delta_3
\end{array}$$

Using this, we see that in the previous array, the terms above the wavy line cancel in the indicated groups with the terms below the wavy line, showing that $\partial^2(u) = 0$ as claimed. This generalises the argument for linear simplices given in Example 10.17.

Proof of Proposition 10.16. Consider a continuous map $u \colon \Delta_k \to X$, or equivalently an element $u \in S_k(X) \subset C_k(X)$. We have

$$\partial^2(u) = \sum_{j=0}^{k} (-1)^j \partial(u \circ \delta_j) = \sum_{i=0}^{k-1} \sum_{j=0}^{k} (-1)^{i+j} u \circ \delta_j \circ \delta_i.$$

We can write this as $A+B$, where

$$A = \sum_{0\leq i<j\leq k} (-1)^{i+j} u\circ\delta_j\circ\delta_i$$
$$B = \sum_{0\leq j\leq i\leq k-1} (-1)^{i+j} u\circ\delta_j\circ\delta_i.$$

Here i and j are just dummy variables, so we can rewrite B as

$$B = \sum_{0\leq q\leq p\leq k-1} (-1)^{p+q} u\circ\delta_q\circ\delta_p.$$

We now reindex again, taking $q=i$ and $p=j-1$. The condition $q\leq p$ becomes $i\leq j-1$ or equivalently $i<j$. The condition $p\leq k-1$ becomes $j-1\leq k-1$ or equivalently $j\leq k$. The sign $(-1)^{p+q}$ becomes $(-1)^{i+j-1}=-(-1)^{i+j}$. This gives

$$B = -\sum_{0\leq i<j\leq k} u\circ\delta_i\circ\delta_{j-1}.$$

However, Lemma 10.18 tells us that $\delta_i\circ\delta_{j-1}=\delta_j\circ\delta_i$ here, so $B=-A$, so $\partial^2(u)=A+B=0$ as claimed.

This proves that $\partial^2(u)=0$ whenever u is a singular k-simplex. More generally, and singular k-chain has the form $u=a_1u_1+\cdots+a_ru_r$ for some integers a_i and singular k-simplices $u_i\colon\Delta_k\to X$. We then have $\partial^2(u_i)=0$ for all i and so $\partial^2(u)=\sum_i a_i\partial^2(u_i)=0$. □

Definition 10.21.

(a) We say that an element $u\in C_k(X)$ is a *k-cycle* if $\partial(u)=0$. We write $Z_k(X)$ for the abelian group of k-cycles, so $Z_k(X)=\ker(\partial\colon C_k(X)\to C_{k-1}(X))$.
(b) We say that an element $u\in C_k(X)$ is a *k-boundary* if there exists $v\in C_{k+1}(X)$ with $\partial(v)=u$. We write $B_k(X)$ for the abelian group of k-boundaries, so $B_k(X)=\mathrm{img}(\partial\colon C_{k+1}(X)\to C_k(X))$.
(c) We note that if $u\in B_k(X)$ then $u=\partial(v)$ for some v, so $\partial(u)=\partial^2(v)=0$ by Proposition 10.16, so $u\in Z_k(X)$. This means that $B_k(X)\leq Z_k(X)$, so we can form the quotient abelian group $H_k(X)=(Z_k(X))/(B_k(X))$. We call this the k'th *homology group* of X.

Remark 10.22. The elements of $H_k(X)$ are cosets $z+B_k(X)$ with $z\in Z_k(X)$, so $z\in C_k(X)$ with $\partial(z)=0$. We will often write $[z]$ for $z+B_k(X)$. Before writing notation like $[z]$ one must check that $\partial(z)=0$; it is an error to use that notation in other cases. Note that $[z]=[z']$ iff $z-z'\in B_k(X)$ iff there exists $w\in C_{k+1}(X)$ with $\partial(w)=z-z'$.

There is essentially only one example that we can calculate directly from the definition.

Proposition 10.23. *If X consists of a single point, then $H_0(X)=\mathbb{Z}$ and $H_k(X)=0$ for $k\neq 0$.*

Proof. There is only one possible map from Δ_k to X, sending all possible points in Δ_k to the unique point of X. We call this map s_k, so $S_k(X)=\{s_k\}$ and $C_k(X)=\mathbb{Z}.s_k$ for all $k\geq 0$ (whereas $C_k(X)=0$ for $k<0$ by definition). For $k>0$ we have $\partial(s_k)=\sum_{i=0}^k(-1)^i s_k\circ\delta_i$. Here $s_k\circ\delta_i$ is a map from Δ_{k-1} to X so it can only be equal to s_{k-1}. This gives

$$\partial(s_1) = s_0 - s_0 = 0$$
$$\partial(s_2) = s_1 - s_1 + s_1 = s_1$$
$$\partial(s_3) = s_2 - s_2 + s_2 - s_2 = 0$$

and so on. In general, we have $\partial(s_{2n+1})=0$ and $\partial(s_{2n+2})=s_{2n+1}$. It follows that $B_{2n+1}(X)=Z_{2n+1}(X)=\mathbb{Z}.s_{2n+1}$ and $B_{2n+2}(X)=Z_{2n+2}(X)=0$. In particular, for all $k>0$ we have $Z_k(X)=B_k(X)$ so the quotient group $H_k(X)=(Z_k(X))/(B_k(X))$ is trivial. On the other hand, $Z_0(X)=\mathbb{Z}.s_0$ and $B_0(X)=0$ so $H_0(X)=(\mathbb{Z}.s_0)/0\simeq\mathbb{Z}$. All this can be tabulated as follows:

C_kX	$\mathbb{Z}.s_0$	$\xleftarrow[0]{\partial}$	$\mathbb{Z}.s_1$	$\xleftarrow[\simeq]{\partial}$	$\mathbb{Z}.s_2$	$\xleftarrow[0]{\partial}$	$\mathbb{Z}.s_3$	$\xleftarrow[\simeq]{\partial}$	$\mathbb{Z}.s_4$	$\xleftarrow[0]{\partial}$	$\mathbb{Z}.s_5$	$\xleftarrow[\simeq]{\partial}$	$\mathbb{Z}.s_6$
Z_kX	$\mathbb{Z}.s_0$		$\mathbb{Z}.s_1$		0		$\mathbb{Z}.s_3$		0		$\mathbb{Z}.s_5$		0
B_kX	0		$\mathbb{Z}.s_1$		0		$\mathbb{Z}.s_3$		0		$\mathbb{Z}.s_5$		0

□

Remark 10.24. We leave the following slight generalisation to the reader. Suppose that X is a finite, discrete set of points, so that every continuous map $\Delta_k \to X$ is constant. Then $H_0(X) = C_0(X) = \mathbb{Z}\{X\}$, and $H_k(X) = 0$ for $k \neq 0$.

We can also calculate $H_0(X)$ for all X.

Proposition 10.25. *There is a canonical isomorphism $H_0(X) \simeq \mathbb{Z}\{\pi_0(X)\}$ for all topological spaces X. Thus, if $|\pi_0(X)| = r$ then $H_0(X) \simeq \mathbb{Z}^r$.*

This should not be a surprise. Both $H_0(X)$ and $\mathbb{Z}\{\pi_0(X)\}$ are ways of constructing an abelian group from X, in such a way that points connected by a path give the same element of the group. We just need to check that the technical differences between these two constructions do not affect the final answer.

Proof. First note that $C_{-1}(X)$ is zero by definition, so the map $\partial\colon C_0(X) \to C_{-1}(X)$ sends everything to zero, so $Z_0(X) = C_0(X)$. This means that the quotient group $H_0(X) = Z_0(X)/B_0(X)$ is the same as $C_0(X)/B_0(X)$.

Next, let $\pi\colon X \to \pi_0(X)$ be the usual quotient map, which sends every point $x \in X$ to the corresponding path component $[x] \in \pi_0(X)$. We can extend this linearly to give a homomorphism $\pi\colon \mathbb{Z}\{X\} = \mathbb{Z}\{S_0(X)\} = C_0(X) \to \mathbb{Z}\{\pi_0(X)\}$, by the rule

$$\pi(n_1x_1 + \cdots + n_px_p) = n_1\pi(x_1) + \cdots + n_p\pi(x_p) = n_1[x_1] + \cdots + n_p[x_p] \in \mathbb{Z}\{\pi_0(X)\}.$$

We will show that π is surjective, with kernel $B_0(X)$. Assuming this, the First Isomorphism Theorem will give us an isomorphism from $C_0(X)/B_0(X) = H_0(X)$ to $\mathbb{Z}\{\pi_0(X)\}$, as required.

Next, for each path component $c \in \pi_0(X)$, we choose a point $\sigma(c) \in c$, so $c = [\sigma(c)]$. This means that the composite

$$\pi_0(X) \xrightarrow{\sigma} X \xrightarrow{\pi} \pi_0(X)$$

is the identity. We can also extend σ linearly to give a homomorphism $\sigma\colon \mathbb{Z}\{\pi_0(X)\} \to C_0(X)$ by the rule $\sigma(n_1c_1 + \cdots + n + pc_p) = n_1\sigma(c_1) + \cdots + n_p\sigma(c_p)$. In this context, we see that the composite

$$\mathbb{Z}\{\pi_0(X)\} \xrightarrow{\sigma} C_0(X) \xrightarrow{\pi} \mathbb{Z}\{\pi_0(X)\}$$

is again the identity. In particular, any element $u \in \mathbb{Z}\{\pi_0(X)\}$ is the same as $\pi(\sigma(u))$, so it is in the image of π; this proves that π is surjective.

Now suppose we have a path $v \in S_1(X)$. We then have $\partial(v) = v(e_1) - v(e_0) \in C_0(X)$, so $\pi(\partial(v)) = \pi(v(e_1)) - \pi(v(e_0)) = [v(e_1)] - [v(e_0)] \in \mathbb{Z}\{\pi_0(X)\}$. However, we have a path v joining $v(e_0)$ to $v(e_1)$, so the corresponding path components are the same, so $\pi(\partial(v)) = 0$. As everything is extended linearly, the rule $\pi(\partial(v)) = 0$ remains valid for all $v \in C_1(X)$. The image of $\partial\colon C_1(X) \to C_0(X)$ is $B_0(X)$, so this means that $\pi(B_0(X)) = 0$, or equivalently $B_0(X) \leq \ker(\pi)$.

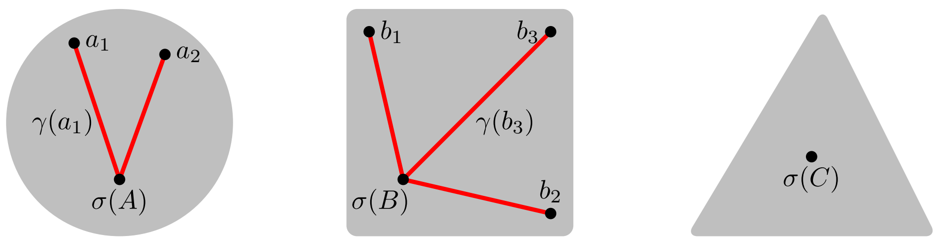
Next, consider a point $x \in X$ and the corresponding path component $c = [x] = \pi(x)$. The points x and $\sigma(c) = \sigma(\pi(x))$ both lie in the same path component c, so there must exist a path from $\sigma(\pi(x))$ to x in X. We choose such a path and call it $\gamma(x)$. This defines a function γ from X to the set $S_1(X)$ of paths in X, which we extend linearly to get a hoomorphism $\gamma\colon C_0(X) \to C_1(X)$. For any point x we know that $\gamma(x)$ runs from $\sigma(\pi(x))$ to x, so $\partial\gamma(x) = x - \sigma(\pi(x))$. As everything is extended linearly, the rule $\partial(\gamma(u)) = u - \sigma(\pi(u))$ is valid for all $u \in C_0(X)$. In particular, if $u \in \ker(\pi)$ then $\pi(u) = 0$ so this simplifies to $\partial(\gamma(u)) = u$, proving that u is in the image of ∂, or in other words $u \in B_0(X)$.

We can now conclude that π is surjective with kernel $B_0(X)$. By the First Isomorphism Theorem, there is a well-defined homomorphism $\overline{\pi}\colon H_0(X) = C_0(X)/B_0(X) \to \mathbb{Z}\{\pi_0(X)\}$ given by $\overline{\pi}(u + B_0(X)) = \pi(u)$ for all $u \in C_0(X)$, and this is in fact an isomorphism. □

Example 10.26. The above proof can be illustrated by the following diagram. It shows a space X with three path components A, B and C, so $\pi_0(X) = \{A, B, C\}$ and

$$\mathbb{Z}\{\pi_0(X)\} = \{kA + nB + mC \mid k, n, m \in \mathbb{Z}\} \simeq \mathbb{Z}^3.$$

We have chosen points $\sigma(A) \in A$ and $\sigma(B) \in B$ and $\sigma(C) \in C$, so $A = [\sigma(A)]$ and $B = [\sigma(B)]$ and $C = [\sigma(C)]$. To say the same thing in different notation, we have $\pi(\sigma(A)) = A$ and $\pi(\sigma(B)) = B$ and $\pi(\sigma(C)) = C$, so $\pi \circ \sigma = \mathrm{id}$. The points a_1 and a_2 also lie in A, so $[a_1] = [a_2] = A$, or equivalently $\pi(a_1) = \pi(a_2) = A$. The path $\gamma(a_1)$ runs from $\sigma(A) = \sigma(\pi(a_1))$ to a_1. Similarly, we have $\pi(b_1) = \pi(b_2) = \pi(b_3) = B$, and we have labelled a path $\gamma(b_3)$ running from $\sigma(\pi(b_3)) = \sigma(B)$ to b_3.



A typical example of an element of $\ker(\pi\colon C_0(X) \to \mathbb{Z}\{\pi_0(X)\})$ could be the element $u = a_1 - a_2 + b_1 + b_3 - 2b_2$. This has $\gamma(u) = \gamma(a_1) - \gamma(a_2) + \gamma(b_1) + \gamma(b_3) - 2\gamma(b_2)$, so

$$\begin{aligned}\partial(\gamma(u)) &= (a_1 - \sigma(A)) - (a_2 - \sigma(A)) + (b_1 - \sigma(B)) + (b_3 - \sigma(B)) - 2(b_2 - \sigma(B)) \\ &= a_1 - a_2 + b_1 + b_3 - 2b_2 = u,\end{aligned}$$

so $u = \partial(\gamma(u)) \in \mathrm{img}(\partial) = B_0(X)$. This illustrates the fact that $\ker(\pi) = \mathrm{img}(\partial)$, which is a key step in our proof of Proposition 10.25.

We next discuss homology classes of paths, revisiting Remark 10.10.

Lemma 10.27. *Let X be a topological space.*

(a) *For any $a \in X$ the constant path $c_a \in S_1(X) \subseteq C_1(X)$ actually lies in $B_1(X)$, so $c_a + B_1(X) = 0$ in the quotient group $C_1(X)/B_1(X)$.*
(b) *For any path $u\colon a \rightsquigarrow b$ in X with reversed path $\overline{u}\colon b \rightsquigarrow a$, we have $u + \overline{u} \in B_1(X)$ so $\overline{u} + B_1(X) = -u + B_1(X)$ in $C_1(X)/B_1(X)$.*
(c) *For any paths $u\colon a \rightsquigarrow b$ and $v\colon b \rightsquigarrow c$ we have $(u * v) + B_1(X) = (u + B_1(X)) + (v + B_1(X))$ in $C_1(X)/B_1(X)$.*

Proof. Exercise.

Video

□

Video (Path homotopy, loop homotopy and homology)

Video (Definition 10.28 and Proposition 10.29)

Definition 10.28. Let X be a topological space. A *loop* in X is a path $u\colon \Delta_1 \to X$ with $u(e_0) = u(e_1)$, so that $\partial(u) = 0$, so we have a coset $[u] = u + B_1(X) \in H_1(X)$. If $u(e_0) = u(e_1) = a$, we say that u is a *loop based at a*.

Proposition 10.29. *Let X be a path connected space, and let a be a point in X. Then for every $h \in H_1(X)$ there exists a loop u based at a with $h = [u]$. Moreover, if u and v are loops based at a then so are c_a, $\overline{u}$ and $u * v$, and we have $[c_a] = 0$ and $[\overline{u}] = -[u]$ and $[u * v] = [u] + [v]$ in $H_1(X)$.*

Proof. Let L be the subset of $H_1(X)$ consisting of classes that can be expressed as $[u]$ for some loop u based at a. We must show that this is all of $H_1(X)$.

It is clear that if u and v are loops based at a, then so are c_a, $\overline{u}$ and $u * v$. By specialising Lemma 10.27, we see that $[c_a] = 0$ and $[\overline{u}] = -[u]$ and $[u * v] = [u] + [v]$ in $H_1(X)$. It follows from this that L is a subgroup of $H_1(X)$.

Now let v be a loop based at a point $b \in X$ which may be different from a. As X is path connected, we can choose a path m from a to b. The path $u = (m * v) * \overline{m}$ is then a loop based at a, and using Lemma 10.27 again we see that

$$u + B_1(X) = m + v - m + B_1(X) = v + B_1(X),$$

or in other word $[u] = [v]$ in $H_1(X)$. This proves that L contains all loops, irrespective of the base point.

Now let h be an arbitrary element of $H_1(X)$. We can write h as $z + B_1(X)$, where z is a $\mathbb{Z}$-linear combination of paths in X. Any term with negative coefficient like $-m.u$ can be replaced by $+m.\overline{u}$ without affecting the coset, so we can assume that all coefficients are positive. Then we can replace any term like $m.u$ by u repeated m times; this gives an expression like

$$h = u_1 + \cdots + u_n + B_1(X)$$

for some list of paths u_i. As this is a homology class, the representing chain must be a cycle, so we must have $\partial(u_1 + \cdots + u_n) = 0$ in $C_0(X)$. As $\partial(u_i) = u_i(e_1) - u_i(e_0)$, this means that

$$u_1(e_1) + \cdots + u_n(e_1) = u_1(e_0) + \cdots + u_n(e_0).$$

As this is happening in the free abelian group $\mathbb{Z}\{X\}$, the terms on the left hand side must just be a permutation of those on the right hand side, so we have a permutation σ of $\{1, \ldots, n\}$ with $u_i(e_1) = u_{\sigma(i)}(e_0)$ for all i. We can now write σ as a product of disjoint cycles. If one of these cycles is $(i\ j\ k\ l)$, for example, then the paths u_i, u_j, u_k and u_l meet end-to-end and so can be joined together to form a loop $((u_i * u_j) * u_k) * u_l$ which is congruent to $u_i + u_j + u_k + u_l$ modulo $B_1(X)$. By doing this for all cycles, we see that h can be expressed as a sum of loops (probably with different basepoints). Our earlier discussion shows that each of these loops lies in L and then that the sum lies in L, so $h \in L$ as claimed. $\square$

Definition 10.30. Let $u \colon \Delta_1 \to X$ be a loop based at a. A *filling in* of u is a map $v \colon \Delta_2 \to X$ with $v \circ \delta_0 = u$ and $v \circ \delta_1 = v \circ \delta_2 = c_a$.

Lemma 10.31. *If u can be filled in, then $[u] = 0$ in $H_1(X)$.*

Proof. Let v be a filling in of u. Then

$$\partial(v) = v \circ \delta_0 - v \circ \delta_1 + v \circ \delta_2 = u - c_a + c_a = u,$$

so $u \in B_1(X)$, so $[u] = u + B_1(X) = 0$. $\square$

11. Homology of the punctured plane

Later we will prove that for all $n \geq 2$ we have

$$H_k(\mathbb{R}^n \setminus \{0\}) = H_k(S^{n-1}) = \begin{cases} \mathbb{Z} & \text{if } k = 0 \text{ or } k = n-1 \\ 0 & \text{otherwise.} \end{cases}$$

For this we will need the Mayer-Vietoris sequence, which is a very important and useful tool, but it will take some work to set that up. In this section, we outline a different approach which is more direct and elementary but which works only for $n = 2$. We will identify $\mathbb{R}^2$ with $\mathbb{C}$ and write $\mathbb{C}^\times = \mathbb{C} \setminus \{0\}$, so our main task will be to prove that $H_1(\mathbb{C}^\times) = \mathbb{Z}$.

Video (Definition 11.2 to Theorem 11.11)

Definition 11.1. Let $z \in \mathbb{C}^\times$ be a nonzero complex number. This can be expressed as $z = r\,e^{i\theta}$ for a unique pair or real numbers r, θ with $r > 0$ and $-\pi < \theta \leq \pi$. We put $\operatorname{plog}(z) = \log(r) + i\theta$, and call this the *principal logarithm* of z.

Note that $\operatorname{plog}\colon \mathbb{C}^\times \to \mathbb{C}$ and $\exp(\operatorname{plog}(z)) = z$ for all z, but plog is not continuous (because for small $\epsilon > 0$ we have $\operatorname{plog}(-1+i\epsilon) \approx i\pi$ but $\operatorname{plog}(-1-i\epsilon) \approx -i\pi$). This cannot be fixed by adjusting the definitions: there is no continuous map $f\colon \mathbb{C}^\times \to \mathbb{C}$ with $\exp(f(z)) = z$ for all z. To work around this we make the following definition:

Definition 11.2. For any $z \in \mathbb{C}^\times$ we put

$$\operatorname{LOG}(z) = \{\widetilde{z} \in \mathbb{C} \mid \exp(\widetilde{z}) = z\} = \operatorname{plog}(z) + 2\pi i\mathbb{Z}.$$

Any element of $\operatorname{LOG}(z)$ will be called a *logarithm* of z. More generally, suppose we have a topological space T and a continuous map $u\colon T \to \mathbb{C}^\times$. By a *continuous logarithm* of u we mean a continuous map $\widetilde{u}\colon T \to \mathbb{C}$ with $\exp \circ \widetilde{u} = u$, or equivalently $\widetilde{u}(t) \in \operatorname{LOG}(u(t))$ for all t.

Note that the cosets $i\pi + 2\pi i\mathbb{Z}$ and $-i\pi + 2\pi i\mathbb{Z}$ are the same, and that $\operatorname{LOG}(-1+i\epsilon)$ is close to this coset for all small ϵ, independent of whether ϵ is positive or negative. Thus, $\operatorname{LOG}(z)$ depends continuously on z even though $\operatorname{plog}(z)$ does not.

Given a continuous map $u\colon T \to \mathbb{C}^\times$, we could attempt to define a continuous logarithm of u by $\widetilde{u} = \operatorname{plog} \circ u$. This works provided that the image $u(T)$ does not touch the negative real axis where plog is discontinuous. If $u(T)$ does touch the negative real axis then it may be possible to find a continuous logarithm by a different method, but in some cases, no continuous logarithm exists.

Lemma 11.3. *Let $u\colon [0,1] \to \mathbb{C}^\times$ be continuous, and suppose that $x \in \operatorname{LOG}(u(0))$. Then there is a unique continuous logarithm $\widetilde{u}\colon [0,1] \to \mathbb{C}$ with $\widetilde{u}(0) = x$.*

We will prove this properly later when we come to discuss covering maps.

Sketch proof. If we choose N sufficently large, then when $|s-t| \leq 1/N$ the points $u(s)/u(t)$ will be close to 1 in $\mathbb{C}^\times$ and so will be far from the negative real axis where plog is discontinuous. We can thus define

$$\widetilde{u}(t) = x + \sum_{k=1}^{N} \operatorname{plog}\left(u\left(\frac{kt}{N}\right)/u\left(\frac{(k-1)t}{N}\right)\right).$$

This is a continuous function of t. When $t = 0$ we see that all the terms in the sum are $\operatorname{plog}(u(0)/u(0)) = \operatorname{plog}(1) = 0$, so $\widetilde{u}(0) = x$. In general we have

$$\exp(\widetilde{u}(t)) = \exp(x).\prod_{i=1}^{N} \frac{u(kt/N)}{u((k-1)t/N)} = u(0).\prod_{i=1}^{N} \frac{u(kt/N)}{u((k-1)t/N)},$$

and most of the terms in the product cancel out leaving only $\exp(\widetilde{u}(t)) = u(t)$. $\square$

Corollary 11.4. *Let $K \subseteq \mathbb{R}^N$ be convex, and suppose that $k \in K$. Let $u\colon K \to \mathbb{C}^\times$ be continuous, and suppose that $x \in \operatorname{LOG}(u(k))$. Then there is a unique continuous logarithm $\widetilde{u}\colon K \to \mathbb{C}$ with $\widetilde{u}(k) = x$. (In particular, this applies when $K = \Delta_d$ for some d.)*

Sketch proof. For $m \in K$ we can define $v_m\colon [0,1] \to \mathbb{C}^\times$ by $v_m(t) = u(tm + (1-t)k)$. By the lemma, there is a unique continuous logarithm $\widetilde{v}_m\colon [0,1] \to \mathbb{C}$ with $\widetilde{v}_m(0) = x$. We define $\widetilde{u}(m) = \widetilde{v}_m(1)$, so $\exp(\widetilde{u}(m)) = \exp(\widetilde{v}_m(1)) = v_m(1) = u(m)$. We have $v_k(t) = u(k)$ for all t so $\widetilde{v}_k$ must be the constant path at x so $\widetilde{u}(k) = \widetilde{v}_k(1) = x$. With some work one can check that $\widetilde{u}$ is continuous. $\square$

Definition 11.5. Given any path $u\colon \Delta_1 \to \mathbb{C}^\times$ we define

$$\omega(u) = (\widetilde{u}(e_1) - \widetilde{u}(e_0))/(2\pi i) \in \mathbb{C},$$

where $\widetilde{u}$ is any continuous logarithm of u. (This is well-defined, because any two continuous logarithms differ by a constant of the form $2n\pi i$, and the constant cancels out when we calculate $\widetilde{u}(e_1) - \widetilde{u}(e_0)$.) We then extend this linearly to get a homomorphism $\omega\colon C_1(\mathbb{C}^\times) \to \mathbb{C}$, given by

$$\omega(n_1u_1 + \cdots + n_ru_r) = n_1\omega(u_1) + \cdots + n_r\omega(u_r).$$

Example 11.6. The standard loop $u_n \colon \Delta_1 \to \mathbb{C}^\times$ of winding number n is given by $u_n(1-t,t) = \exp(2\pi i n t)$. The obvious continuous logarithm is $\widetilde{u}_n(t) = 2\pi i n t$, and using this we get $\omega(u_n) = n$.

Lemma 11.7. *For $u \colon \Delta_2 \to \mathbb{C}^\times$ we have $\omega(\partial(u)) = 0$. Thus, we have $\omega(B_1(\mathbb{C}^\times)) = 0$.*

Proof. For $i = 0,1,2$ we put $v_i = u \circ \delta_i \colon \Delta_1 \to \mathbb{C}^\times$, so $\partial(u) = v_0 - v_1 + v_2$. By Corollary 11.4, we can choose a continuous logarithm $\widetilde{u} \colon \Delta_2 \to \mathbb{C}$ for u. We then note that the map $\widetilde{v}_i = \widetilde{u} \circ \delta_i \colon \Delta_1 \to \mathbb{C}$ is a continuous logarithm for for v_i, so $\omega(v_i) = (\widetilde{v}_i(e_1) - \widetilde{v}_i(e_0))/(2\pi i)$. This gives

$$2\pi i \omega(\partial(u)) = (\widetilde{v}_0(e_1) - \widetilde{v}_0(e_0)) - (\widetilde{v}_1(e_1) - \widetilde{v}_1(e_0)) + (\widetilde{v}_2(e_1) - \widetilde{v}_2(e_0)).$$

However, we have

$$\begin{aligned} \delta_0(e_1) &= e_2 & \delta_1(e_1) &= e_2 & \delta_2(e_1) &= e_1 \\ \delta_0(e_0) &= e_1 & \delta_1(e_0) &= e_0 & \delta_2(e_0) &= e_0, \end{aligned}$$

so the above expression becomes

$$2\pi i \omega(\partial(u)) = (\widetilde{u}(e_2) - \widetilde{u}(e_1)) - (\widetilde{u}(e_2) - \widetilde{u}(e_0)) + (\widetilde{u}(e_1) - \widetilde{u}(e_0)) = 0,$$

so $\omega(\partial(u)) = 0$ as claimed. More generally, if $u \in C_2(\mathbb{C}^\times)$ then $u = n_1 u_1 + \dots + n_r u_r$ for some integers n_i and maps $u_i \colon \Delta_2 \to \mathbb{C}^\times$, and this gives

$$\omega(\partial(u)) = \sum_i n_i \omega(\partial(u_i)) = \sum_i n_i.0 = 0$$

as before. Thus, if $w \in B_1(\mathbb{C}^\times)$ then $w = \partial(u)$ for some $u \in C_2(\mathbb{C}^\times)$ giving $\omega(w) = \omega(\partial(u)) = 0$ as claimed. □

Definition 11.8. We now define homomorphisms $\beta \colon \mathbb{C} \to \mathbb{C}^\times$ and $\gamma \colon C_0(\mathbb{C}^\times) \to \mathbb{C}^\times$ by $\beta(z) = \exp(2\pi i z)$ and

$$\gamma(n_1 z_1 + \dots + n_r z_r) = z_1^{n_1} z_2^{n_2} \cdots z_r^{n_r}.$$

(Here we regard $\mathbb{C}$ and $C_0(\mathbb{C}^\times)$ as groups under addition and $\mathbb{C}^\times$ as a group under multiplication, so to say that β and γ are homomorphisms means that $\beta(w+z) = \beta(w)\beta(z)$ and $\gamma(u+v) = \gamma(u)\gamma(v)$; it is easy to see that both of these identities are valid.)

Lemma 11.9. *The following square commutes (or in other words, $\beta(\omega(u)) = \gamma(\partial(u))$ for all $u \in C_1(\mathbb{C}^\times)$).*

$$\begin{tikzcd} C_1(\mathbb{C}^\times) \arrow[r, "\partial"] \arrow[d, "\omega"'] & C_0(\mathbb{C}^\times) \arrow[d, "\gamma"] \\ \mathbb{C} \arrow[r, "\beta"'] & \mathbb{C}^\times \end{tikzcd}$$

Proof. In general u will be a $\mathbb{Z}$-linear combination of paths in $\mathbb{C}^\times$, but all the maps are homomorphisms, so it will be enough to consider the case where $u \colon \Delta_1 \to \mathbb{C}^\times$ is just a single path. We then have $\partial(u) = u(e_1) - u(e_0) \in C_0(\mathbb{C}^\times)$ and so $\gamma(\partial(u)) = u(e_1)/u(e_0) \in \mathbb{C}^\times$. Now choose a continuous logarithm $\widetilde{u} \colon \Delta_1 \to \mathbb{C}$ for u. By definition we have $\omega(u) = (\widetilde{u}(e_1) - \widetilde{u}(e_0))/(2\pi i)$, so

$$\beta(\omega(u)) = \exp(\widetilde{u}(e_1) - \widetilde{u}(e_0)) = \exp(\widetilde{u}(e_1))/\exp(\widetilde{u}(e_0)) = u(e_1)/u(e_0) = \gamma(\partial(u)).$$

□

Corollary 11.10. *We have $\omega(Z_1(\mathbb{C}^\times)) = \mathbb{Z}$ and $\omega(B_1(\mathbb{C}^\times)) = 0$, so ω induces a homomorphism $\overline{\omega} \colon H_1(\mathbb{C}^\times) \to \mathbb{Z}$ given by $\overline{\omega}(z + B_1(\mathbb{C}^\times)) = \omega(z)$.*

Proof. Suppose that $z \in Z_1(\mathbb{C}^\times)$, so $\partial(z) = 0$, so $\gamma(\partial(z)) = \gamma(0) = 1$. By the Lemma we then have $\beta(\omega(z)) = 1$, or in other words $\exp(2\pi i \omega(z)) = 1$, so $\omega(z) \in \mathbb{Z}$. As in Example 11.6 we also have standard loops $u_n \in Z_1(\mathbb{C}^\times)$ with $\omega(u_n) = n$, so the image $\omega(Z_1(\mathbb{C}^\times))$ is the whole group $\mathbb{Z}$. We saw in Lemma 11.7 that $\omega(B_1(\mathbb{C}^\times)) = 0$, and it follows that the rule $\overline{\omega}(z + B_1(\mathbb{C}^\times)) = \omega(z)$ gives a well-defined homomorphism from the quotient group $H_1(\mathbb{C}^\times) = Z_1(\mathbb{C}^\times)/B_1(\mathbb{C}^\times)$ to $\mathbb{Z}$. □

Theorem 11.11. *The homomorphism $\overline{\omega} \colon H_1(\mathbb{C}^\times) \to \mathbb{Z}$ is an isomorphism.*

Proof. We have already remarked that $\overline{\omega}(u_n + B_1(\mathbb{C}^\times)) = \omega(u_n) = n$ for all $n \in \mathbb{Z}$; this shows that $\overline{\omega}$ is surjective. Now suppose we have $h \in H_1(\mathbb{C}^\times)$ with $\overline{\omega(h)} = 0$. By Proposition 10.29, we can find a loop $u\colon \Delta_1 \to \mathbb{C}^\times$ based at $1 \in \mathbb{C}^\times$ with $h = [u]$. By Lemma 11.3, there is a unique continuous logarithm $\widetilde{u}\colon \Delta_1 \to \mathbb{C}$ with $\widetilde{u}(e_0) = 0$. We then have $\omega(u) = (\widetilde{u}(e_1) - \widetilde{u}(e_0))/(2\pi i) = \widetilde{u}(e_1)/(2\pi i)$. However, we also know that $\omega(u) = \overline{\omega}([u]) = \overline{\omega}(h) = 0$, so we must have $\widetilde{u}(e_1) = 0$ as well. We now define $\widetilde{v}\colon \Delta_2 \to \mathbb{C}$ by

$$\widetilde{v}(t_0, t_1, t_2) = \begin{cases} (1 - t_0)\widetilde{u}(t_1/(1 - t_0), t_2/(1 - t_0)) & \text{if } t_0 < 1 \\ 0 & \text{if } t_0 = 1. \end{cases}$$

We leave it to the reader to check that $\widetilde{v}$ is continuous even at e_0. (A full proof of a more general fact will be given later.) We then define $v = \exp \circ \widetilde{v}\colon \Delta_2 \to \mathbb{C}^\times$. It is easy to see that $\widetilde{v}$ is a filling in for $\widetilde{u}$, and thus that v is a filling in for u, so $[u] = 0$ in $H_1(X)$ by Lemma 10.31, or in other words $h = 0$. This proves that $\overline{\omega}$ is also injective, and so is an isomorphism as claimed. $\square$

12. Abelian groups

To go further with homology, we will need some additional theory of abelian groups.

Video (Lemma 12.1 to Corollary 12.3)

We will almost always use additive notation for abelian groups, so the group operation will be denoted by $a + b$, the identity element by 0, and the inverse of a by $-a$. The product of groups A and B will usually be written $A \oplus B$ rather than $A \times B$.

The basic examples of finitely generated abelian groups are $\mathbb{Z}$ and $\mathbb{Z}/n$. Recall that the elements of $\mathbb{Z}/n$ are the cosets $\overline{i} = i + n\mathbb{Z}$. These can be defined for all $i \in \mathbb{Z}$, but they repeat with period n, so $\{\overline{0}, \dots, \overline{n-1}\}$ is a complete list of elements.

Lemma 12.1. *Let n, m and k be integers such that $n, m > 0$ and kn is divisible by m. Then there is a well-defined homomorphism $\phi\colon \mathbb{Z}/n \to \mathbb{Z}/m$ given by $\phi(i + n\mathbb{Z}) = ik + m\mathbb{Z}$.*

Proof. We can certainly define a homomorphism $\phi_0\colon \mathbb{Z} \to \mathbb{Z}/m$ by $\phi_0(i) = ik + m\mathbb{Z}$. By assumption we have $kn = qm$ for some q, so $\phi_0(in) = ink + m\mathbb{Z} = iqm + m\mathbb{Z}$, and this is the same as $0 + m\mathbb{Z}$ because $iqm \in m\mathbb{Z}$. This shows that $n\mathbb{Z} \leq \ker(\phi_0)$, so we have a well-defined homomorphism ϕ as described. $\square$

Proposition 12.2 (The Chinese Remainder Theorem)**.** *Suppose that n and m are positive integers that are coprime. Then there is an isomorphism $\phi\colon \mathbb{Z}/nm \to \mathbb{Z}/n \oplus \mathbb{Z}/m$ given by*

$$\phi(i + nm\mathbb{Z}) = (i + n\mathbb{Z},\ i + m\mathbb{Z}).$$

Proof. It is clear that the above formula gives a well-defined homomorphism. Next, as n and m are coprime, we can find integers a, b such that $an + bm = 1$. By the lemma, there are well-defined homomorphisms $\alpha\colon \mathbb{Z}/n \to \mathbb{Z}/nm$ and $\beta\colon \mathbb{Z}/m \to \mathbb{Z}/nm$ given by $\alpha(j + n\mathbb{Z}) = bmj + nm\mathbb{Z}$ and $\beta(k + m\mathbb{Z}) = ank + nm\mathbb{Z}$. We can combine these to define $\psi\colon \mathbb{Z}/n \oplus \mathbb{Z}/m \to \mathbb{Z}/nm$ by $\psi(u, v) = \alpha(u) + \beta(v)$. Taking account of the identity $an + bm = 1$, this can be written more explicitly as

$$\psi(j + n\mathbb{Z}, k + m\mathbb{Z}) = bmj + ank + nm\mathbb{Z} = j + an(k - j) + nm\mathbb{Z} = k + bm(j - k) + nm\mathbb{Z}.$$

From the third and fourth expressions we see that this is equal to j mod n and equal to k mod m, so $\phi \circ \psi = \mathrm{id}$. On the other hand, we have

$$\psi(\phi(i + nm\mathbb{Z})) = \psi(i + n\mathbb{Z}, i + m\mathbb{Z}) = bmi + ani + nm\mathbb{Z} = (an + bm)i + nm\mathbb{Z} = i + nm\mathbb{Z},$$

so $\psi \circ \phi$ is also the identity. $\square$

Corollary 12.3. *Suppose that $n = p_1^{v_1} \cdots p_r^{v_r}$, where $p_1, \dots, p_r$ are distinct primes and $v_1, \dots, v_r \geq 0$. Then there is an isomorphism*

$$\phi\colon \mathbb{Z}/n \to \mathbb{Z}/p_1^{v_1} \oplus \cdots \oplus \mathbb{Z}/p_r^{v_r}$$

given by

$$\phi(i + n\mathbb{Z}) = (i + p_1^{v_1}\mathbb{Z}, \dots, i + p_r^{v_r}\mathbb{Z}).$$

Proof. This follows from Proposition 12.2 by induction on r. $\square$

Video (Definition 12.4 to Example 12.11)

Definition 12.4. Let A be an abelian group, and let $L = (a_1, \dots, a_r)$ be a finite list of elements of A. Put

$$B = \{n_1a_1 + \cdots + n_ra_r \mid n_1, \dots, n_r \in \mathbb{Z}\}.$$

This is easily seen to be a subgroup of A. We say that L *generates* A if $B = A$. We say that A is *finitely generated* if there is a finite list that generates it.

Example 12.5.

(a) The list (1) generates $\mathbb{Z}$.
(b) For $r \geq 0$, the standard basis vectors $(e_1, \dots, e_r)$ generate $\mathbb{Z}^r$.
(c) The list $(1 + n\mathbb{Z})$ generates $\mathbb{Z}/n$.
(d) If $(a_1, \dots, a_r)$ generates A and $(b_1, \dots, b_s)$ generates B then the list

$$((a_1, 0), \dots, (a_r, 0),\ (0, b_1), \dots, (0, b_s))$$

generates $A \oplus B$. Thus, if A and B are finitely generated, then so is $A \oplus B$.
(e) We can now see by induction that any group of the form

$$A = \mathbb{Z}^r \oplus \mathbb{Z}/n_1 \oplus \cdots \oplus \mathbb{Z}/n_s$$

is finitely generated.
(f) Suppose again that $(a_1, \dots, a_r)$ generates A. Let B be a subgroup of A, and let $\pi\colon A \to A/B$ be the usual projection homomorphism, given by $\pi(x) = x + B$. It is then easy to see that the list $(\pi(a_1), \dots, \pi(a_r))$ generates A/B, so A/B is again finitely generated.
(g) If A is a finite group then it is certainly finitely generated, because we can just take the full list of elements as generators.

Example 12.6. Consider the polynomial ring $\mathbb{Z}[x]$ as an abelian group under addition; we will show that this is not finitely generated. Let $L = (f_1, \dots, f_r)$ be a finite list of elements of $\mathbb{Z}[x]$, and let B be the subgroup of $\mathbb{Z}$-linear combinations of this list. Let d be the maximum of the degrees of all the polynomials f_i. Any element of B has the form $\sum_i n_if_i$ for some integers n_i, and so has degree at most d. It follows that $x^{d+1} \notin B$, so B is not all of $\mathbb{Z}[x]$, so L does not generate $\mathbb{Z}[x]$.

Example 12.7. Consider the field $\mathbb{Q}$ as an abelian group under addition; we will show that this is not finitely generated. Let $L = (q_1, \dots, q_r)$ be a finite list of elements of $\mathbb{Q}$, and let B be the subgroup of $\mathbb{Z}$-linear combinations of this list. We can write q_i as a_i/b_i for some $a_i, b_i \in \mathbb{Z}$ with $b_i > 0$. Put $b = b_1b_2 \cdots b_r$, so $bq_i \in \mathbb{Z}$ for all i. It follows easily that $bx \in \mathbb{Z}$ for all $x \in B$, so $1/(2b) \notin B$, so B is not all of $\mathbb{Q}$, so L does not generate $\mathbb{Q}$.

We now quote two theorems whose proofs can be found in almost any textbook on abstract algebra.

Proposition 12.8. *Let A be a finitely generated abelian group, and let B be a subgroup of A. Then B is also finitely generated.* □

Theorem 12.9. *Let A be a finitely generated abelian group. Then A can be expressed (up to isomorphism) as the direct sum of a finite list of summands of the form $\mathbb{Z}$ or $\mathbb{Z}/p^v$ (with p prime and $v > 0$). Moreover, the list of summands is unique up to order.* □

Example 12.10. Put $A = \mathbb{Z}^3$ and $B = \{(2n, 2n, 2n) \mid n \in \mathbb{Z}\} < A$ and $C = A/B$. We claim that $C \simeq \mathbb{Z}^2 \oplus \mathbb{Z}/2$. Indeed, we can define maps

$$\mathbb{Z}^2 \oplus \mathbb{Z}/2 \xrightarrow{\phi} C \xrightarrow{\psi} \mathbb{Z}^2 \oplus \mathbb{Z}/2$$

by

$$\phi(i, j, k + 2\mathbb{Z}) = (i + k, j + k, k) + C$$
$$\psi((p, q, r) + C) = (p - r, q - r, r + 2\mathbb{Z})$$

It is an exercise to check that these are well-defined and inverse to each other.

Example 12.11. Consider an abelian group A with $|A| = 72 = 2^3 3^2$. Theorem 12.9 tells us that this can be decomposed as a direct sum of groups of the form $\mathbb{Z}/p^v$, where p is prime and $v > 0$. If $\mathbb{Z}/p^v$ is a summand, then the order p^v must divide $|A|$. Thus, the only possible summands are $\mathbb{Z}/2$, $\mathbb{Z}/4$, $\mathbb{Z}/8$, $\mathbb{Z}/3$ and $\mathbb{Z}/9$. The only possibilities for the 2-power summands are $\mathbb{Z}/8$ or $\mathbb{Z}/2 \oplus \mathbb{Z}/4$ or $\mathbb{Z}/2 \oplus \mathbb{Z}/2 \oplus \mathbb{Z}/2$. The only possibilities for the 3-power summands are $\mathbb{Z}/9$ or $\mathbb{Z}/3 \oplus \mathbb{Z}/3$. Thus, there are six possibilities for A:

$$\begin{aligned}
A_1 &= \mathbb{Z}/8 \oplus \mathbb{Z}/9 \\
A_2 &= \mathbb{Z}/2 \oplus \mathbb{Z}/4 \oplus \mathbb{Z}/9 \\
A_3 &= \mathbb{Z}/2 \oplus \mathbb{Z}/2 \oplus \mathbb{Z}/2 \oplus \mathbb{Z}/9 \\
A_4 &= \mathbb{Z}/8 \oplus \mathbb{Z}/3 \oplus \mathbb{Z}/3 \\
A_5 &= \mathbb{Z}/2 \oplus \mathbb{Z}/4 \oplus \mathbb{Z}/3 \oplus \mathbb{Z}/3 \\
A_6 &= \mathbb{Z}/2 \oplus \mathbb{Z}/2 \oplus \mathbb{Z}/2 \oplus \mathbb{Z}/3 \oplus \mathbb{Z}/3.
\end{aligned}$$

One might think that there were additional possibilities like $\mathbb{Z}/36 \oplus \mathbb{Z}/2$, but the Chinese Remainder Theorem gives $\mathbb{Z}/36 \simeq \mathbb{Z}/4 \oplus \mathbb{Z}/9$, so $\mathbb{Z}/36 \oplus \mathbb{Z}/2 \simeq A_2$. Similarly, we have $\mathbb{Z}/72 \simeq A_1$.

Video (Definition 12.12 to Lemma 12.20)

The following definition will turn out to be very important.

Definition 12.12. Consider a pair of homomorphisms $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$ of abelian groups. Recall that we have subgroups

$$\begin{aligned}
\operatorname{img}(\alpha) &= \{\alpha(a) \mid a \in A\} = \{b \in B \mid b = \alpha(a) \text{ for some } a \in A\} \leq B \\
\ker(\beta) &= \{b \in B \mid \beta(b) = 0\} \leq B.
\end{aligned}$$

We say that the pair is *exact* if $\operatorname{img}(\alpha) = \ker(\beta)$. We also say that the pair is *short exact* if it is exact, and α is injective, and β is surjective.

Remark 12.13. It is standard that α is injective iff $\ker(\alpha) = 0$ and β is surjective iff $\operatorname{img}(\beta) = C$. Thus, the sequence is short exact iff we have $\ker(\alpha) = 0$ and $\operatorname{img}(\alpha) = \ker(\beta)$ and $\operatorname{img}(\beta) = C$.

Example 12.14.

(a) For any abelian groups A and B we have a short exact sequence $A \xrightarrow{j} A \oplus B \xrightarrow{q} B$ given by $j(a) = (a, 0)$ and $q(a, b) = b$.
(b) For any $n > 0$ we have a short exact sequence $\mathbb{Z} \xrightarrow{\mu} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/n$ given by $\mu(x) = nx$ and $\pi(y) = y + n\mathbb{Z}$.
(c) For any $n, m > 0$ we have a short exact sequence $\mathbb{Z}/n \xrightarrow{\alpha} \mathbb{Z}/nm \xrightarrow{\beta} \mathbb{Z}/m$ given by $\alpha(i + n\mathbb{Z}) = mi + nm\mathbb{Z}$ and $\beta(j + nm\mathbb{Z}) = j + m\mathbb{Z}$.
(d) If we define $\alpha \colon \mathbb{Z}^2 \to \mathbb{Z}^2$ by $\alpha(x, y) = (y, 0)$, then the sequence $\mathbb{Z}^2 \xrightarrow{\alpha} \mathbb{Z}^2 \xrightarrow{\alpha} \mathbb{Z}^2$ is exact but not short exact.

Proposition 12.15.

(a) *A sequence $0 \to B \xrightarrow{\beta} C$ is exact iff β is injective.*
(b) *A sequence $A \xrightarrow{\alpha} B \to 0$ is exact iff α is surjective.*
(c) *If $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$ is exact, then $\alpha = 0$ iff β is injective.*
(d) *If $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$ is exact, then $\beta = 0$ iff α is surjective.*
(e) *A sequence $0 \to A \to 0$ is exact iff $A = 0$.*

Proof.

(a) Consider a sequence $0 \to B \to C$. The image of the first map can only be zero, so the sequence is exact iff $\ker(\beta) = 0$, which means that β is injective.
(b) Consider a sequence $A \xrightarrow{\alpha} B \to 0$. The kernel of the second map is all of B, so the sequence is exact iff $\operatorname{img}(\alpha) = B$, which means that α is surjective.

(c) Consider an exact sequence $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$. It is clear that α is the zero homomorphism iff $\mathrm{img}(\alpha)$ is the zero subgroup of B, and β is injective iff $\ker(\beta) = 0$. As $\mathrm{img}(\alpha)$ and $\ker(\beta)$ are the same, we see that $\alpha = 0$ iff β is injective.
(d) Similarly, it is clear that α is surjective iff $\mathrm{img}(\alpha) = B$, and $\beta = 0$ iff $\ker(\beta) = B$. As $\mathrm{img}(\alpha)$ and $\ker(\beta)$ are the same, we see that α is surjective iff $\beta = 0$.
(e) Consider a sequence $0 \to A \to 0$. The image of the first map is 0, and the kernel of the second map is A. The sequence is exact iff these two subgroups are the same, iff $A = 0$.

□

Definition 12.16. Consider a longer sequence

$$\cdots A_{i-2} \xrightarrow{\alpha_{i-2}} A_{i-1} \xrightarrow{\alpha_{i-1}} A_i \xrightarrow{\alpha_i} A_{i+1} \xrightarrow{\alpha_{i+1}} A_{i+2} \cdots$$

We say that this is *exact at* A_i if $\mathrm{img}(\alpha_{i-1}) = \ker(\alpha_i)$, so the subsequence

$$A_{i-1} \xrightarrow{\alpha_{i-1}} A_i \xrightarrow{\alpha_i} A_{i+1}$$

is exact. We say that the whole sequence is exact if it is exact at A_i for all i.

Example 12.17. Consider a sequence

$$0 \to A \xrightarrow{\alpha} B \xrightarrow{\beta} C \to 0.$$

This is exact at A iff α is injective, and exact at C iff β is surjective. Thus, the whole sequence is exact iff $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$ is short exact.

Example 12.18. Consider the group $\mathbb{Z}/4 = \{0, 1, 2, 3\}$ and the homomorphism $\alpha \colon \mathbb{Z}/4 \to \mathbb{Z}/4$ given by $\alpha(a) = 2a$. We then find that $\mathrm{img}(\alpha) = \ker(\alpha) = \{0, 2\}$, so the sequence

$$\cdots \xrightarrow{\alpha} \mathbb{Z}/4 \xrightarrow{\alpha} \mathbb{Z}/4 \xrightarrow{\alpha} \mathbb{Z}/4 \xrightarrow{\alpha} \mathbb{Z}/4 \xrightarrow{\alpha} \mathbb{Z}/4 \xrightarrow{\alpha} \mathbb{Z}/4 \xrightarrow{\alpha} \cdots$$

is exact.

Lemma 12.19. *A sequence* $0 \to A \xrightarrow{\alpha} B \to 0$ *is exact iff* α *is an isomorphism.*

Proof. The sequence is exact iff $\ker(\alpha)$ is the image of the map $0 \to A$ (so $\ker(\alpha) = 0$) and $\mathrm{img}(\alpha)$ is the kernel of the map $B \to 0$ (which is all of B). The first condition means that α is injective, and the second means that α is surjective, so the two conditions together mean that α is an isomorphism. □

Lemma 12.20. *Let* $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$ *be a short exact sequence. Then* A *is isomorphic to the subgroup* $\mathrm{img}(\alpha) = \alpha(A) \leq B$*, and the quotient group* $B/\alpha(A)$ *is isomorphic to* C*. Moreover, the group* B *is finite iff* A *and* C *are both finite, and if so, we have* $|B| = |A||C|$*.*

Proof. As the sequence is assumed to be short exact, we know that α is injective and β is surjective and $\mathrm{img}(\alpha) = \ker(\beta)$.

We can considar α as a homomorphism from A to $\alpha(A)$. In this context it is surjective (by definition of $\alpha(A)$) and injective (by assumption) so it is an isomorphism, as required.

Next, as $\beta \colon B \to C$ is a surjective homomorphism, the First Isomorphism Theorem tells us that it induces an isomorphism $B/\ker(\beta) \to C$. By the exactness assumption we have $\ker(\beta) = \alpha(A)$, so we have $B/\alpha(A) \simeq C$ as claimed.

Suppose that A and C are finite. We can divide B into cosets for the subgroup $\alpha(A)$. As $\alpha \colon A \to \alpha(A)$ is an isomorphism we have $|\alpha(A)| = |A|$, so each coset has size $|A|$. There is one coset for each element of the group $B/\alpha(A) \simeq C$, so the number of cosets is $|C|$. Thus, the total number of elements is $|B| = |A||C| < \infty$.

Conversely, suppose that B is finite. As A is isomorphic to a subgroup of B, it is also finite. As C is isomorphic to a quotient of B, it is also finite. We can then go back to the last paragraph to see that $|B| = |A||C|$ again. □

Video (Definition 13.1 to Example 13.7)

Given a topological space X, we previously defined a system of groups $C_k(X) = \mathbb{Z}\{S_k(X)\}$ and homomorphisms $\partial\colon C_k(X) \to C_{k-1}(X)$. In Proposition 10.16 we showed that $\partial^2 = 0$, and this allowed us to define the homology groups $H_k(X)$. In this section we will place all those constructions in a wider context, which will be useful for calculating $H_k(X)$.

Definition 13.1. A *chain complex* is a sequence of abelian groups and homomorphisms like

$$\cdots \leftarrow A_{-2} \xleftarrow{d_{-1}} A_{-1} \xleftarrow{d_0} A_0 \xleftarrow{d_1} A_1 \xleftarrow{d_2} A_2 \xleftarrow{d_3} \cdots,$$

such that $d_i \circ d_{i+1} = 0\colon A_{i+1} \to A_{i-1}$ for all i. We will often suppress the indices and just write $d^2 = 0$ instead of $d_i \circ d_{i+1} = 0$. We now put

$$Z_i(A) = \ker(d_i\colon A_i \to A_{i-1}) \leq A_i$$
$$B_i(A) = \operatorname{img}(d_{i+1}\colon A_{i+1} \to A_i) \leq A_i.$$

The elements of $Z_i(A)$ are called *cycles*, and the elements of $B_i(A)$ are called *boundaries*. As in Remark 12.13, the condition $d_i \circ d_{i+1} = 0$ means that $B_i(A) \leq Z_i(A)$, so every boundary is a cycle. Because of this, it is legitimate to define

$$H_i(A) = Z_i(A)/B_i(A).$$

This is called the *i'th homology group* of the complex. Given a cycle $z \in Z_i(A)$, we write $[z]$ for the coset $z + B_i(A) \in H_i(A)$, and call this the *homology class* of z. Thus $[z]$ is defined iff $dz = 0$, and $[z] = [z']$ iff $z' = z + dy$ for some y. The homomorphisms d_i are called *differentials*.

Remark 13.2. We use the notation A_* to refer to the whole chain complex as a single object. We also write $Z_*(A)$ for the whole system of groups $Z_i(A)$, and similarly for $B_*(A)$ and $H_*(A)$.

Remark 13.3. The content of Section 10 can now be expressed as follows: the groups $C_*(X)$ form a chain complex (with differential ∂), and the homology groups of the space X are defined to be the homology groups of the chain complex $C_*(X)$.

Remark 13.4. Note that the quotient $H_i(A) = Z_i(A)/B_i(A)$ is zero iff $Z_i(A) = B_i(A)$ iff $\ker(d_i\colon A_i \to A_{i-1}) = \operatorname{img}(d_{i+1}\colon A_{i+1} \to A_i)$ iff the sequence $A_{i+1} \xrightarrow{d_{i+1}} A_i \xrightarrow{d_i} A_{i-1}$ is exact. Thus, the size of the group $H_i(A)$ can be regarded as a measure of how badly the chain complex fails to be exact at A_i.

Example 13.5. Consider a chain complex in which $A_i = 0$ for $i > 0$, so the complex just has the form

$$\leftarrow 0 \xleftarrow{d_0} A_0 \xleftarrow{d_1} 0 \xleftarrow{d_2} 0 \xleftarrow{d_3} \cdots$$

Every differential either starts or ends at the zero group, so all differentials are zero. We find that $Z_0(A) = A_0$ and $B_0(A) = 0$ so $H_0(A) = A_0/0 = A_0$. For $i \neq 0$ we just have $Z_i(A) = 0$ and $B_i(A) = 0$ and $H_i(A) = 0$.

Example 13.6. Now consider a chain complex in which $A_i = 0$ for $i > 1$, so the complex just has the form

$$\leftarrow 0 \xleftarrow{d_0} A_0 \xleftarrow{d_1} A_1 \xleftarrow{d_2} 0 \xleftarrow{d_3} \cdots$$

It is clear that $d_i = 0$ for all $i \neq 1$, but d_1 may be nonzero. It follows that

$$\begin{array}{lll} Z_0(A) = A_0 & B_0(A) = d_1(A_1) & H_0(A) = A_0/d_1(A_1) \\ Z_1(A) = \ker(d_1) & B_1(A) = 0 & H_1(A) = \ker(d_1). \end{array}$$

Example 13.7. We now consider a specific chain complex A of the type discussed in Example 13.6, so $A_k = 0$ for $k > 1$. We fix $n > 0$, and take A_0 and A_1 to be free abelian groups as follows:

$$A_0 = \mathbb{Z}\{v_0, \ldots, v_{n-1}\}$$
$$A_1 = \mathbb{Z}\{e_0, \ldots, e_{n-1}\}.$$

These indices are supposed to be read modulo n, so $v_n = v_0$ and $v_{-1} = v_{n-1}$ and so on. We define $d_1\colon A_1 \to A_0$ by $d_1(e_i) = v_{i+1} - v_i$ (so in particular $d_1(e_{n-1}) = v_0 - v_{n-1}$). We claim that the groups $H_0(A)$ and $H_1(A)$ are both isomorphic to $\mathbb{Z}$.

The easiest way to see this is to introduce some alternative bases for these groups. We put

$$\begin{aligned} e'_i &= e_0 + e_1 + \cdots + e_i = \sum_{j=0}^{i} e_j \qquad (\text{ for } 0 \leq i < n) \\ v'_0 &= v_0 \\ v'_i &= v_i - v_0 \qquad (\text{ for } 0 < i < n). \end{aligned}$$

For example, when $n = 4$ we have

$$\begin{array}{llll} e'_0 = e_0 & e_0 = e'_0 & v'_0 = v_0 & v_0 = v'_0 \\ e'_1 = e_0 + e_1 & e_1 = e'_1 - e'_0 & v'_1 = v_1 - v_0 & v_1 = v'_1 + v'_0 \\ e'_2 = e_0 + e_1 + e_2 & e_2 = e'_2 - e'_1 & v'_2 = v_2 - v_0 & v_2 = v'_2 + v'_0 \\ e'_3 = e_0 + e_1 + e_2 + e_3 & e_3 = e'_3 - e'_2 & v'_3 = v_3 - v_0 & v_3 = v'_3 + v'_0 \end{array}$$

It is easy to see that the list $(e'_0, \ldots, e'_{n-1})$ is a basis for A_1 over $\mathbb{Z}$, and $(v'_0, \ldots, v'_{n-1})$ is a basis for A_0. We also have

$$d_1(e'_i) = (v_1 - v_0) + (v_2 - v_1) + \cdots + (v_{i+1} - v_i).$$

Most of the terms cancel, giving $d_1(e'_i) = v_{i+1} - v_0$. For $0 \leq i < n - 1$ we can write this as $d_1(e'_i) = v'_{i+1}$. However, v_n is the same as v_0, so $d'_1(e'_{n-1}) = 0$. In summary, we have

$$d_1(e'_0) = v'_1 \qquad d_1(e'_1) = v'_2 \qquad d_1(e'_2) = v'_3 \qquad \cdots \qquad d_1(e'_{n-2}) = v'_{n-1} \qquad d_1(e'_{n-1}) = 0.$$

From this it is clear that $\ker(d_1) = \mathbb{Z}.e'_{n-1}$ and $\operatorname{img}(d_1) = \mathbb{Z}\{v'_1, \ldots, v'_{n-1}\}$ so $A_0/\operatorname{img}(d_1) = \mathbb{Z}.v'_0$. In other words, we have $H_1(A) = \mathbb{Z}.e'_{n-1}$ and $H_0(A) = \mathbb{Z}.v'_0$.

We will typically express this answer by writing $H_*(A) = (\mathbb{Z}, \mathbb{Z})$. The first group listed is H_0, the second one is H_1, and it is implicit that all subsequent groups are zero.

Remark 13.8. The method used above is fairly typical for calculation of the homology of small chain complexes. To calculate $H_*(A)$, we try to find bases for all the groups A_i such that the differential sends every basis element to another basis element or to zero. We can then take the basis for A and discard all pairs of basis elements (a, a') with $da = a'$. The remaining elements will then give a basis for $H_*(A)$ over $\mathbb{Z}$, so in particular all the groups $H_i(A)$ are free abelian groups.

There are some chain complexes A for which the groups $H_i(A)$ are not free abelian groups, and in those cases we will not be able to find a basis with properties as above. There is a more complicated algorithm that will still work in those cases, but we will not explain it here.

Video (Definition 13.9 to Remark 13.13)

Definition 13.9. Let A and A' be chain complexes. A *chain map* from A to A' is a sequence of group homomorphisms $f_n\colon A_n \to A'_n$ such that for all $n \geq 0$ we have $f_n \circ d^A_{n+1} = d^{A'}_{n+1} \circ f_{n+1}$. In other words, the following diagram must commute:

$$\begin{tikzcd} A_{n+1} \arrow[r, "f_{n+1}"] \arrow[d, "d^A_{n+1}"'] & A'_{n+1} \arrow[d, "d^{A'}_{n+1}"] \\ A_n \arrow[r, "f_n"'] & A'_n. \end{tikzcd}$$

We will typically suppress all subscripts and superscripts and just write $fd = df$ rather than $f_n \circ d^A_{n+1} = d^{A'}_{n+1} \circ f_{n+1}$.

Definition 13.10. We write id_{A_*} for the sequence of identity maps $\mathrm{id}_{A_n}\colon A_n \to A_n$. This is clearly a chain map from A to itself. Now suppose we have chain maps $A \xrightarrow{f} A' \xrightarrow{f'} A''$. We define a chain map $f' \circ f\colon A \to A''$ by the obvious rule $(f' \circ f)_n = f'_n \circ f_n$; this satisfies

$$d \circ (f' \circ f) = d \circ f' \circ f = f' \circ d \circ f = (f' \circ f) \circ d$$

as required. It is easy to see that this kind of composition is associative and that $\mathrm{id} \circ f = f = f \circ \mathrm{id}$, so we have a category Chain of chain complexes and chain maps.

Proposition 13.11. *Let $f\colon A \to A'$ be a chain map. Then*

(a) *For all n we have $f_n(Z_n(A)) \leq Z_n(A')$*
(b) *For all n we have $f_n(B_n(A)) \leq B_n(A')$*
(c) *There is a well-defined map $f_*\colon H_n(A) \to H_n(A')$ given by $f_*[z] = [f(z)]$.*

Moreover, these constructions give functors $Z_n, B_n, H_n\colon$ Chain $\to$ Ab.

Proof. (a) Suppose that $z \in Z_n(A)$, so $d(z) = 0$. We then have $f(z) \in A'_n$ with $d(f(z)) = f(d(z)) = f(0) = 0$, so $f(z) \in Z_n(A')$ as required. (As mentioned in Remark 10.22, the notation $[z']$ is only meaningful if $d(z') = 0$. Now we have checked that $d(f(z)) = 0$, we see that the notation $[f(z)]$ used in (c) is valid.)

(b) Suppose that $b \in B_n(A)$, so $b = d(x)$ for some $x \in A_{n+1}$. We then have $f(b) = fd(x) = df(x) \in d(A'_{n+1}) = B_n(A')$ as required.

(c) From (a) and (b) we see that if $z+B_n(A) = z'+B_n(A)$ then $z-z' \in B_n(A)$ so $f(z)-f(z') = f(z-z') \in B_n(A')$ so $f(z) + B_n(A') = f(z') + B_n(A')$. It follows that there is an induced homomorphism $f_*\colon H_n(A) \to H_n(A')$ given by

$$f_*(z + B_n(A)) = f(z) + B_n(A'),$$

or in other words $f_*[z] = [f(z)]$.

We can make Z_n into a functor by defining

$$Z_n f = f_n|_{Z_n(A)}\colon Z_n(A) \to Z_n(A').$$

As this is just a restriction of f_n, it is clearly compatible with composition and identity morphisms, as required. We can make B_n into a functor in the same way. Finally, suppose we have chain maps

$$A \xrightarrow{f} A' \xrightarrow{f'} A''.$$

For a homology class $[z] \in H_n(A)$ we have

$$(f' \circ f)_*([z]) = [(f' \circ f)(z)] = [f'(f(z))] = f'_*[f(z)] = f'_* f_*[z].$$

From this it is clear that $H_n\colon$ Chain $\to$ Ab is also a functor. □

Construction 13.12. We now want to make homology into a functor from topological spaces to abelian groups. Let $f\colon X \to Y$ be a continuous map. Let u be an element of the set $S_k(X)$, or equivalently, a continuous map $u\colon \Delta_k \to X$. We define $f_\#(u) \in S_k(Y)$ to be the composite function $f \circ u\colon \Delta_k \to Y$. Next, given an element $u = \sum_{i=1}^r n_i u_i \in C_k(X)$ we define $f_\#(u) = \sum_{i=1}^r n_i f_\#(u_i) \in C_k(Y)$. This extends the map $f_\#\colon S_k(X) \to S_k(Y)$ linearly to give a homomorphism $f_\#\colon C_k(X) \to C_k(Y)$. We claim that this is a chain map, or in other words that $f_\#(\partial(u)) = \partial(f_\#(u))$ whenever $u \in C_k(X)$. As everything is extended linearly, it will be enough to prove this when $u \in S_k(X)$, or equivalently $u\colon \Delta_k \to X$. We then have

$$\begin{aligned}
f_\#(u) &= f \circ u \\
\partial(u) &= \sum_{i=0}^k (-1)^i (u \circ \delta_i) \\
f_\#(\partial(u)) &= \sum_{i=0}^k (-1)^i (f \circ u \circ \delta_i) = \partial(f_\#(u)),
\end{aligned}$$

as required. It is clear that $\mathrm{id}_\# = \mathrm{id}$ and $(g \circ f)_\# = g_\# \circ f_\#$, so we have defined a functor $C_*\colon$ Top $\to$ Chain. We can compose this with the functor $H_n\colon$ Chain $\to$ Ab to get a functor Top $\to$ Ab which is traditionally also denoted by H_n.

Remark 13.13. Now that we know that H_n is a functor, we can use Propositions 6.17 and 6.25. These tell us that:

(a) If X is homeomorphic to Y, then $H_n(X)$ is isomorphic to $H_n(Y)$.
(b) If X is a retract of Y, then $H_n(X)$ is a retract of $H_n(Y)$.

Remark 13.14. Suppose that $f\colon X \to Y$ is constant, say $f(x) = b$ for all $x \in X$. We then claim that $f_* = 0\colon H_n(X) \to H_n(Y)$ for all $n > 0$. Indeed, we can define $X \xrightarrow{p} \{0\} \xrightarrow{q} Y$ by $p(x) = 0$ and $q(0) = b$, so $f = q \circ p$. Thus, the map $f_*\colon H_n(X) \to H_n(Y)$ is the composite of the maps $p_*\colon H_n(X) \to H_n(\{0\})$ and $q_*\colon H_n(\{0\}) \to H_n(Y)$. However, Proposition 10.23 tells us that $H_n(\{0\}) = 0$, and the claim is clear from this.

14. Chain homotopy

We next need to build a connection between the concept of homotopy in topology, and the behaviour of chain complexes and homology in algebra.

Video (Definition 14.1 to Proposition 14.7)

Definition 14.1. Let A_* and A'_* be chain complexes, and let $f, g\colon A_* \to A'_*$ be chain maps (so $df = fd$ and $dg = gd$). A *chain homotopy* between f and g is a system of maps $s_r\colon A_r \to A'_{r+1}$ such that $g_r - f_r = d^{A'}_{r+1} s_r + s_{r-1} d^A_r$ for all r (or more briefly, $g - f = ds + sd$). In the case $r = 0$, we should interpret s_{-1} as 0, so the condition is $g_0 - f_0 = d_1 s_0$. We say that f and g are *chain homotopic* if there is a chain homotopy between them. If so, we write $f \cong g$.

Remark 14.2. In Section 2 we introduced a crude intuitive version of homology involving chains as subsets of a space X. For such a chain $u \subseteq X$ we can define $\sigma(u) = [0,1] \times u$, which is a chain in $[0,1] \times X$; we call this the *thickening* of u. If u is a filled triangle in X, then $\sigma(u)$ is a triangular prism. Now $\partial\sigma(u)$ is the boundary of this triangular prism, which consists of the top, the bottom and the sides. On the other hand, $\sigma\partial(u)$ is what we get by thickening the boundary of u, which is just the sides of the prism. After adjusting the $\pm$-signs to account for orientations, we end up with the relation

$$\partial\sigma(u) + \sigma\partial(u) = \text{top} - \text{bottom}.$$

This is a relation between chains in $[0,1] \times X$, but if we have two maps $f, g\colon X \to Y$ and a homotopy $h\colon [0,1] \times X \to Y$ between them, then we can apply h_* to get a relation between chains in Y. The main point of this section is to provide a rigorous and general version of this picture.

Example 14.3. In Example 13.7 we introduced a chain complex A with

$$A_k = \begin{cases} \mathbb{Z}\{e_i \mid i \in \mathbb{Z}/n\} & \text{if } k = 1 \\ \mathbb{Z}\{v_i \mid i \in \mathbb{Z}/n\} & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

The differential is given by $d(e_i) = v_{i+1} - v_i$ and $d(v_i) = 0$. Define $f\colon A_* \to A_*$ by $f(e_i) = e_{i+1}$ and $f(v_i) = v_{i+1}$. This is a chain map because $d(f(e_i)) = v_{i+2} - v_{i+1} = f(d(e_i))$. We claim that f is chain homotopic to the identity. Indeed, we can define $s_0\colon A_0 \to A_1$ by $s_0(v_i) = e_i$, and we define $s_i\colon A_i \to A_{i+1}$ to be zero for all $i \neq 0$. We then find that

$$\begin{aligned} (ds + sd)(e_i) &= d(0) + s(v_{i+1} - v_i) = e_{i+1} - e_i = f(e_i) - \mathrm{id}(e_i) \\ (ds + sd)(v_i) &= d(e_i) + s(0) = v_{i+1} - v_i = f(v_i) - \mathrm{id}(v_i) \end{aligned}$$

Proposition 14.4. *The relation of being chain homotopic is an equivalence relation.*

Proof. The zero map is a chain homotopy from f to itself. If s is a chain homotopy from f to g, then $-s$ is a chain homotopy from g to f. If t is also a chain homotopy from g to h, then $s + t$ is a chain homotopy from f to h. ∎

Proposition 14.5. *Suppose we have chain maps*

$$A \underset{f_1}{\overset{f_0}{\rightrightarrows}} B \underset{g_1}{\overset{g_0}{\rightrightarrows}} C$$

Suppose that f_0 is chain homotopic to f_1, and that g_0 is homotopic to g_1. Then $g_1 \circ f_1$ is chain homotopic to $g_0 \circ f_0$.

Proof. We are assuming that f_0 is chain homotopic to f_1, which means that there is a chain homotopy s with $f_1 - f_0 = ds + sd$. Similarly, there is a chain homotopy t with $g_1 - g_0 = dt + td$. Put $u = g_1 s + t f_0$. Using $dg_1 = g_1 d$ we get $du = g_1 ds + dt f_0$. Using $f_0 d = df_0$ we get $ud = g_1 sd + tdf_0$. By adding these, we get

$$du + ud = g_1(ds + sd) + (td + dt)f_0 = g_1(f_1 - f_0) + (g_1 - g_0)f_0 = g_1 f_1 - g_1 f_0 + g_1 f_0 - g_0 f_0 = g_1 f_1 - g_0 f_0,$$

as required. □

Definition 14.6. We write $\mathrm{hChain}(A_*, A'_*)$ for the set of chain homotopy classes of chain maps from A_* to A'_*, or in other words equivalence classes under the equivalence relation defined above. Using Proposition 14.5, we see that these are the morphism sets of a well-defined category hChain, whose objects are chain complexes. This is analogous to the category hTop introduced in Definition 9.9.

Proposition 14.7. *Let A_* and A'_* be chain complexes, and let $f, g\colon A_* \to A'_*$ be chain maps that are chain homotopic. Then the induced maps $f_*, g_*\colon H_*(A) \to H_*(A')$ are the same.*

Proof. Let s be a chain homotopy from f to g, so $g - f = ds + sd$. Consider an element $[z] \in H_r(A)$, so $z \in A_r$ with $dz = 0$. Recall from Proposition 13.11 that $f(z), g(z) \in Z_r(A')$ so that the expressions $[f(z)]$ and $[g(z)]$ are meaningful and refer to elements of $H_r(A')$. By definition we have $f_*[z] = [f(z)]$ and $g_*[z] = [g(z)]$, so we need to check that these are the same. We have $g(z) - f(z) = d(s(z)) + s(d(z))$ but $d(z) = 0$ so $g(z) = f(z) + d(s(z)) \in f(z) + B_r(A')$. It follows that $[g(z)] = [f(z)]$ in $H_r(A')$, or in other words $f_*[z] = g_*[z]$. This means that $f_* = g_*$ as claimed. □

Proposition 14.8. *Let X and Y be topological spaces, and let $f, g\colon X \to Y$ be continuous maps that are homotopic to each other. Then the chain maps $f_\#, g_\#\colon C_*(X) \to C_*(Y)$ are chain homotopic to each other, so the induced maps $f_*, g_*\colon H_*(X) \to H_*(Y)$ are the same.*

For the proof, we will first choose a homotopy h from f to g, so h is a continuous map $[0,1] \times X \to Y$ with $h(0,x) = f(x)$ and $h(1,x) = g(x)$ for all $x \in X$. We need to use this to construct a chain homotopy between $f_\#$ and $g_\#$, or equivalently a system of maps $\sigma_k\colon C_k(X) \to C_{k+1}(Y)$ with $\partial\sigma + \sigma\partial = g_\# - f_\#$. Before giving the general proof, we will discuss the cases $k = 0$ and $k = 1$.

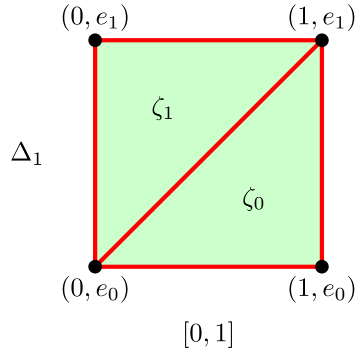
Consider a point $a \in S_0(X) = X$. We can define a continuous map $v\colon \Delta_1 \to Y$ by $v(1-t, t) = h(t, a)$. This can be regarded as an element of $S_1(Y) \subset C_1(Y)$, and it satisfies

$$\partial(v) = v(e_1) - v(e_0) = h(1,a) - h(0,a) = g(a) - f(a) = g_\#(a) - f_\#(a).$$

We define $\sigma_0(a) = v$ and extend linearly to get a homomorphism $\sigma_0\colon C_0(X) \to C_1(Y)$ with $\partial(\sigma_0(u)) = g_\#(u) - f_\#(u)$ for all $u \in C_0(X)$.

Now consider instead an element $u \in S_1(X)$, or in other words, a continuous map $u\colon \Delta_1 \to X$. We want to define $\sigma_1(u) \in C_2(Y)$, so $\sigma_1(u)$ should be a $\mathbb{Z}$-linear combination of continuous maps from the triangle Δ_2 to Y. We have a map $m\colon [0,1] \times \Delta_1 \to Y$ given by $m(t,s) = h(t, u(s))$. Here Δ_1 is homeomorphic to $[0,1]$ so $[0,1] \times \Delta_1$ is a square, with corners $(0,e_0)$, $(0,e_1)$, $(1,e_0)$ and $(1,e_1)$. We can divide this square into two triangles and restrict k to these triangles, giving two different maps $\Delta_2 \to Y$, which will be the terms in $\sigma_1(u)$. In detail, we define maps $\zeta_0, \zeta_1\colon \Delta_2 \to [0,1] \times \Delta_1$ by

$$\zeta_0(t_0,t_1,t_2) = t_0(0,e_0) + t_1(1,e_0) + t_2(1,e_1)$$
$$\zeta_1(t_0,t_1,t_2) = t_0(0,e_0) + t_1(0,e_1) + t_2(1,e_1).$$



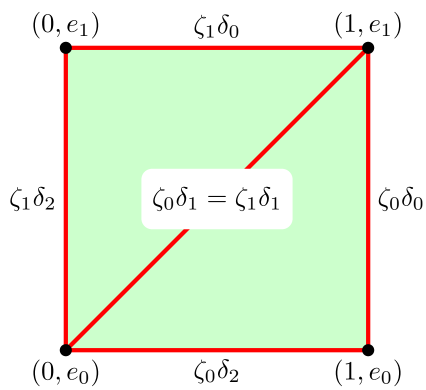
The composites $(\Delta_2 \xrightarrow{\zeta_i} [0,1] \times \Delta_1 \xrightarrow{m} Y)$ (for $i = 0, 1$) can be regarded as elements of $S_2(Y) \subset C_2(Y)$. We define $\sigma_1(u) = m\zeta_0 - m\zeta_1 \in C_2(Y)$. This can be extended linearly to give a homomorphism $\sigma_1 \colon C_1(X) \to C_2(Y)$. We claim that $\partial\sigma_1(u) + \sigma_0\partial(u) = g_\#(u) - f_\#(u)$ for all $u \in C_1(X)$. Indeed, it will be enough to prove this when $u \in S_1(X)$. We then have

$$\partial\sigma_1(u) = \partial(m\zeta_0) - \partial(m\zeta_1) = m\zeta_0\delta_0 - m\zeta_0\delta_1 + m\zeta_0\delta_2 - m\zeta_1\delta_0 + m\zeta_1\delta_1 - m\zeta_1\delta_2$$

(where $m(t, s) = h(t, u(s))$ as before). Here, for $t = (t_0, t_1) = (1 - t_1, t_1) \in \Delta_1$ we have

$$\begin{aligned}
(\zeta_0\delta_0)(t) &= \zeta_0(0, t_0, t_1) = t_0(1, e_0) + t_1(1, e_1) = (1, t) \\
(\zeta_0\delta_1)(t) &= \zeta_0(t_0, 0, t_1) = t_0(0, e_0) + t_1(1, e_1) = (t_1, t) \\
(\zeta_0\delta_2)(t) &= \zeta_0(t_0, t_1, 0) = t_0(0, e_0) + t_1(1, e_0) = (t_1, e_0) \\
(\zeta_1\delta_0)(t) &= \zeta_1(0, t_0, t_1) = t_0(0, e_1) + t_1(1, e_1) = (t_1, e_1) \\
(\zeta_1\delta_1)(t) &= \zeta_1(t_0, 0, t_1) = t_0(0, e_0) + t_1(1, e_1) = (t_1, t) \\
(\zeta_1\delta_2)(t) &= \zeta_1(t_0, t_1, 0) = t_0(0, e_0) + t_1(0, e_1) = (0, t).
\end{aligned}$$

This can be displayed as follows:



It follows that

$$\begin{aligned}
m\zeta_0\delta_0(t) &= h(1, u(t)) = g(u(t)) \\
m\zeta_0\delta_1(t) &= h(t_1, u(t)) = m\zeta_1\delta_1(t_0, t_1) \\
m\zeta_0\delta_2(t) &= h(t_1, u(e_0)) = \sigma_0(u(e_0))(t) \\
m\zeta_1\delta_0(t) &= h(t_1, u(e_1)) = \sigma_0(u(e_1))(t) \\
m\zeta_1\delta_2(t) &= h(0, u(t)) = f(u(t)).
\end{aligned}$$

Thus, in our formula for $\partial(\sigma_1(u))$ we see that the first and last terms give $g_\#(u) - f_\#(u)$, the second and fifth terms cancel out, and the third and fourth terms give $\sigma_0(u(e_0) - u(e_1)) = -\sigma_0(\partial(u))$. Putting this together, we get $\partial\sigma_1(u) + \sigma_0\partial(u) = g_\#(u) - f_\#(u)$ as required.

We now extend the above discussion to cover $k > 1$.

Video (Definition 14.9 to Lemma 14.12)

Definition 14.9. For $0 \leq i \leq k$ we define $\zeta_i \colon \Delta_{k+1} \to [0,1] \times \Delta_k$ by

$$\begin{aligned}\zeta_i(t_0, \dots, t_{k+1}) &= \sum_{j=0}^{i} t_j.(0, e_j) + \sum_{j=i+1}^{k+1} t_j.(1, e_{j-1}) \\ &= t_0(0, e_0) + \cdots + t_i(0, e_i) + t_{i+1}(1, e_i) + \cdots + t_{k+1}(1, e_k) \\ &= (t_{i+1} + \cdots + t_{k+1}, (t_0, \dots, t_{i-1}, t_i + t_{i+1}, t_{i+2}, \dots, t_{k+1}))\end{aligned}$$

If it is necessary to specify k, we will write $\zeta_{k,i}$ instead of ζ_i.

Example 14.10. When $k = 2$ the maps $\zeta_i \colon \Delta_3 \to [0,1] \times \Delta_2$ are given by

$$\begin{aligned}\zeta_0(x_0, x_1, x_2, x_3) &= (x_1 + x_2 + x_3, (x_0 + x_1, x_2, x_3)) \\ \zeta_1(x_0, x_1, x_2, x_3) &= (x_2 + x_3 + x_4, (x_0, x_1 + x_2, x_3)) \\ \zeta_2(x_0, x_1, x_2, x_3) &= (x_3 + x_4, (x_0, x_1, x_2 + x_3)).\end{aligned}$$

We saw above that $[0,1] \times \Delta_1$ is the union of the triangles $\zeta_0(\Delta_2)$ and $\zeta_1(\Delta_2)$, which fit together nicely along one edge. In the same way, it can be shown that $[0,1] \times \Delta_2$ is the union of the images of the maps $\zeta_i \colon \Delta_3 \to [0,1] \times \Delta_2$, and the intersection of any two of these images is another simplex of lower dimension. However, we will not need this so we omit the proof.

Interactive demo

When $k = 3$ we have

$$\begin{aligned}\zeta_0(x_0, x_1, x_2, x_3, x_4) &= (x_1 + x_2 + x_3 + x_4, (x_0 + x_1, x_2, x_3, x_4)) \\ \zeta_1(x_0, x_1, x_2, x_3, x_4) &= (x_2 + x_3 + x_4, (x_0, x_1 + x_2, x_3, x_4)) \\ \zeta_2(x_0, x_1, x_2, x_3, x_4) &= (x_3 + x_4, (x_0, x_1, x_2 + x_3, x_4)) \\ \zeta_3(x_0, x_1, x_2, x_3, x_4) &= (x_4, (x_0, x_1, x_2, x_3 + x_4)).\end{aligned}$$

Definition 14.11. Given $h \colon [0,1] \times X \to Y$ and $u \colon \Delta_k \to X$ as before, we put

$$\sigma_k(u) = \sum_{i=0}^{k} (-1)^i (h \circ (\mathrm{id} \times u) \circ \zeta_i) \in C_{k+1}(Y).$$

We extend this linearly to define $\sigma_k \colon C_k(X) \to C_{k+1}(Y)$.

Lemma 14.12.

(a) *Suppose that $0 \leq i \leq k$ and $0 \leq j \leq k+1$, so we can form the composite*

$$\Delta_k \xrightarrow{\delta_j} \Delta_{k+1} \xrightarrow{\zeta_i} [0,1] \times \Delta_k.$$

If $j < i$ then $\zeta_i \delta_j$ is the same as the composite

$$\Delta_k \xrightarrow{\zeta_{i-1}} [0,1] \times \Delta_{k-1} \xrightarrow{\mathrm{id} \times \delta_j} [0,1] \times \Delta_k,$$

or in other words $\zeta_i \delta_j = (\mathrm{id} \times \delta_j) \zeta_{i-1}$.

(b) *On the other hand, if $j \geq i + 2$ then $\zeta_i \delta_j = (\mathrm{id} \times \delta_{j-1}) \zeta_i$.*

(c) *For $1 \leq i \leq k$ we also have $\zeta_i \delta_i = \zeta_{i-1} \delta_i$.*

(d) *Finally, we have $\zeta_0 \delta_0(x) = (1, x)$ and $\zeta_k \delta_{k+1}(x) = (0, x)$.*

Proof. All the maps under discussion are affine maps from Δ_k to $[0,1] \times \Delta_k$. To check that two such maps agree, it suffices to check that they have the same effect on the vertices of Δ_k. The map δ_j sends e_p to e_p (if $p < j$) or e_{p+1} (if $p \geq j$). The map ζ_i sends e_p to $(0, e_p)$ (if $p \leq i$) or $(1, e_{p-1})$ (if $p > i$).

(a) When $j < i$ we find that both $\zeta_i\delta_j$ and $(\mathrm{id}\times\delta_j)\zeta_{i-1}$ have the following effect:

$$e_p \mapsto \begin{cases} (0, e_p) & \text{if } p < j \\ (0, e_{p+1}) & \text{if } j \leq p < i \\ (1, e_p) & \text{if } i \leq p. \end{cases}$$

(b) When $j \geq i+2$ we find that both $\zeta_i\delta_j$ and $(\mathrm{id}\times\delta_{j-1})\zeta_i$ have the following effect:

$$e_p \mapsto \begin{cases} (0, e_p) & \text{if } p \leq i \\ (1, e_{p-1}) & \text{if } i < p < j \\ (1, e_p) & \text{if } j \leq p. \end{cases}$$

(c) When $1 \geq i \leq k$ we find that both $\zeta_i\delta_i$ and $\zeta_{i-1}\delta_i$ have the following effect:

$$e_p \mapsto \begin{cases} (0, e_p) & \text{if } p < i \\ (1, e_p) & \text{if } i \leq p. \end{cases}$$

(d) We also have $\zeta_0\delta_0(e_p) = \zeta_0(e_{p+1}) = (1, e_p)$ and $\zeta_k\delta_{k+1}(e_p) = \zeta_k(e_p) = (0, e_p)$

□

Video (Proposition 14.13 to Proposition 14.17)

Proposition 14.13. *For all $u \in C_k(X)$ we have*

$$\partial(\sigma_k(u)) + \sigma_{k-1}(\partial(u)) = g_\#(u) - f_\#(u).$$

Proof. It will be enough to prove this when $u \in S_k(X) \subset C_k(X)$, so $u\colon \Delta_k \to X$. We can then define $m = h \circ (\mathrm{id}\times u)\colon [0,1] \times \Delta_k \to X$, so $\sigma_k(u) = \sum_{i=0}^k (-1)^i m\zeta_i$. This gives

$$\partial(\sigma_k(u)) = \sum_{i=0}^k \sum_{j=0}^{k+1} (-1)^{i+j} m\zeta_i\delta_j.$$

We can divide this sum into four parts:

- A is the sum of the terms where $j < i$
- B is the sum of the terms where $i+2 \leq j$
- C is the sum of the terms where $j = i$ with $1 \leq i \leq k$
- D is the sum of the terms where $j = i+1$ with $0 \leq i < k$
- E consists of the terms with $(i,j) = (0,0)$ or $(i,j) = (k, k+1)$.

We thus have $\partial(\sigma_k(u)) = A + B + C + D + E$.

Now note that $\partial(u) = \sum_{q=0}^k (-1)^q u\delta_q$, and

$$\sigma_{k-1}(u\delta_q) = \sum_{p=0}^k (-1)^p (h \circ (\mathrm{id}\times u\delta_q) \circ \zeta_p) = \sum_{p=0}^k (-1)^p m(\mathrm{id}\times\delta_q)\zeta_p.$$

This gives

$$\sigma_{k-1}(\partial(u)) = \sum_{p=0}^k \sum_{q=0}^k (-1)^{p+q} m(\mathrm{id}\times\delta_q)\zeta_p.$$

We let A' be the sum of the terms where $q \leq p$, and we let B' be the sum of the terms where $p < q$.

Each term $(-1)^{i+j}m\zeta_i\delta_j$ in A can be rewritten (using part (a) of the lemma) as $(-1)^{i+j}m(\mathrm{id}\times\delta_j)\zeta_{i-1}$. This can then be written as $-(-1)^{p+q}m(\mathrm{id}\times\delta_q)\delta_p$, where $p = i-1$ and $q = j$. Because $j < i$ for terms in A, we see that $q \leq p$, so the rewritten term is the negative of a term in A'. Similarly, each term $(-1)^{i+j}m\zeta_i\delta_j$ in B can be rewritten (using part (b) of the lemma) as $(-1)^{i+j}m(\mathrm{id}\times\delta_{j-1})\zeta_i$. This can then be written as $-(-1)^{p+q}m(\mathrm{id}\times\delta_q)\delta_p$, where $p = i$ and $q = j-1$. Because $j \geq i+2$ for terms in B, we see that $q > p$, so the rewritten term is the negative of a term in B'. Using this we see that $A' = -A$ and $B' = -B$. A similar argument with part (c) of the lemma shows that $D = -C$. We now have

$$\begin{aligned}\partial(\sigma_k(u)) + \sigma_{k-1}(\partial(u)) &= (A + B + C + D + E) + (A' + B') \\ &= (A + B + C - C + E) + (-A - B) = E.\end{aligned}$$

Also, part (d) of the lemma gives

$$\begin{aligned}m\zeta_0\delta_0(x) &= m(1, x) = h(1, u(x)) = g(u(x)) \\ m\zeta_k\delta_{k+1}(x) &= m(0, x) = h(0, u(x)) = f(u(x)).\end{aligned}$$

The first of these terms has sign $(-1)^{0+0} = +1$, and the second has sign $(-1)^{k+(k+1)} = -1$. We therefore have $E = g_\#(u) - f_\#(u)$ as required. $\square$

We can gain some insight into the above proof by considering a simple special case. Suppose that $X = \mathbb{R}^N$ and $Y = \mathbb{R}^M$. Suppose that $f\colon X \to Y$ is affine, i.e. $f(x) = Ax + b$ for some matrix A and vector b. Suppose that $g\colon X \to Y$ is also affine, and that h is just the linear homotopy $h(t, x) = (1 - t)f(x) + tg(x)$. Consider a linear 3-simplex $u = \langle a_0, a_1, a_2, a_3\rangle \in C_3(X)$. As all the maps involved are affine, we see that

$$\begin{aligned}f_*(u) &= \langle f(a_0), f(a_1), f(a_2), f(a_3)\rangle \\ g_*(u) &= \langle g(a_0), g(a_1), g(a_2), g(a_3)\rangle.\end{aligned}$$

As in Example 10.17, we will use abbreviated notation, writing i for a_i or $f(a_i)$, and $\bar{i}$ for $g(a_i)$, so the above equations become $f_*(0123) = 0123$ and $g_*(0123) = \overline{0123}$. It is then not hard to check that

$$h \circ (\mathrm{id} \times u) \circ \zeta_2 = \langle f(a_0), f(a_1), f(a_2), g(a_2), g(a_3)\rangle = 012\overline{23},$$

and similarly for the other terms in $\sigma(u)$. The terms in $\partial(u)$, $\sigma(u)$, $\partial\sigma(u)$ and $\sigma\partial(u)$ can now be laid out as follows:

$$\sigma(x) = \quad +0\overline{0123} \quad -01\overline{123} \quad +012\overline{23} \quad -0123\overline{3} \qquad\qquad \partial(x) = \quad +123 \quad -023 \quad +013 \quad -012$$

$$\partial\sigma(x) = \begin{array}{cccc} +\overline{0123} & -1\overline{123} & +12\overline{23} & -123\overline{3} \\ -0\overline{123} & +0\overline{123} & -02\overline{23} & +023\overline{3} \\ +0\overline{023} & -01\overline{23} & +01\overline{23} & -013\overline{3} \\ -0\overline{013} & +01\overline{13} & -012\overline{3} & +012\overline{3} \\ +0\overline{012} & -01\overline{12} & +012\overline{2} & -0123 \end{array} \qquad \sigma\partial(x) = \begin{array}{ccc} +1\overline{123} & -12\overline{23} & +123\overline{3} \\ -0\overline{023} & +02\overline{23} & -023\overline{3} \\ +0\overline{013} & -01\overline{13} & +013\overline{3} \\ -0\overline{012} & +01\overline{12} & -012\overline{2} \end{array}$$

Most terms cancel in the indicated groups, which correspond to the expressions $A, \dots, E$ in the proof of Proposition 14.13, leaving $\partial\sigma(u) + \sigma\partial(u) = \overline{0123} - 0123 = g_*(u) - f_*(u)$ as expected. Here we have just displayed the case $k = 3$, but the pattern generalises in an obvious way to other values of k. This presentation is only directly relevant for linear simplices and affine maps. However, in the general case, most of the work involves linear simplices in the space $[0, 1] \times \Delta_k \subseteq \mathbb{R}^{k+2}$, and then we finish up by applying the map $h \circ (\mathrm{id} \times u)\colon [0, 1] \times \Delta_k \to Y$. Because of this, it is possible to deduce the general case from the linear case, although we will not spell out the details here.

Corollary 14.14. *If $f\colon X \to Y$ is a homotopy equivalence, then $f_*\colon H_*(X) \to H_*(Y)$ is an isomorphism.*

Proof. Choose a map $g\colon Y \to X$ which is homotopy inverse to f, so $g \circ f$ is homotopic to id_X and $f \circ g$ is homotopic to id_Y. As homology is a functor, the composite $g_* \circ f_*\colon H_*(X) \to H_*(X)$ is the same as $(g \circ f)_*$. As $g \circ f$ is homotopic to id_X, Proposition 14.8 tells us that $(g \circ f)_* = (\mathrm{id}_X)_*$. Using functoriality again, we have $(\mathrm{id}_X)_* = \mathrm{id}_{H_*(X)}$. Putting this together, we see that $g_* \circ f_* = \mathrm{id}\colon H_*(X) \to H_*(X)$, and essentially the same argument shows that $f_* \circ g_* = \mathrm{id}\colon H_*(Y) \to H_*(Y)$. Thus, f_* and g_* are mutually inverse isomorphisms. $\square$

Remark 14.15. Another way to organise the above argument is as follows. Propositions 14.7 and 14.8 tell us that H_n can be regarded as a functor hTop $\to$ Ab. Any homotopy equivalence $f\colon X \to Y$ becomes an isomorphism in hTop, and Corollary 6.18 tells us that functors send isomorphisms to isomorphisms, so $H_n(f)$ must be an isomorphism.

Proposition 14.16. *If $f\colon X \to Y$ is homotopic to a constant map, then the map $f_*\colon H_n(X) \to H_n(Y)$ is zero for all $n > 0$.*

Proof. Let $g\colon X \to Y$ be a constant map that is homotopic to f. Then $f_* = g_*$ by Proposition 14.8, but $g_* = 0$ by Remark 13.14. □

Proposition 14.17. *Suppose that X is contractible. Then $H_0(X) = \mathbb{Z}$ but $H_n(X) = 0$ for all $n \neq 0$. In particular, this applies if X is a convex subset of $\mathbb{R}^N$ for some n.*

Proof. Proposition 9.17 tells us that X is homotopy equivalent to a point, so it has the same homology as a point, which is given by Proposition 10.23. □

15. Homology of spheres

Consider a topological space X with open subsets U and V such that $X = U \cup V$. There are many cases like this where we already understand three of the groups $H_*(U)$, $H_*(V)$, $H_*(U\cap V)$ and $H_*(U\cup V) = H_*(X)$, and we want to determine the fourth one.

Video (Theorem 15.1 and Proposition 15.2)

We name the corresponding inclusion maps as shown on the left below; they induce homomorphisms of homology groups as shown on the right.

$$\begin{tikzcd} U\cap V \arrow[r, "i"] \arrow[d, "j"'] & U \arrow[d, "k"] \\ V \arrow[r, "l"'] & U\cup V = X \end{tikzcd} \qquad \begin{tikzcd} H_*(U\cap V) \arrow[r, "i_*"] \arrow[d, "j_*"'] & H_*(U) \arrow[d, "k_*"] \\ H_*(V) \arrow[r, "l_*"'] & H_*(U\cup V) = H_*(X) \end{tikzcd}$$

As the left hand diagram commutes and homology is functorial, we have $k_*i_* = l_*j_*\colon H_*(U\cap V) \to H_*(U\cup V)$.

We can now combine the above maps to get maps as follows:

$$H_n(U\cap V) \xrightarrow{\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right]} H_n(U)\oplus H_n(V) \xrightarrow{[\,k_*\ l_*\,]} H_n(U\cup V)$$

Here the notation $\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right]$ refers to the map sending a to $(i_*(a), -j_*(a))$, and the notation $[\,k_*\ l_*\,]$ refers to the map sending (b,c) to $k_*(b)+l_*(c)$. The composite of these two maps therefore sends a to $k_*i_*(a) - l_*j_*(a)$, but this is zero because $k_*i_* = l_*j_*$. Thus, the above sequence has a chance to be exact. In fact, we have the following result:

Theorem 15.1 (Mayer-Vietoris)**.** *In the above context, there are natural maps $\delta\colon H_n(U\cup V) \to H_{n-1}(U\cap V)$ such that the resulting sequence*

$$H_{n+1}(U\cup V) \xrightarrow{\delta} H_n(U\cap V) \xrightarrow{\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right]} H_n(U)\oplus H_n(V) \xrightarrow{[\,k_*\ l_*\,]} H_n(U\cup V) \xrightarrow{\delta} H_{n-1}(U\cap V)$$

is exact for all n.

In this section, we will explain some homology calculations based on the above theorem. The proof will be given in Section 19, using preliminary results from Sections 17 and 18.

We first note that the exact sequences in the theorem can be chained together to make an infinite sequence as follows:

$$\begin{tikzcd}
\cdots \arrow[r] & H_3(U\cap V) \arrow[r] & H_3(U)\oplus H_3(V) \arrow[r] & H_3(X) \arrow[dll, rounded corners, "\delta"] \\
 & H_2(U\cap V) \arrow[r] & H_2(U)\oplus H_2(V) \arrow[r] & H_2(X) \arrow[dll, rounded corners, "\delta"] \\
 & H_1(U\cap V) \arrow[r] & H_1(U)\oplus H_1(V) \arrow[r] & H_1(X) \arrow[dll, rounded corners, "\delta"] \\
 & H_0(U\cap V) \arrow[r] & H_0(U)\oplus H_1(V) \arrow[r] & H_0(X) \arrow[r] & 0
\end{tikzcd}$$

The following slightly modified form is often convenient.

Proposition 15.2. *Suppose that $U \cap V$ is path connected (and therefore nonempty). Then we can modify the Mayer-Vietoris sequence by replacing all the H_0 terms by zero, and the resulting sequence is still exact. Moreover, if U and V are also path connected then the same is true of $U \cup V$, so all the H_0 groups are isomorphic to $\mathbb{Z}$.*

Proof. Pick a point $a \in U \cap V$, so $H_0(U \cap V) = \mathbb{Z}.[a]$. Now $H_0(U)$ has a basis corresponding to the elements of $\pi_0(U)$, and $i_*[a]$ is one of these basis elements, so the map $i_* \colon H_0(U \cap V) \to H_0(U)$ is injective. It follows that the map $\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right] : H_0(U \cap V) \to H_0(U) \oplus H_0(V)$ is also injective. Now consider the tail end of the Mayer-Vietoris sequence:

$$H_1(U) \oplus H_1(V) \xrightarrow{\left[\begin{smallmatrix} k_* & l_* \end{smallmatrix}\right]} H_1(U \cup V) \xrightarrow{\delta} H_0(U \cap V) \xrightarrow{\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right]} H_0(U) \oplus H_0(V).$$

The map $\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right]$ is injective, and so the kernel is zero. As the sequence is exact, we see that the image of δ is also zero, so δ is the zero homomorphism. Using exactness again, we deduce that the map $\left[\begin{smallmatrix} k_* & l_* \end{smallmatrix}\right]$ must be surjective. Using this together with the exactness of the original MV sequence, we see that the sequence

$$H_1(U \cap V) \xrightarrow{\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right]} H_1(U) \oplus H_1(V) \xrightarrow{\left[\begin{smallmatrix} k_* & l_* \end{smallmatrix}\right]} H_1(U \cup V) \to 0 \to 0 \to 0$$

is exact. The rest of the modified MV sequences is the same as the original MV sequence, and so is also exact, as claimed.

Now suppose that U and V are also path connected. Any point in U can be joined to a by a path in U, and any point in V can be joined to a by a path in V, so any point in the space $X = U \cup V$ can be joined to a by a path in X. This shows that X is also path connected. We therefore see that the groups $H_0(X)$, $H_0(U)$, $H_0(V)$ and $H_0(U \cap V)$ are all the same: a copy of $\mathbb{Z}$, generated by $[a]$. $\square$

We now want to calculate the homology groups of spheres S^n for $n \geq 0$. We will do this by induction, starting with $n = 0$. Note that the point $e_0 = (1, 0, \cdots, 0)$ always lies in S^n and so gives an element $a_n = [e_0] \in H_0(S^n)$. Note also that

$$S^0 = \{x \in \mathbb{R} \mid |x| = 1\} = \{e_0, -e_0\}.$$

As in Remark 10.24, it follows that the elements $a_0 = [e_0]$ and $[-e_0]$ give a basis for $H_0(S^0)$, and that $H_k(S^0) = 0$ for $k \neq 0$. However, it turns out to be more convenient to put $b_0 = [-e_0] - [e_0] \in H_0(S^0)$; the elements a_0 and b_0 also form a basis.

Theorem 15.3. *For all $n \geq 1$ there is an element $b_n \in H_n(S^n)$ such that*

$$H_k(S^n) = \begin{cases} \mathbb{Z}a_n & \text{if } k = 0 \\ \mathbb{Z}b_n & \text{if } k = n \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Video

Recall that

$$S^n = \{(x_0, \dots, x_n) \in \mathbb{R}^{n+1} \mid \sum_i x_i^2 = 1\}.$$

We will think of a point $x \in \mathbb{R}^{n+1}$ as a pair (y, z) with $y \in \mathbb{R}^n$ and $z \in \mathbb{R}$, so $\|x\|^2 = \|y\|^2 + z^2$, and x lies in S^n iff $\|y\|^2 + z^2 = 1$. We put $N = (0, 1)$ (the "North pole") and $U = S^n \setminus \{N\} = \{(y, z) \in S^n \mid z \neq 1\}$ and $V = S^n \setminus \{-N\} = \{(y, z) \in S^n \mid z \neq -1\}$. These are open sets with $U \cup V = S^n$ and

$$U \cap V = S^n \setminus \{N, -N\} = \{(y, z) \in S^n \mid -1 < z < 1\} = \{(y, z) \in S^n \mid y \neq 0\}.$$

Just as in Example 4.13 we have a homeomorphism $f\colon U \to \mathbb{R}^n$ given by $f(y, z) = y/(1 - z)$ with $f^{-1}(y) = (2y, \|y\|^2 - 1)/(\|y\|^2 + 1)$. This proves that U is contractible. The map $(y, z) \mapsto (y, -z)$ gives a homeomorphism between U and V, so V is also contractible. It follows that $H_0(U) = H_0(V) = \mathbb{Z}$ and $H_d(U) = H_d(V) = 0$ for $d > 0$.

We also have a homeomorphism $g\colon U \cap V \to S^{n-1} \times (-1, 1)$ given by $g(y, z) = y/\|y\|$ and $g^{-1}(y, z) = (\sqrt{1 - z_2}y, z)$. (It is valid to divide by $\|y\|$ here because $y \neq 0$ whenever $(y, z) \in U \cap V$.) As $(-1, 1)$ is contractible, it follows that $U \cap V$ is homotopy equivalent to S^{n-1}. More precisely, we can define $p\colon S^{n-1} \to U \cap V$ by $p(y) = (y, 0)$, and we find that p is a homotopy equivalence, so the map $p_*\colon H_*(S^{n-1}) \to H_*(U \cap V)$ is an isomorphism.

Now consider the Mayer-Vietoris sequence

$$H_d(U) \oplus H_d(V) \to H_d(S^n) \xrightarrow{\delta} H_{d-1}(U \cap V) \to H_{d-1}(U) \oplus H_{d-1}(V).$$

(a) For $d \geq 2$ we have $H_d(U) = H_d(V) = H_{d-1}(U) = H_{d-1}(V) = 0$. Thus, Lemma 12.19 tells us that in these cases the map $\delta\colon H_d(S^n) \to H_{d-1}(U \cap V) = H_{d-1}(S^{n-1})$ is an isomorphism. This almost proves the induction step, apart from tweaks needed when $d = 0, 1$. We are assuming that $n \geq 1$ so S^n is path connected so $H_0(S^n) = \mathbb{Z}.a_n$. Thus, we just need to deal with the case $d = 1$.

(b) Using the known values of $H_d(U)$ and $H_d(V)$ for $d = 0, 1$, the bottom end of the Mayer-Vietoris sequence is as follows:

$$0 \to H_1(S^n) \xrightarrow{\delta} H_0(S^{n-1}) \xrightarrow{[\,k_* \;\; l_*\,]} \mathbb{Z} \oplus \mathbb{Z} \xrightarrow{\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right]} H_0(S^n) \to 0.$$

(c) Consider the case where $n = 1$. Point (a) tells us that for $d \geq 2$ we have $H_d(S^1) = H_{d-1}(S^0) = 0$. Point (b) gives an exact sequence

$$0 \to H_1(S^1) \xrightarrow{\delta} \mathbb{Z}\{a_0, b_0\} \xrightarrow{[\,k_* \;\; l_*\,]} \mathbb{Z} \oplus \mathbb{Z} \xrightarrow{\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right]} H_0(S^1) \to 0.$$

This means that δ gives an isomorphism from $H_1(S^1)$ to the kernel of the map $[\,k_* \;\; l_*\,]$.

Here we have identified $H_0(U)$ with $\mathbb{Z}$. More precisely, as U is connected and contains both e_0 and $-e_0$, we see that $H_0(U) = \mathbb{Z}.[e_0]$ and that $[e_0] = [-e_0]$ in $H_0(U)$. It follows that $k_*(a_0) = [e_0]$ and $k_*(b_0) = [-e_0] - [e_0] = 0$ in $H_0(U)$. For essentially the same reason we have $l_*(a_0) = [e_0]$ and $l_*(b_0) = 0$ in $H_0(V)$. Thus, the kernel of $[\,k_* \;\; l_*\,]$ is $\mathbb{Z}.b_0$. The exact sequence therefore means that there is a unique element $b_1 \in H_1(S^1)$ such that $\delta(b_1) = b_0$, and that $H_1(S^1) = \mathbb{Z}.b_1$. As S^1 is path connected, we also know that $H_0(S^1) = \mathbb{Z}.a_1$. This proves that $H_*(S^1)$ is as claimed.

(d) Now suppose instead that $n \geq 2$, and that we have already proved the claim for $H_*(S^{n-1})$. Point (a) gives an isomorphism $\delta\colon H_n(S^n) \to H_{n-1}(S^{n-1}) = \mathbb{Z}.b_{n-1}$, so there is a unique element $b_n \in H_n(S^n)$ with $\delta(b_n) = b_{n-1}$, and we have $H_n(S^n) = \mathbb{Z}.b_n$. Point (a) also shows that $H_d(S^n) = 0$ for $d \geq 2$ with $d \neq n$. As S^n is path connected, we have $H_0(S^n) = \mathbb{Z}.a_n$. This just leaves $H_1(S^n)$. The bottom end of the modified Mayer-Vietoris sequence has the form $0 = H_1(U) \oplus H_1(V) \to H_1(S^n) \to 0$, and exactness forces $H_1(S^n)$ to be zero as expected.

$\square$

Remark 15.4. If $u \in H_d(X)$ then we sometimes say that $u \in H_*(X)$ with $|u| = d$, and we call d the *degree* of u. We could then say that the elements a_n and b_n form a basis for $H_*(S^n)$ with $|a_n| = 0$ and $|b_n| = n$. This formulation is valid for $n = 0$ as well as for $n > 0$.

Video (Theorems 16.1 and 16.2)

Theorem 16.1. *No sphere S^n is contractible. Moreover, if $n \neq m$ then S^n is not homotopy equivalent to S^m.*

Proof. As homotopy equivalent spaces have isomorphic homology, it will suffice to prove that $H_n(S^n) \not\simeq H_n(\text{point})$ and that $H_n(S^m) \not\simeq H_n(S^m)$ when $n \neq m$. This is clear from the above calculations. $\square$

Theorem 16.2. *Suppose that $n, m \geq 0$ with $n \neq m$. Then $\mathbb{R}^n$ is not homeomorphic to $\mathbb{R}^m$.*

Note here that $\mathbb{R}^n$ and $\mathbb{R}^m$ are certainly homotopy equivalent, as they are both contractible. This is perfectly consistent: homeomorphism implies homotopy equivalence but not conversely.

Proof. Suppose that $n, m \geq 0$ and that we have a homeomorphism $f\colon \mathbb{R}^n \to \mathbb{R}^m$. We must show that $n = m$. If $n = 0$ then $\mathbb{R}^n$ is a single point so $\mathbb{R}^m$ is a single point so $m = 0$. Similarly, if $m = 0$ then $n = 0$. Thus, we can restrict attention to the case where $n, m \geq 1$.

Choose $a \in \mathbb{R}^n$ and put $b = f(a) \in \mathbb{R}^m$. It is easy to see that f restricts to give a homeomorphism $f_0\colon \mathbb{R}^n \setminus \{a\} \to \mathbb{R}^m \setminus \{b\}$. We can now define maps

$$S^{n-1} \underset{p}{\overset{i}{\rightleftarrows}} \mathbb{R}^n \setminus \{a\} \underset{f_0^{-1}}{\overset{f_0}{\rightleftarrows}} \mathbb{R}^m \setminus \{b\} \underset{j}{\overset{q}{\rightleftarrows}} S^{m-1}$$

by

$$\begin{aligned} i(x) &= x + a & p(x) &= (x-a)/\|x-a\| \\ j(y) &= y + b & q(y) &= (y-b)/\|y-b\|. \end{aligned}$$

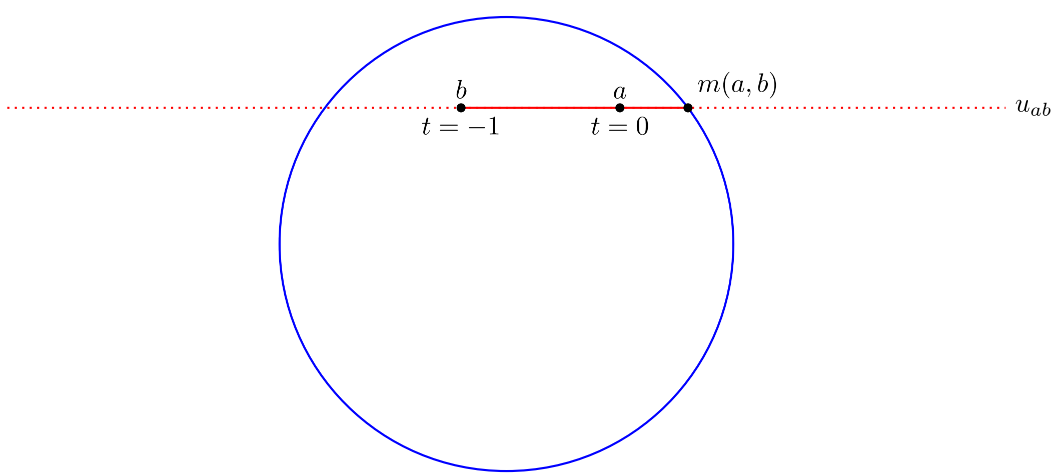
A tiny adaptation of Proposition 9.12 shows that p and q are homotopy inverses for i and j respectively, so that all maps in the above diagrams are homotopy equivalences, so S^n and S^m are homotopy equivalent. It follows by Theorem 16.1 that $n = m$ as required. $\square$

Video (Theorem 16.3 to Lemma 16.5)

Theorem 16.3 (Brouwer Fixed Point Theorem)**.** *Let $f\colon B^n \to B^n$ be a continuous map (for some $n > 0$). Then there is a point $a \in B^n$ such that $f(a) = a$.*

The proof will rely on the following construction.

Definition 16.4. Put $X_n = \{(a, b) \in B^n \times B^n \mid a \neq b\}$. For $(a, b) \in X_n$ we consider the map $u_{ab}\colon \mathbb{R} \to \mathbb{R}^n$ given by $u_{ab}(t) = a + t(a - b)$, so u_{ab} traces out the straight line joining b to a, with $u_{ab}(-1) = b$ and $u_{ab}(0) = a$.



It is geometrically clear that this line crosses the sphere S^{n-1} at precisely two points, one with $t \leq -1$ and the other with $t \geq 0$. We define $m(a,b)$ to be the intersection point with $t \geq 0$.

Lemma 16.5. *The map $m\colon X_n \to S^{n-1}$ is continuous, and it satisfies $m(a,b) = a$ if $\|a\| = 1$.*

Proof. It is possible to argue geometrically, but more efficient to just find the formula for the point $c = m(a,b)$. Put $v = a - b$, so $c = tv + a$ for some $t \geq 0$, and must satisfy $\langle c, c\rangle = 1$. Expanding this out, we get

$$\|v\|^2 t^2 + 2\langle v, a\rangle t + \|a\|^2 - 1 = 0.$$

The quadratic formula tells us that the positive root is

$$t_+ = (\sqrt{\langle v, a\rangle^2 + (1 - \|a\|^2)\|v\|^2} - \langle v, a\rangle)/\|v\|^2.$$

Note that the quantity under the square root is nonnegative, because squares are always nonnegative and $a \in B^n$ so $1 - \|a\|^2 \geq 0$. Also, the vector $v = a - b$ is nonzero by the definition of X_n so $\|v\|^2 > 0$ so it is harmless to divide by $\|v\|^2$. This shows that t_+ is a well-defined continuous function of the pair (a,b). It follows that the function $m(a,b) = a + t_+ b$ is also continuous.

It is clear from the geometry that if $\|a\| = 1$ (so a lies on the unit sphere S^{n-1}) then $m(a,b) = a$. Alternatively, it is clear in this case that $t = 0$ is a nonnegative root of our quadratic, so it must be the same as t_+. □

Proof of Theorem 16.3. Suppose, for a contradiction, that we have a continuous map $f\colon B^n \to B^n$ with no fixed points. This means that for any $a \in B^n$ we have $(a, f(a)) \in X_n$ so we can define $r(a) = m(a, f(a)) \in S^{n-1}$. If $\|a\| = 1$, then this is just $r(a) = a$. This means that S^{n-1} is a retract (and thus a homotopy retract) of B^n. As B^n is contractible, we can use Proposition 9.24 to see that S^{n-1} is also contractible, but this contradicts Theorem 16.1. □

Video (Theorem 16.6)

Theorem 16.6 (Fundamental Theorem of Algebra)**.** *Let $p(x) \in \mathbb{C}[x]$ be a non-constant complex polynomial. Then $p(x)$ has a complex root.*

There are many different ways to prove this theorem. We will give a proof using homology.

Proof. Consider a non-constant polynomial $p(x)$ of degree $n > 0$, so

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$

for some coefficients a_i with $a_n \neq 0$. Suppose, for a contradiction that $p(x)$ is never zero. Choose some very large radius R and define $h\colon [0,1] \times S^1 \to \mathbb{C} \setminus \{0\}$ by $h(s,z) = p(Rsz)/p(Rs)$. As we are assuming that p is never zero, the division is valid and $h(s,z)$ lies in $\mathbb{C} \setminus \{0\}$ as required. Put $f(z) = h(0,z) = 1$ and $g(z) = h(1,z) = p(Rz)/p(R)$. As f is constant, the map $f_*\colon H_1(S^1) \to H_1(\mathbb{C} \setminus \{0\})$ is zero. As h gives a homotopy between f and g, the map $g_*\colon H_1(S^1) \to H_1(\mathbb{C} \setminus \{0\})$ must also be zero. However, as R is very large and $a_n \neq 0$, the term $a_n(Rz)^n$ will be much larger than all the other terms in $p(Rz)$, so $g(z) = p(Rz)/p(R) \approx a_n(Rz)^n/(a_n R^n) = z^n$. Thus, if we put $q(z) = z^n$ then $g(z)$ will be very close to $q(z)$ for all $z \in S^1$, so the straight line from $g(z)$ to $q(z)$ will never pass through 0, so g will be linearly homotopic to q, so $q_* = g_* = 0\colon H_1(S^1) \to H_1(\mathbb{C} \setminus \{0\})$. However, it is clear from our earlier discussions in Section 11 that $H_1(S^1) = H_1(\mathbb{C} \setminus \{0\}) = \mathbb{Z}$ and q_* sends 1 to $n \neq 0$, so we have a contradiction. □

17. The Snake Lemma

We now return to the task of constructing the Mayer-Vietoris sequence. There are two key ingredients: the Snake Lemma (in this section) and subdivision (in the next section). The videos cover this material in a slightly different order than the notes: the first video is attached to Definition 17.3 below.

The basic input for the Snake Lemma is as follows: we have chain complexes U_*, V_* and W_* and chain maps

$$U_* \xrightarrow{i} V_* \xrightarrow{p} W_*$$

which form a short exact sequence. One might hope that the resulting sequence

$$H_*(U) \xrightarrow{i_*} H_*(V) \xrightarrow{p_*} H_*(W)$$

would also be a short exact sequence, but that is not quite right. We will show that the above sequence is exact (in the sense that $\operatorname{img}(i_*) = \ker(p_*)$), but i_* need not be injective, and p_* need not be surjective. In other words, $\ker(i_*)$ need not be zero, and $\operatorname{img}(p_*)$ need not be all of $H_*(W)$. We can still obtain a great deal of information about $\ker(i_*)$ and $\operatorname{img}(p_*)$, but that will require some preparation. For the moment we will just prove the easier statement mentioned above.

Proposition 17.1. *Let $U_* \xrightarrow{i} V_* \xrightarrow{p} W_*$ be a short exact sequence of chain maps between chain complexes. Then in the resulting sequence $H_*(U) \xrightarrow{i_*} H_*(V) \xrightarrow{p_*} H_*(W)$ we have $\operatorname{img}(i_*) = \ker(p_*)$.*

Proof. First, as the sequence $U_* \xrightarrow{i} V_* \xrightarrow{p} W_*$ is exact we have $p \circ i = 0$. It follows that $p_* \circ i_* = (p \circ i)_* = 0_* = 0$, so $\operatorname{img}(i_*) \leq \ker(p_*)$.

Conversely, suppose we are given an element $b \in \ker(p_*)$; must show that it lies in $\operatorname{img}(i_*)$. If $b \in H_r(V)$ then we have $b = [v]$ for some $v \in V_r$ with $d(v) = 0$. We are assuming that $p_* b = 0$, which means that $[p(v)]$ is zero in the quotient group $Z_r(W)/B_r(W)$, which means that $p(v) \in B_r(W)$, which means that $p(v) = d(w')$ for some $w' \in W_{r+1}$. Also, we are assuming that the sequence $U_* \xrightarrow{i} V_* \xrightarrow{p} W_*$ is short exact, which means in particular that p is surjective. We can therefore choose $v' \in V_{r+1}$ with $p(v') = w'$. We now have

$$p(v - d(v')) = p(v) - p(d(v')) = p(v) - d(p(v')) = p(v) - d(w') = 0,$$

so $v - d(v') \in \ker(p)$. We also have $\ker(p) = \operatorname{img}(i)$ by our exactness assumption, so we can find $u \in U_r$ with $i(u) = v - d(v')$. From our initial assumptions we have $d(v) = 0$, and also $d^2 = 0$ so $d(d(v')) = 0$, so $d(i(u)) = 0$. As i is a chain map this gives $i(d(u)) = 0$, and i is injective so $d(u) = 0$. This means we have an element $a = [u] \in H_r(U)$. This satisfies $i_*(a) = [i(u)] = [v - d(v')]$ but $d(v') \in B_r(V)$ so $[v - d(v')]$ is the same as $[v]$, which is b. We conclude that $i_*(a) = b$, so $b \in \operatorname{img}(i_*)$ as claimed.

We can display the relevant groups and elements as follows:

$$\begin{tikzcd}
 & & & v' \arrow[rr, mapsto] \arrow[ddd, dotted] & & w' \arrow[ddd, mapsto] \\
U_{r+1} \arrow[rr, tail, "i"] \arrow[d, "d"] & & V_{r+1} \arrow[rr, two heads, "p"] \arrow[d, "d"] & & W_{r+1} \arrow[d, "d"] & \\
U_r \arrow[rr, tail, "i"] & & V_r \arrow[rr, two heads, "p"] & & W_r & \\
 & u \arrow[rr, dotted] & & v \arrow[rr, mapsto] & & p(v) = d(w')
\end{tikzcd}$$

The two dotted arrows are supposed to indicate the relation $d(v') + i(u) = v$. ∎

Theorem 17.2. *For $U_* \xrightarrow{i} V_* \xrightarrow{p} W_*$ as above, there is a natural map $\delta \colon H_n(W) \to H_{n-1}(U)$ such that the sequence*

$$\begin{tikzcd}
H_{n+1}U \arrow[r, "i_*"] & H_{n+1}V \arrow[r, "p_*"] & H_{n+1}W \arrow[dll, rounded corners, "\delta"] \\
H_nU \arrow[r, "i_*"] & H_nV \arrow[r, "p_*"] & H_nW \arrow[dll, rounded corners, "\delta"] \\
H_{n-1}U \arrow[r, "i_*"] & H_{n-1}V \arrow[r, "p_*"] & H_{n-1}W
\end{tikzcd}$$

is exact for all n.

The proof will be broken into a number of steps. The map δ will be defined in Definition 17.6, and Propositions 17.1, 17.10 and 17.11 will show that the resulting long sequence is exact.

Definition 17.3. Video

A *snake* for the above sequence is a system (c, w, v, u, a) such that

- $c \in H_n(W)$;
- $w \in Z_n(W)$ is a cycle such that $c = [w]$;
- $v \in V_n$ is an element with $p(v) = w$;
- $u \in Z_{n-1}(U)$ is a cycle with $i(u) = d(v) \in V_{n-1}$;
- $a = [u] \in H_{n-1}(U)$.

More specifically, we say that a system (c, w, v, u, a) as above is a *snake from c to a*.

Video (Lemma 17.4 to Remark 17.8)

Lemma 17.4. *For any $c \in H_n(W)$, there is a snake starting with c.*

Proof. Consider an element $c \in H_n(W)$. As $H_n(W) = Z_n(W)/B_n(W)$ by definition, we can certainly choose $w \in Z_n(W)$ such that $c = [w]$. As the sequence $U_* \xrightarrow{i} V_* \xrightarrow{p} W_*$ is short exact, we know that $p\colon V_n \to W_n$ is surjective, so we can choose $v \in V_n$ with $p(v) = w$. As p is a chain map we have $p(d(v)) = d(p(v)) = d(w) = 0$ (the last equation because $w \in Z_n(W)$). This means that $d(v) \in \ker(p)$, but $\ker(p) = \operatorname{img}(i)$ because the sequence is exact, so we have $u \in U_{n-1}$ with $i(u) = d(v)$. Note also that $i(d(u)) = d(i(u)) = d(d(v)) = 0$ (because i is a chain map and $d^2 = 0$). On the other hand, exactness means that i is injective, so the relation $i(d(u)) = 0$ implies that $d(u) = 0$. This shows that $u \in Z_{n-1}(U)$, so we can put $a = [u] \in H_{n-1}(U)$. We now have a snake (c, w, v, u, a) starting with c as required. $\square$

Lemma 17.5. *Suppose we have two snakes that have the same starting point; then they also have the same endpoint.*

Proof. Suppose we have two snakes that start with c. We can then subtract them to get a snake $(0, w, v, u, a)$ starting with 0. It will be enough to show that this ends with 0 as well, or equivalently that $a = 0$. The first snake condition says that $[w] = 0$, which means that $w = d(w')$ for some $w' \in W_{n+1}$. Because p is surjective we can also choose $v' \in V_{n+1}$ with $w' = p(v')$, and this gives $w = d(w') = d(p(v')) = p(d(v'))$. The next snake condition says that $p(v) = w$. We can combine these facts to see that $p(v - d(v')) = 0$, so $v - d(v') \in \ker(p) = \operatorname{img}(i)$. We can therefore find $u' \in U_n$ with $v - d(v') = i(u')$. We can apply d to this using $d^2 = 0$ and $di = id$ to get $d(v) = i(d(u'))$. On the other hand, the third snake condition tells us that $d(v) = i(u)$. Subtracting these gives $i(u - d(u')) = 0$, but i is injective, so $u = d(u')$, so $u \in B_{n-1}(U)$. The final snake condition now says that $a = [u] = u + B_{n-1}(U)$, but $u \in B_{n-1}(U)$ so $a = [u] = 0$. $\square$

Definition 17.6. For any $c \in H_n(W)$, we define $\delta(c) \in H_{n-1}(U)$ to be the endpoint of any snake that starts with c. (This is well-defined by the last two lemmas.)

Remark 17.7. It is easy to see that the sum of two snakes is a snake, and from that we can deduce that δ is a homomorphism.

Remark 17.8. The slogan behind the definition is that $\delta = i^{-1}dp^{-1}$. In more detail, suppose we have $c \in H_n(W)$. To calculate $\delta(c)$, we must find a snake of the form (c, w, v, u, a), then $\delta(c) = a$. The slogan glosses over the distinction between w and $c = [w]$, and the distinction between u and $a = [u]$. The condition $p(v) = w$ means that v is a choice of $p^{-1}w$, and the condition $i(u) = d(v)$ means that u is essentially $i^{-1}(d(v)) = i^{-1}(d(p^{-1}(w)))$. The point of the above definitions and lemmas is to make this slogan precise.

Remark 17.9. The Snake Lemma (in a slightly different incarnation) is probably the most advanced piece of mathematics ever to appear in a mainstream movie:

Video

Video (Proposition 17.1, 17.10 and 17.11)

Proposition 17.10. *The sequence $H_n(V) \xrightarrow{p_*} H_n(W) \xrightarrow{\delta} H_{n-1}(U)$ is exact (or equivalently, $\operatorname{img}(p_*) = \ker(\delta)$).*

Proof. First, suppose that $b \in H_n(V)$, so $b = [v]$ for some $v \in V_n$ with $d(v) = 0$. We find that $(p_*(b), p(v), v, 0, 0)$ is a snake starting with $p_*(b)$, so $\delta(p_*(b)) = 0$. From this we get $\delta \circ p_* = 0$ and $\mathrm{img}(p_*) \leq \ker(\delta)$.

Conversely, consider an element $c \in \ker(\delta) \leq H_n(W)$. As $c \in \ker(\delta)$, there must exists a snake of the form $(c, w, v, u, 0)$. The last snake condition says that $[u] = 0$, so we must have $u = d(u')$ for some $u' \in U_n$. Another snake condition says that $d(v) = i(u) = i(d(u')) = d(i(u'))$, so we have $d(v - i(u')) = 0$. This means that $v - i(u')$ is a cycle, so we have a homology class $b = [v - i(u')] \in H_n(V)$. This satisfies $p_*(b) = [p(v - i(u'))]$, but $pi = 0$ and $p(v) = w$ so this simplifies to $p_*(b) = [w] = c$, so $c \in \mathrm{img}(p_*)$. □

Proposition 17.11. *The sequence* $H_n(W) \xrightarrow{\delta} H_{n-1}(U) \xrightarrow{i_*} H_{n-1}(V)$ *is exact (or equivalently,* $\mathrm{img}(\delta) = \ker(i_*)$*).*

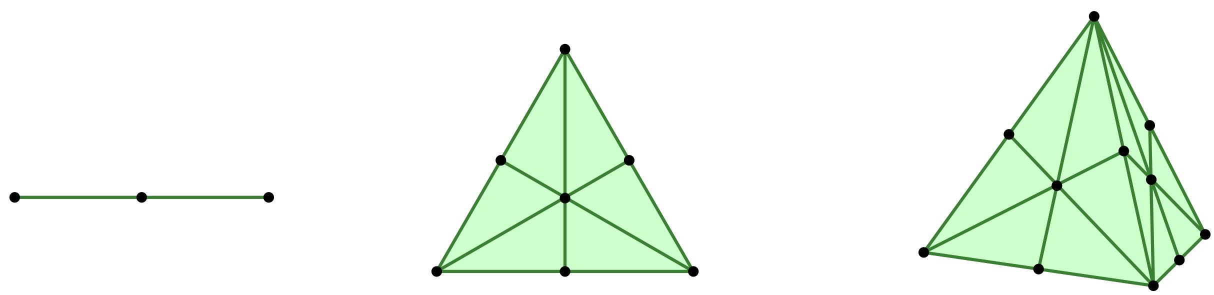
Proof. First suppose we have an element $c \in H_n(W)$. Choose a snake (c, w, v, u, a) starting with c, so $\delta(c) = a = [u]$. We then have $i_*\delta(c) = i_*[u] = [i(u)]$, but one of the snake conditions says that $i(u) = d(v) \in B_{n-1}(V)$, so $[i(u)] = 0$, so $i_*\delta(c) = 0$. This proves that $i_* \circ \delta = 0$ and so $\mathrm{img}(\delta) \leq \ker(i_*)$.

Conversely, suppose that $a \in \ker(i_*)$. We can choose $u \in Z_{n-1}(U)$ such that $a = [u]$. Now $[i(u)] = i_*(a) = 0$, so $i(u) \in B_{n-1}(V)$, so there exists $v \in V_n$ with $d(v) = i(u)$. Put $w = p(v) \in W_n$. We then have $d(w) = d(p(v)) = p(d(v)) = p(i(u))$, and this is zero because $p \circ i = 0$. This means that $w \in Z_n(W)$, so we can define $c = [w] \in H_n(W)$. We now see that (c, w, v, u, a) is a snake, so $a = \delta(c)$, so $a \in \mathrm{img}(\delta)$. □

18. Subdivision

Video (Introduction to subdivision)

Consider the following pictures:



Δ_1 Δ_2 Δ_3

Interactive demo

On the left we have the 1-simplex Δ_1, divided into two pieces. In Lemma 10.27 we showed that if u and v are joinable paths in X, then $u * v = u + v \pmod{B_1(X)}$. Equivalently, if we start with a path w and split it in the middle to get two paths u and v, then $w = u + v \pmod{B_1(X)}$. Thus, there is a sense in which subdivision of paths acts as the identity in homology.

In the middle picture we have divided the simplex Δ_2 into 6 pieces. In the right hand picture, and the interactive demonstration, we have divided Δ_3 into 24 pieces. It will again turn out that this kind of subdivision acts as the identity in homology. To prove this, we need to study the combinatorics of the subdivision process.

We subdivide Δ_1 by introducing a new vertex in the middle (which we call the *barycentre*), giving two copies of Δ_1. To subdivide Δ_2, we first subdivide each of the 3 edges in the same way as Δ_1, giving $3 \times 2 = 6$ edges altogether. We then take each of these subdivided edges and connect it to the barycentre of Δ_2; this divides Δ_2 into 2×3 copies of Δ_2. Next, we divide all 4 of the faces of Δ_3 in the same way as Δ_2, giving $4 \times 3 \times 2 = 24$ triangles on the surface of Δ_3. We connect all of these to the barycentre of Δ_3; this divides Δ_3 into $4 \times 3 \times 2$ copies of Δ_3. We can continue in the same way to divide Δ_n into $(n+1)!$ copies of Δ_n. We now start to make this more formal.

Video (Definition 18.1 to Proposition 18.10)

Definition 18.1.

(a) The *barycentre* of Δ_n is the point $b_n = (1, \dots, 1)/(n+1) \in \Delta_n$ (so $b_3 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, for example). We will write b instead of b_n if there is no danger of confusion.

(b) Given any linear k-simplex $u = \langle a_0, \dots, a_k \rangle \in C_k(\Delta_n)$, we define

$$\beta\langle a_0, \dots, a_k \rangle = \langle b, a_0, \dots, a_k \rangle \in C_{k+1}(\Delta_n).$$

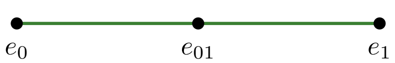
More generally, if $u = n_1 u_1 + \cdots + n_r u_r$ with $n_1, \dots, n_r \in \mathbb{Z}$ and each u_i being a linear k-simplex, we define $\beta(u) = n_1\beta(u_1) + \cdots + n_r\beta(u_r)$.

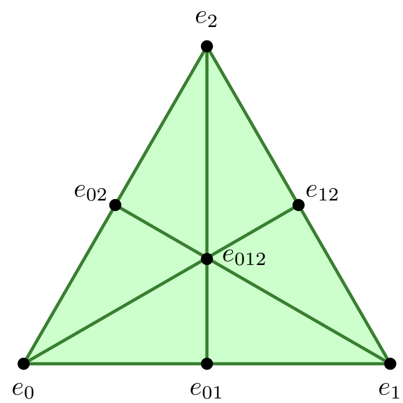
Remark 18.2. It is possible to define β for nonlinear k-simplices, but a little work is required to check that the resulting maps $\Delta_{k+1} \to \Delta_n$ are always continuous. We do not need the general case so we omit it.

Lemma 18.3. *Let u be a linear combination of linear k-simplices in Δ_n with $k > 0$. Then $\partial\beta(u) + \beta\partial(u) = u$.*

Proof. We can easily reduce to the case where u is a single linear k-simplex, say $u = \langle a_0, \dots, a_k \rangle$. Let u_i be the same as u except that a_i is omitted, so $\partial(u) = \sum_i (-1)^i u_i$, so $\beta\partial(u) = \sum_i (-1)^i \beta(u_i)$. On the other hand, we have $\beta(u) = \langle b, a_0, a_1, \dots, a_k \rangle$. For the initial term in $\partial\beta(u)$ we omit the b and we have a sign $(-1)^0$; this just gives us u. For each of the remaining terms in $\partial\beta(u)$ we omit the a_i appearing in position $i+1$ of $\beta(u)$, and multiply by $(-1)^{i+1}$; this gives $-(-1)^i\beta(u_i)$, which cancels with a term in $\beta\partial(u)$. Putting everything together gives $\partial\beta(u) + \beta\partial(u) = u$ as claimed. □

We now want to define certain elements $\theta_n \in C_n(\Delta_n)$ for all n. The idea is that we subdivide Δ_n into smaller copies of Δ_n as sketched previously, and take θ_n to be the sum of these smaller copies with suitable $\pm$-signs to make the orientations match up correctly. We can mark some points in Δ_1 and Δ_2 as follows:





It will work out that

$\theta_0 = \langle e_0 \rangle$

$\theta_1 = \langle e_{01}, e_1 \rangle - \langle e_{01}, e_0 \rangle$

$\theta_2 = \langle e_{012}, e_{12}, e_2 \rangle - \langle e_{012}, e_{12}, e_1 \rangle - \langle e_{012}, e_{02}, e_2 \rangle + \langle e_{012}, e_{02}, e_0 \rangle + \langle e_{012}, e_{01}, e_1 \rangle - \langle e_{012}, e_{01}, e_0 \rangle$

The general picture is as follows.

Definition 18.4. We start with $\theta_0 = \langle e_0 \rangle \in C_0(\Delta_0)$. Now suppose that we have $n > 0$ and we have already defined an element $\theta_{n-1} \in C_{n-1}(\Delta_{n-1})$ which is a $\mathbb{Z}$-linear combination of linear simplices. For $i = 0, \dots, n$ we have an affine map $\delta_i \colon \Delta_{n-1} \to \Delta_n$ and thus a chain $(\delta_i)_\#(\theta_{n-1}) \in C_{n-1}(\Delta_n)$, which is again a $\mathbb{Z}$-linear combination of linear simplices. We put

$$\theta'_n = \sum_{i=0}^{n} (-1)^i (\delta_i)_\#(\theta_{n-1}) \in C_{n-1}(\Delta_n)$$

$$\theta_n = \beta(\theta'_n) = \sum_{i=0}^{n} (-1)^i \beta((\delta_i)_\#(\theta_{n-1})) \in C_n(\Delta_n).$$

This defines θ_n for all n by recursion. Next, suppose we have a space X and a map $u\colon \Delta_n \to X$, so $u \in S_n(X) \subset C_n(X)$. The map $u\colon \Delta_n \to X$ then gives a map $u_\#\colon C_n(\Delta_n) \to C_n(X)$, and we define $\operatorname{sd}(u) = u_\#(\theta_n)$.

Example 18.5. The map $\delta_0\colon \Delta_1 \to \Delta_2$ sends e_0, e_1 and e_{01} to e_1, e_2 and e_{12} respectively. It follows that

$$\begin{aligned}(\delta_0)_\#(\theta_1) &= (\delta_0)_\#(\langle e_{01}, e_1\rangle - \langle e_{01}, e_0\rangle) = \langle e_{12}, e_2\rangle - \langle e_{12}, e_1\rangle \\ \beta((\delta_0)_\#(\theta_1)) &= \langle e_{012}, e_{12}, e_2\rangle - \langle e_{012}, e_{12}, e_1\rangle.\end{aligned}$$

After expressing $\beta((\delta_1)_\#(\theta_1))$ and $\beta((\delta_2)_\#(\theta_1))$ in the same way, we obtain the advertised formula for θ_2:

$$\theta_2 = \langle e_{012}, e_{12}, e_2\rangle - \langle e_{012}, e_{12}, e_1\rangle - \langle e_{012}, e_{02}, e_2\rangle + \langle e_{012}, e_{02}, e_0\rangle + \langle e_{012}, e_{01}, e_1\rangle - \langle e_{012}, e_{01}, e_0\rangle.$$

Example 18.6. Suppose we have a path $u\colon \Delta_1 \to X$. We identify Δ_1 with $[0,1]$ as usual, so the points e_0, e_1 and e_{01} become 0, 1 and $\frac{1}{2}$ respectively. The map $\langle e_{01}, e_1\rangle\colon \Delta_1 \to \Delta_1$ is thus $t \mapsto (1+t)/2$, and the map $\langle e_{01}, e_0\rangle\colon \Delta_1 \to \Delta_1$ is $t \mapsto (1-t)/2$. This means that $\operatorname{sd}(u) = v - w$, where $v(t) = u((1+t)/2)$ and $w(t) = ((1-t)/2)$. In other words, v is the second half of u and w is the reverse of the first half of u, so $u = \overline{w} * v$.

Remark 18.7. An alternative approach is as follows. Let π be a permutation of $\{0,\dots,n\}$. For $0 \leq i \leq n$ we put $e_i^\pi = (n-i+1)^{-1}\sum_{j=i}^n e_{\pi(j)} \in \Delta_n$. This gives a linear n-simplex $u_\pi = \langle e_0^\pi, \dots, e_n^\pi\rangle \in S_n\Delta_n$. It can be shown that $\theta_n = \sum_\pi \operatorname{sgn}(\pi) u_\pi \in C_n(\Delta_n)$. We could instead have taken this formula as the definition of θ_n; that would make some things easier and some other things harder.

Lemma 18.8. *For any $f\colon X \to Y$ and any $u \in C_n(X)$ we have $f_\#(\operatorname{sd}(u)) = \operatorname{sd}(f_\#(u))$ in $C_n(Y)$.*

Proof. We can easily reduce to the case where $u \in S_n(X)$, or in other words $u\colon \Delta_n \to X$. We then have

$$f_\#(\operatorname{sd}(u)) = f_\#(u_\#(\theta_n)) = (f \circ u)_\#(\theta_n) = \operatorname{sd}(f_\#(u)).$$

□

Lemma 18.9. *If we let ι_n denote the identity map $\Delta_n \to \Delta_n$ considered as an element of $C_n(\Delta_n)$, then we have $\theta_n' = \operatorname{sd}(\partial(\iota_n))$, and therefore $\theta_n = \beta(\operatorname{sd}(\partial(\iota_n)))$.*

Proof. We have $\partial(\iota_n) = \sum_{i=0}^n (-1)^i(\iota_n \circ \delta_i) = \sum_{i=0}^n (-1)^i \delta_i$. By definition sd is linear and has $\operatorname{sd}(\delta_i) = (\delta_i)_\#(\theta_{n-1})$. It follows that $\operatorname{sd}(\partial(\iota_n)) = \sum_{i=0}^n (-1)^i \operatorname{sd}(\delta_i) = \sum_{i=0}^n (\delta_i)_\#(\theta_{n-1}) = \theta_n'$ as claimed. □

Proposition 18.10. *The map $\operatorname{sd}\colon C_*(X) \to C_*(X)$ is a chain map.*

Proof. We must show that for all $n \geq 0$ and all $u \in C_n(X)$ we have $\partial(\operatorname{sd}(u)) = \operatorname{sd}(\partial(u))$ in $C_{n-1}(X)$. If $n = 0$ then $C_{n-1}(X) = 0$ and so the claim is clear. For $n > 0$ we will argue by induction. We can easily reduce to the case where $u \in S_n(X) \subset C_n(X)$, or in other words $u\colon \Delta_n \to X$. We then have $\operatorname{sd}(u) = u_\#(\theta_n)$, and $u_\#$ is a chain map, so $\partial(\operatorname{sd}(u)) = \partial(u_\#(\theta_n)) = u_\#(\partial(\theta_n))$. On the other hand, we have

$$\operatorname{sd}(\partial(u)) = \sum_{i=0}^n (-1)^i \operatorname{sd}(u \circ \delta_i) = \sum_{i=0}^{n-1} (-1)^i (u \circ \delta_i)_\#(\theta_{n-1}) = u_\#(\theta_n').$$

It will therefore be enough to prove that $\partial(\theta_n) = \theta_n'$.

We are assuming inductively that $\partial(\operatorname{sd}(v)) = \operatorname{sd}(\partial(v))$ for all $v \in C_{n-1}(X)$. We can take $v = \partial(\iota_n)$, so $\partial(v) = \partial^2(\iota_n) = 0$; it follows that $\partial(\operatorname{sd}(\partial(\iota_n))) = 0$, or in other words $\partial(\theta_n') = 0$. We also know from Lemma 18.3 that $\partial(\beta(\theta_n')) + \beta(\partial(\theta_n')) = \theta_n'$. As $\beta(\theta_n') = \theta_n$ and $\partial(\theta_n') = 0$ this can be rewritten as $\partial(\theta_n) = \theta_n'$, as required. □

Proposition 18.11. *The chain map $\operatorname{sd}\colon C_*(X) \to C_*(X)$ is chain-homotopic to the identity.*

Proof. Video

We define chains $\kappa_n \in C_{n+1}(\Delta_n)$ recursively as follows. We start with $\kappa_0 = 0$. Now suppose that $n > 0$ and we have already defined $\kappa_{n-1} \in C_n(\Delta_{n-1})$. For $0 \leq i \leq n$ we have a face inclusion $\delta_i\colon \Delta_{n-1} \to \Delta_n$, using which we can form $(\delta_i)_\#(\kappa_{n-1}) \in C_n(\Delta_n)$. We put

$$\kappa_n' = \iota_n - \theta_n - \sum_{i=0}^{n} (-1)^i (\delta_i)_\#(\kappa_{n-1}) \in C_n(\Delta_n),$$

and $\kappa_n = \beta(\kappa_n') \in C_{n+1}(\Delta_n)$; this completes the recursion step.

Next, given $u \in S_n(X)$ we note that $u\colon \Delta_n \to X$ and $\kappa_n \in C_{n+1}(\Delta_n)$ so $u_\#(\kappa_n) \in C_{n+1}(X)$. We define $\sigma(u) = u_\#(\kappa_n)$, and extend this linearly to get $\sigma\colon C_n(X) \to C_{n+1}(X)$. We will prove that this gives the required chain homotopy.

As a first step, we will reformulate the definition of κ_n'. We have $(\delta_i)_\#(\kappa_{n-1}) = \sigma(\delta_i)$ so

$$\sum_{i=0}^{n} (-1)^i (\delta_i)_\#(\kappa_{n-1}) = \sum_{i=0}^{n} (-1)^i \sigma(\delta_i) = \sigma\left(\sum_{i=0}^{n} (-1)^i \delta_i\right) = \sigma(\partial(\iota_n)).$$

It follows that

$$\kappa_n' = \iota_n - \theta_n - \sigma(\partial(\iota_n)) = \iota_n - \mathrm{sd}(\iota_n) - \sigma(\partial(\iota_n)).$$

We will now prove by induction that $\partial(\sigma(u)) + \sigma(\partial(u)) = u - \mathrm{sd}(u)$ for all spaces X and all $u \in C_n(X)$. When $n = 0$ the claim is just that $0 = 0$, which is true. Suppose we have proved the claim for $n-1$. We can then apply it to the element $\partial(\iota_n) \in C_{n-1}(\Delta_n)$; we find that

$$\partial(\sigma(\partial(\iota_n))) + \sigma(\partial(\partial(\iota_n)) = \partial(\iota_n) - \mathrm{sd}(\partial(\iota_n)).$$

Using $\partial^2 = 0$ and $\partial\,\mathrm{sd} = \mathrm{sd}\,\partial$ we can rewrite this as

$$\partial\left(\iota_n - \mathrm{sd}(\iota_n) - \sigma(\partial(\iota_n))\right) = 0,$$

or in other words $\partial(\kappa_n') = 0$. We can therefore take $u = \kappa_n'$ in Lemma 18.3 to get $\partial(\beta(\kappa_n')) = \kappa_n'$. After recalling our formula above for κ_n' and the definition $\sigma(\iota_n) = \kappa_n = \beta(\kappa_n')$ we get

$$\partial(\sigma(\iota_n)) = \iota_n - \mathrm{sd}(\iota_n) - \sigma(\partial(\iota_n)) \in C_n(\Delta_n).$$

Now suppose we have a map $u\colon \Delta_n \to X$. We apply $u_\#$ to the above equation, noting that $u_\#\partial = \partial u_\#$ and $u_\#\,\mathrm{sd} = \mathrm{sd}\,u_\#$ and $u_\#\sigma = \sigma u_\#$ and $u_\#(\iota_n) = u$. We get

$$\partial(\sigma(u)) = u - \mathrm{sd}(u) - \sigma(\partial(u)),$$

or equivalently $\partial(\sigma(u)) + \sigma(\partial(u)) = u - \mathrm{sd}(u)$. We have proved this for $u \in S_n(X)$, but it follows by linearity for all $u \in C_n(X)$, as required. □

Video (Definition 18.12 and Lemma 18.13)

Definition 18.12. Let $u\colon \Delta_n \to \mathbb{R}^N$ be a linear simplex. We define

$$\mathrm{diam}(u) = \max\{\|u(s) - u(t)\| \mid s, t \in \Delta_n\} = \max\{\|u(e_i) - u(e_j)\| \mid 0 \leq i, j \leq n\},$$

and we call this the *diameter* of u. More generally, given a chain $u = m_1u_1 + \cdots + m_ru_r \in C_n(\mathbb{R}^N)$ we put $\mathrm{diam}(u) = \max(\mathrm{diam}(u_1), \ldots, \mathrm{diam}(u_r))$.

Lemma 18.13. *If $u \in C_n(\mathbb{R}^N)$ is a $\mathbb{Z}$-linear combination of linear simplices then we have* $\mathrm{diam}(\mathrm{sd}(u)) \leq \frac{n}{n+1}\,\mathrm{diam}(u)$.

Proof. In the case $n = 0$ all diameters are zero so the claim is clear. We can therefore assume that $n > 0$ and argue by induction. The claim involves the number $c_n = n/(n+1) = 1 - (n+1)^{-1}$; from the second form it is clear that $0 \leq c_n < c_{n+1} < 1$.

We can easily reduce to the case where u is a single linear simplex, say $u = \langle a_0, \ldots, a_n\rangle$. Put $d = \mathrm{diam}(u)$, so $\|a_i - a_j\| \leq d$ for all i, j. Put $b = (a_0 + \cdots + a_n)/(n+1)$, which is the barycentre of u. For any i we can write a_i as $(n+1)^{-1}\sum_{j=0}^{n} a_i$. Using this, we get $a_i - b = (n+1)^{-1}\sum_{j=0}^{n}(a_i - a_j)$. In the sum on right hand side, the term for $j = i$ is zero and the other n terms have norm at most d; it follows that $\|a_i - b\| \leq \frac{n}{n+1}d = c_nd$. More generally, consider a point $x \in u(\Delta_n)$, say $x = \sum_{i=0}^{n} t_ia_i$ with $t_i \geq 0$ and $\sum_i t_i = 1$. We can write b as $\sum_i t_ib$, so

$$\|x - b\| = \|\sum_i t_i(a_i - b)\| \leq \sum_i t_i\|a_i - b\| \leq \sum_i t_ic_nd = c_nd.$$

Now let v be a simplex occuring in $\mathrm{sd}(u)$. Then $v = \beta(w)$ for some w occuring in $\mathrm{sd}(u \circ \delta_i)$ for some i, so the vertices of v are the vertices of w together with b. It is clear that $\mathrm{diam}(u \circ \delta_i) \leq d$ so by induction we have $\mathrm{diam}(w) \leq c_{n-1}d \leq c_n d$. Also, from the discussion above, any vertex in w has distance at most $c_n d$ from b. It follows that v has diameter at most $c_n d$ as required. □

19. Construction of the Mayer-Vietoris sequence

Video (All of Section 19)

This section will constitute the proof of Theorem 15.1. As before, let X be a topological space, and let U and V be open sets with $X = U \cup V$. We name the inclusion maps as shown below:

$$\begin{tikzcd} U \cap V \arrow[r, "i"] \arrow[d, "j"'] & U \arrow[d, "k"] \\ V \arrow[r, "l"'] & U \cup V = X \end{tikzcd}$$

This gives a sequence of chain maps as follows:

$$C_*(U \cap V) \xrightarrow{\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right]} C_*(U) \oplus C_*(V) \xrightarrow{[\,k_*\ l_*\,]} C_*(U \cup V).$$

If this was a short exact sequence, then we could apply Theorem 17.2 to get the Mayer-Vietoris sequence. Unfortunately, however, this is not quite a short exact sequence: we will see that the first map is injective and the image of the first map is the kernel of the second one, but the second map is not surjective. We will need an extra step involving subdivision to deal with this issue.

Definition 19.1. Put

$$\begin{aligned} S'_n(U) &= \{u \colon \Delta_n \to U \mid u(\Delta_n) \not\subseteq U \cap V\} \\ S'_n(V) &= \{u \colon \Delta_n \to V \mid u(\Delta_n) \not\subseteq U \cap V\} \\ S'_n(X) &= \{u \colon \Delta_n \to X \mid u(\Delta_n) \not\subseteq U \text{ and } u(\Delta_n) \not\subseteq V\} \\ S_n(U,V) &= \{u \colon \Delta_n \to X \mid u(\Delta_n) \subseteq U \text{ or } u(\Delta_n) \subseteq V\}. \end{aligned}$$

We also define $C'_n(U) = \mathbb{Z}\{S'_n(U)\}$ and similarly for $C'_n(V)$, $C'_n(X)$ and $C_n(U,V)$.

Remark 19.2. If $u \in S'_n(U)$ then u does not send the whole of Δ_n into $U \cap V$, but it may send some faces of Δ_n into $U \cap V$. Because of this, $C'_*(U)$ need not be closed under ∂, so it need not be a subcomplex of $C_*(U)$. However, this will not matter for our immediate purposes.

We now note that

$$\begin{aligned} S_n(U) &= S_n(U \cap V) \cup S'_n(U) \\ S_n(V) &= S_n(U \cap V) \cup S'_n(V) \\ S_n(U,V) &= S_n(U \cap V) \cup S'_n(U) \cup S'_n(V) \\ S_n(U \cup V) &= S_n(U \cap V) \cup S'_n(U) \cup S'_n(V) \cup S'_n(X), \end{aligned}$$

and all these unions involve disjoint sets. It follows that

$$\begin{aligned} C_n(U) &= C_n(U \cap V) \oplus C'_n(U) \\ C_n(V) &= C_n(U \cap V) \oplus C'_n(V) \\ C_n(U,V) &= C_n(U \cap V) \oplus C'_n(U) \oplus C'_n(V) \\ C_n(U \cup V) &= C_n(U \cap V) \oplus C'_n(U) \oplus C'_n(V) \oplus C'_n(X), \end{aligned}$$

Thus, in our earlier sequence, the map $\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right] : C_n(U \cap V) \to C_n(U) \oplus C_n(V)$ becomes the map

$$C_n(U \cap V) \to C_n(U \cap V) \oplus C'_n(U) \oplus C_n(U \cap V) \oplus C'_n(V)$$

given by $a \mapsto (a, 0, -a, 0)$. Similarly, the map $\left[\begin{smallmatrix} k_* & l_* \end{smallmatrix}\right] : C_n(U) \oplus C_n(V) \to C_n(U \cup V)$ becomes the map

$$C_n(U \cap V) \oplus C'_n(U) \oplus C_n(U \cap V) \oplus C'_n(V) \to C_n(U \cap V) \oplus C'_n(U) \oplus C'_n(V) \oplus C'_n(X)$$

given by $(a, b, a', c) \mapsto (a + a', b, c, 0)$. From this it is clear that we have a short exact sequence

$$C_*(U \cap V) \xrightarrow{\left[\begin{smallmatrix} i_* \\ -j_* \end{smallmatrix}\right]} C_*(U) \oplus C_*(V) \xrightarrow{\left[\begin{smallmatrix} k_* & l_* \end{smallmatrix}\right]} C_*(U, V),$$

and an inclusion $C_*(U, V) \to C_*(X)$ of chain complexes. By applying Theorem 17.2 to the short exact sequence, we get something which is essentially the Mayer-Vietoris sequence except that it involves $H_*(C_*(U, V))$ instead of $H_*(U \cup V) = H_*(X)$. To complete the construction of the Mayer-Vietoris sequence, we need to show that $H_*(C_*(U, V))$ is actually the same as $H_*(X)$.

Lemma 19.3. *For any $u \in C_n(X)$ there exists $k \geq 0$ such that $\mathrm{sd}^k(u) \in C_n(U, V)$.*

Proof. We can easily reduce to the case where u is a single map $\Delta_n \to X$. Put $U' = u^{-1}(U)$ and $V' = u^{-1}(V)$, so U' and V' are open in Δ_n with $U' \cup V' = \Delta_n$. This means that $\{U', V'\}$ is an open cover of Δ_n, so Proposition 8.31 tells us that there is a Lebesgue number, say $\epsilon > 0$. The identity chain $\iota_n \in C_n(\Delta_n)$ has diameter $\sqrt{2}$, so $\mathrm{sd}^k(\iota_n)$ has diameter at most $(n/(n+1))^k\sqrt{2}$. If we choose k large enough, then this diameter will be less than ϵ, and it will follow that every simplex involved in $\mathrm{sd}^k(\iota_n)$ is either contained in U' or contained in V'. It follows that every simplex involved in the chain $\mathrm{sd}^k(u) = u_*(\mathrm{sd}^k(\iota_n))$ is either contained in U or contained in V, so $\mathrm{sd}^k(u) \in C_n(U, V)$ as claimed. □

Corollary 19.4. *The homology of the quotient complex $Q_* = C_*(X)/C_*(U, V)$ is zero.*

Proof. First, note that the subdivision map $\mathrm{sd}\colon C_*(X) \to C_*(X)$ sends $C_*(U)$ to $C_*(U)$ and $C_*(V)$ to $C_*(V)$ so it also sends the subcomplex $C_*(U, V) = C_*(U) + C_*(V)$ to itself. We therefore have an induced map $\mathrm{sd}\colon Q_* \to Q_*$ given by $\mathrm{sd}(u + C_n(U, V)) = \mathrm{sd}(u) + C_n(U, V)$. Similarly, the chain homotopy $\sigma\colon C_n(X) \to C_{n+1}(X)$ induces a chain homotopy $\sigma_n\colon Q_n \to Q_{n+1}$. We showed previously that $\partial\sigma + \sigma\partial = \mathrm{id} - \mathrm{sd}$ on $C_*(X)$, and it follows that we have the same relation on Q_*. We therefore deduce from Proposition 14.7 that the map $\mathrm{sd}_*\colon H_n(Q_*) \to H_n(Q_*)$ is the identity. Consider an element $q \in H_n(Q_*)$. This has the form $q = z + B_n(Q_*)$ for some $z \in Z_n(Q_*) \leq Q_n = C_n(X)/C_n(U, V)$. This in turn has the form $z = u + C_n(U, V)$ for some $u \in C_n(X)$. For sufficiently large k we have $\mathrm{sd}^k(u) \in C_n(U, V)$, so $\mathrm{sd}^k(z) = 0$, so $\mathrm{sd}^k_*(q) = 0$. As sd_* is the identity this means that $q = 0$. Thus, we have $H_n(Q_*) = 0$ as claimed. □

Corollary 19.5. *The inclusion $C_*(U, V) \to C_*(X)$ induces an isomorphism $H_*(C_*(U, V)) \to H_*(X)$.*

Proof. We have a short exact sequence of chain complexes $C_*(U, V) \to C_*(X) \to Q_*$. Theorem 17.2 therefore gives us exact sequences

$$H_{n+1}(Q_*) \to H_n(C_*(U, V)) \to H_n(C_*(X)) \to H_n(Q_*).$$

The first and last groups are zero by Corollary 19.5, so exactness forces the middle map to be an isomorphism. □

This completes the construction of the Mayer-Vietoris sequence.

20. Further calculations

Video (Lemma 20.1 and Proposition 20.2)

Suppose that $n \geq 2$. We previously noted that $\mathbb{R}^n \setminus \{0\}$ is homotopy equivalent to S^{n-1} and so has $H_0 \simeq H_{n-1} \simeq \mathbb{Z}$, with all other homology groups zero. One might ask what happens if we remove several points $a_1, \dots, a_m$ from $\mathbb{R}^n$ instead of just removing the origin. One can answer this question by induction on m, but to make the induction work smoothly, it is convenient to generalise slightly. Instead of just considering $\mathbb{R}^n \setminus \{a_1, \dots, a_m\}$, we will consider $W \setminus \{a_1, \dots, a_m\}$, where W is an arbitrary convex open subset of $\mathbb{R}^n$. We first need the following lemma:

Lemma 20.1. *Suppose that $X = U \cup V$, where U and V are open and connected, and $U \cap V$ is contractible. Then $H_0(X) = \mathbb{Z}$ and $H_k(X) = H_k(U) \oplus H_k(V)$ for all $k \geq 1$.*

Proof. This follows from Proposition 15.2 (the truncated Mayer-Vietoris sequence). That result includes the fact that $H_0(X) = \mathbb{Z}$, and that there are exact sequences

$$H_k(U \cap V) \to H_k(U) \oplus H_k(V) \to H_k(X) \to H_{k-1}(U \cap V)$$

for all $k \geq 2$, and also an exact sequence

$$H_1(U \cap V) \to H_1(U) \oplus H_1(V) \to H_1(X) \to 0.$$

As $H_j(U \cap V) = 0$ for all $j > 0$, these exact sequences show that the map $H_k(U) \oplus H_k(V) \to H_k(X)$ is an isomorphism for all $k \geq 1$. $\square$

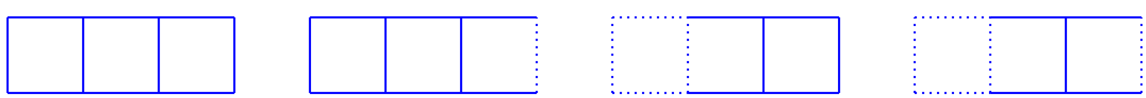
Proposition 20.2. *Let W be a convex open subset of $\mathbb{R}^n$ (with $n \geq 2$). Let A be a finite subset of W, with $|A| = m$ say, and put $X = W \setminus A$. Then $H_0(X) = \mathbb{Z}$ and $H_{n-1}(X) = \mathbb{Z}^m$ and $H_k(X) = 0$ for $k \notin \{0, n-1\}$.*

Proof. If $m = 0$ then $X = W$ so X is contractible so the claim is clear. Suppose instead that $m = 1$ so $A = \{a\}$ say. Define $r \colon X \to S^{n-1}$ by $r(x) = (x - a)/\|x - a\|$. Next, as W is open we can choose $\epsilon > 0$ such that the open ball of radius ϵ around a is contained in W. We can then define $j \colon S^{n-1} \to X$ by $j(y) = a + \epsilon y/2$. We find that rj is the identity, and that jr is homotopic to the identity by a straight-line homotopy, so X is homotopy equivalent to S^{n-1}; the claim follows from this.

We now suppose that $m > 1$ and that we have already proved the claim for sets of size less than m. We have only a finite number of vectors $a - a'$ with $a, a' \in A$ and $a \neq a'$; choose any unit vector u that is not perpendicular to any of these. We then find that the dot products $u.a$ (for $a \in A$) are all different, so we can list the elements of A as $a_1, \dots, a_m$ with $u.a_1 < \cdots < u.a_m$. Choose constants p, q with $u.a_{m-1} < p < q < u.a_m$. Put $V = \{x \in X \mid u.x < q\}$ and $U = \{x \in X \mid u.x > p\}$, so U and V are open with $U \cup V = X$. The space V is obtained by removing $\{a_1, \dots, a_{m-1}\}$ from the convex set $\{w \in W \mid u.w < q\}$, so the homology of V is given by the induction hypothesis. The space U is obtained by removing a_m from the set $\{w \in W \mid u.w > p\}$, so the homology of U is given by the case $m = 1$. The intersection $U \cap V$ is just the convex set $\{w \in W \mid p < u.w < q\}$ (with no points removed), so it is contractible. Lemma 20.1 therefore gives $H_0(X) = \mathbb{Z}$ and $H_k(X) = H_k(U) \oplus H_k(V)$ for all $k > 0$, and the induction step is clear from this. $\square$

Video (Examples 20.3 and 20.4)

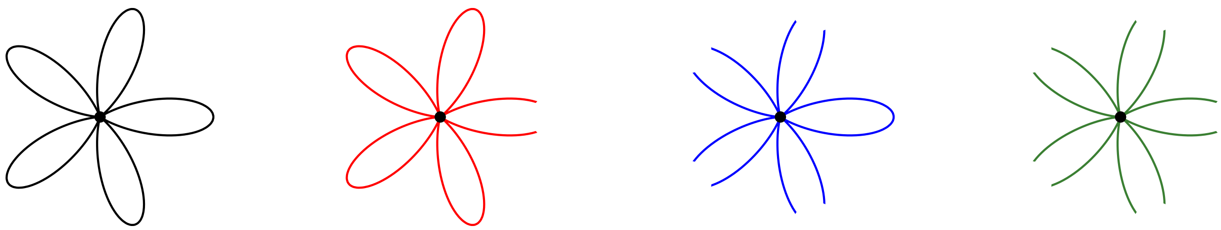
Example 20.3. Let L_n be the union of n adjacent squares arranged horizontally. The case $n = 3$ is illustrated below.



L_3 $\qquad$ $U \cong L_2$ $\qquad$ $V \cong S^1$ $\qquad$ $U \cap V \cong 1$

Let U be the space obtained by removing the rightmost vertical edge; this is easily seen to be homotopy equivalent to L_{n-1}. Let V be the space obtained by removing L_{n-2}, leaving just the rightmost square with two extra edges attached; this is easily seen to be homotopy equivalent to S^1. The intersection $U \cap V$ is a sideways H shape, and is easily seen to be contractible. We can therefore apply Lemma 20.1 to see that $H_k(L_n) = H_k(L_{n-1}) \oplus H_k(S^1)$ for all $k > 0$. It follows inductively that $H_0(L_n) = \mathbb{Z}$ and $H_1(L_n) = \mathbb{Z}^n$ and $H_k(L_n) = 0$ for $k > 1$. Alternatively, we can let a_p be the centre of the p'th square, so $L_n \subseteq \mathbb{R}^2 \setminus \{a_1, \dots, a_n\}$. It can be shown that the inclusion $i \colon L_n \to \mathbb{R}^2 \setminus \{a_1, \dots, a_n\}$ is a homotopy equivalence. The homology of L_n can therefore be obtained from Proposition 20.2; it is easy to see that this gives the same answer.

Example 20.4. Let W_n consist of n circles joined together at a single point. We can write this as $U \cup V$ as illustrated below. Given this, essentially the same argument as in Example 20.3 gives $H_0(W_n) = \mathbb{Z}$ and $H_1(W_n) = \mathbb{Z}^n$ and $H_k(W_n) = 0$ for $k \geq 2$.

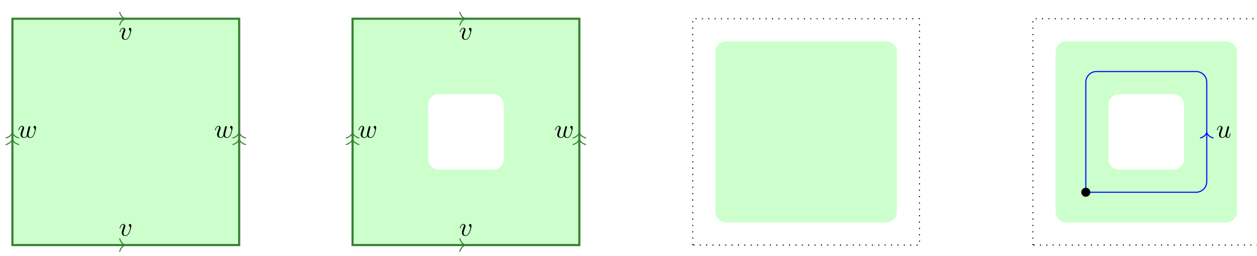


$W_5 = U \cup V$ $\qquad$ $U \cong W_4$ $\qquad$ $V \cong S^1$ $\qquad$ $U \cap V \cong 1$

Video (Proposition 20.5 and Remark 20.6)

Proposition 20.5. *The homology of the torus $T = S^1 \times S^1$ is given by $H_0(T) = H_2(T) = \mathbb{Z}$ and $H_1(T) = \mathbb{Z}^2$ and $H_k(T) = 0$ for $k \geq 3$.*

Proof. We can form the torus by taking a square, gluing the top to the bottom and the left edge to the right edge. This is illustrated by the left hand picture below. The other pictures show two open sets $U, V \subseteq T$ with $U \cup V = T$.



$T = U \cup V$ $\qquad$ U $\qquad$ V $\qquad$ $U \cap V$

Interactive demo

The set V is contractible, so $H_0(V) = \mathbb{Z}$ and $H_k(V) = 0$ for $k > 0$. The set $U \cap V$ is homotopy equivalent to S^1, so $H_0(U \cap V) = \mathbb{Z}$ and $H_1(U \cap V) = \mathbb{Z}.[u]$ and $H_k(U \cap V) = 0$ for $k > 1$. It might appear that U is also homotopy equivalent to S^1, but that is misleading because the edges are glued together. The linked demonstration makes it clear that U is the union of two circular bands, whose intersection is a filled square (and so is contractible). It therefore follows from Lemma 20.1 that $H_0(U) = \mathbb{Z}$ and $H_1(U) = \mathbb{Z} \oplus \mathbb{Z}$ and $H_k(U) = 0$ for $k > 1$. More specifically, the two bands are thickenings of the loops v and w, so $H_1(U) = \mathbb{Z}\{[v], [w]\}$. All the relevant spaces are connected, so $H_0(T) = \mathbb{Z}$ and we have a truncated Mayer-Vietoris sequence as in Proposition 15.2. After filling in the known groups, we see that the tail end of this sequence is as follows:

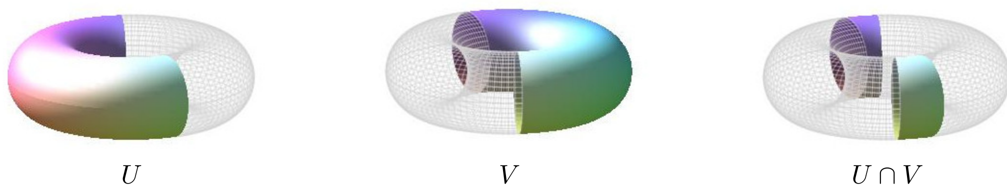
$$0 \to H_2(T) \xrightarrow{\delta} \mathbb{Z}.[u] \xrightarrow{i_*} \mathbb{Z}\{[v], [w]\} \xrightarrow{k_*} H_1(T) \to 0.$$

In U, we can deform u outwards to the edge of the square (which does not change the homology class). It then becomes equal to the join $v * w * \overline{v} * \overline{w}$. As discussed in Proposition 10.29, this is homologous to $v + w - v - w = 0$. This proves that $i_* = 0$, so $\ker(i_*) = \mathbb{Z}.[u]$ and $\mathrm{img}(i_*) = 0$. As the sequence is exact, it follows that $\mathrm{img}(\delta) = \mathbb{Z}.[u]$ and $\ker(k_*) = 0$, so the maps δ and k_* are isomorphisms, so $H_1(T) \simeq \mathbb{Z}^2$ and $H_2(T) \simeq \mathbb{Z}$. For $k \geq 3$ we have an exact sequence

$$0 = H_k(U) \oplus H_k(V) \to H_k(X) \xrightarrow{\delta} H_{k-1}(U \cap V) = 0,$$

which shows that $H_k(T) = 0$. $\square$

Remark 20.6. As an alternative, we could write the torus as the union of the following open sets:

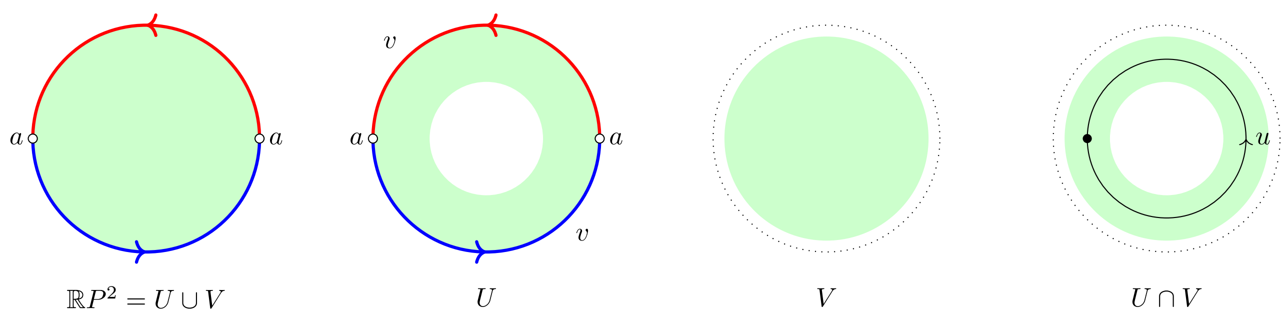


Here U and V are both homotopy equivalent to S^1, whereas $U \cap V$ is homotopy equivalent to the union of two disjoint copies of S^1. We leave it to the reader to understand how the resulting Mayer-Vietoris sequence gives the same answer as before. This method can be generalised to calculate the homology of the d-dimensional torus $T_d = (S^1)^d = S^1 \times \cdots \times S^1$. The answer is that $H_k(T_d) \simeq \mathbb{Z}^{\binom{d}{k}}$ for all k, but we will not give the details here.

Proposition 20.7. *For the real projective plane $\mathbb{R}P^2$ we have $H_0(\mathbb{R}P^2) = \mathbb{Z}$ and $H_1(\mathbb{R}P^2) = \mathbb{Z}/2$ and $H_k(\mathbb{R}P^2) = 0$ for $k \geq 2$.*

Proof. Video

Recall (perhaps from the Knots and Surfaces course) that we can form $\mathbb{R}P^2$ by taking a disc as shown on the left below, and identifying opposite points on the boundary circle. In particular, the two points marked a are the same, and each point on the upper red semicircle is identified with the corresponding point on the lower blue semicircle, so the two paths marked v are the same. We consider $\mathbb{R}P^2$ as the union of the indicated open sets U and V.

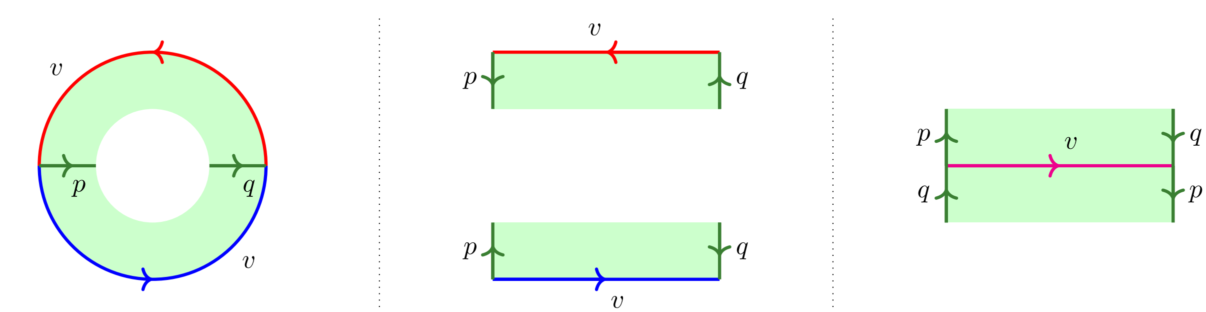


It is clear that V is contractible, so $H_0(V) = \mathbb{Z}$ and $H_k(V) = 0$ for $k \geq 1$. It is also clear that that $U \cap V$ is homotopy equivalent to a circle, so $H_0(U \cap V) = \mathbb{Z}$ and $H_1(U \cap V) = \mathbb{Z}.[u]$ and $H_k(U \cap V) = 0$ for $k \geq 2$.

Next, we claim that U is homeomorphic to a Möbius strip, and therefore homotopy equivalent to a circle. One way to see this is to use the following animated diagram:

Interactive demo

Alternatively, we can argue using the pictures below. The first one shows U together with two extra edges marked p and q. If we cut along these, we get the middle picture. If we flip the top half over and glue it to the bottom half along v, we get the third picture, which is the standard gluing diagram for a Möbius strip.



Either way, it follows that $H_0(U) = \mathbb{Z}$ and $H_1(U) = \mathbb{Z}.[v]$ and $H_k(U) = 0$ for $k \geq 2$. We next need to understand the map $i_*\colon H_1(U \cap V) \to H_1(U)$. In U, we can push the loop u out to the boundary by a loop homotopy, which does not change the homology class. The deformed loop then covers both copies of v, so we have $i_*([u]) = 2[v]$. After taking account of the known homology groups, the end of the truncated Mayer-Vietoris sequence looks like this:

$$\begin{tikzcd} 0 \arrow[r] & 0 \arrow[r] & H_2(\mathbb{R}P^2) \arrow[dll, rounded corners, "\delta"] \\ \mathbb{Z}.[u] \arrow[r, "i_*"] & \mathbb{Z}.[v] \arrow[r, "k_*"] & H_1(\mathbb{R}P^2) \arrow[r] & 0 \end{tikzcd}$$

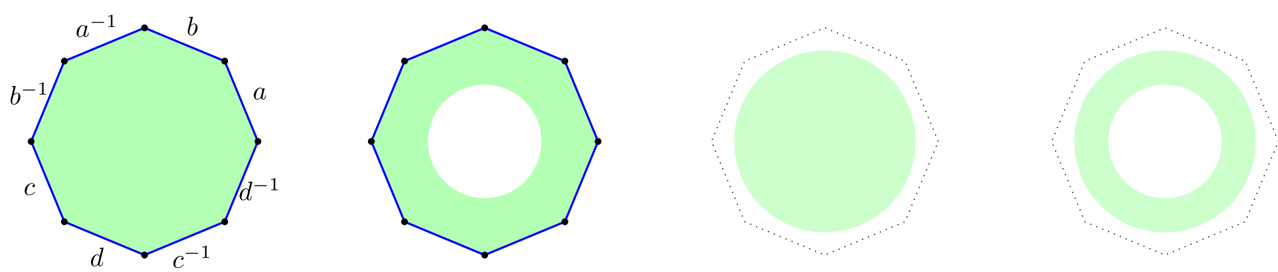
As $i_*([u]) = 2[v]$ we see that $\ker(i_*) = 0$ and so exactness forces $H_2(\mathbb{R}P^2)$ to be zero. Also, the map k_* is surjective with kernel given by the image of i_*, which is $2\mathbb{Z}.[v]$, so $H_1(\mathbb{R}P^2) \simeq \mathbb{Z}/2$. For $k \geq 3$ we also have exact sequences

$$0 = H_k(U) \oplus H_k(V) \to H_k(\mathbb{R}P^2) \to H_{k-1}(U \cap V) = 0,$$

so $H_k(\mathbb{R}P^2) = 0$. As $\mathbb{R}P^2$ is connected, we have $H_0(\mathbb{R}P^2) = \mathbb{Z}$. □

Example 20.8. Video

Now consider a closed surface X presented as in the Knots and Surfaces course, by gluing edges of a polygon according to a surface word. We illustrate this using the standard word $aba^{-1}b^{-1}cdc^{-1}d^{-1}$ for an orientable surface of genus 2, but the method works for any word satisfying the usual conditions of surface theory. We can calculate $H_*(X)$ using essentially the same method that we used for the torus. The conclusion is that $H_0(X) = H_2(X) = \mathbb{Z}$, and $H_k(X) = 0$ for $k \geq 3$, and $H_1(X) = \mathbb{Z}^{2g}$, where g is the genus. In particular, in the illustrated case we have $H_1(X) = \mathbb{Z}^4$. To see this, we use open sets U and V illustrated below.



$X = U \cup V$ $\qquad$ U $\qquad$ V $\qquad$ $U \cap V$

As with the case of the torus, the space V is contractible and the space $U \cap V$ is homotopy equivalent to a circle, so their homology is easy to understand. The space U apparently has 8 edges and 8 marked vertices. However, the edges are glued together in pairs, in such a way that all vertices get glued together; the result is just the space W_4 from Example 20.4. The full space U consists of W_4 with a fringe attached, but that does not affect the homotopy type, so we have $H_*(U) = H_*(W_4)$. This means that $H_0(U) = \mathbb{Z}$ and $H_1(U) = \mathbb{Z}^4$ and $H_k(U) = 0$ for $k \geq 2$. If we let u be a loop once around $U \cap V$, then in U we see that u becomes $a * b * \overline{a} * \overline{b} * c * d * \overline{c} * \overline{d}$. This has homology class $[a] + [b] - [a] - [b] + [c] + [d] - [c] - [d] = 0$, so the homomorphism $i_*\colon H_1(U \cap V) \to H_1(U)$ is zero. We now have an exact sequence

$$0 \to H_2(X) \xrightarrow{\delta} \mathbb{Z} \xrightarrow{i_*=0} \mathbb{Z}^4 \to H_1(X) \to 0,$$

and the claimed description of $H_*(X)$ follows easily from this.

Proposition 20.9. *The group $H_1(\mathbb{R}P^1)$ is isomorphic to $\mathbb{Z}$, but for $n \geq 2$ the group $H_1(\mathbb{R}P^n)$ is isomorphic to $\mathbb{Z}/2$. Moreover, for $k > n$ we have $H_k(\mathbb{R}P^n) = 0$.*

Proof. Video

We saw in Example 7.23 that $\mathbb{R}P^1$ is homeomorphic to S^1, so $H_1(\mathbb{R}P^1) \simeq \mathbb{Z}$. The case $n = 2$ is given by Proposition 20.7. We have an inclusion $i\colon S^n \to S^{n+1}$ given by $i(x_0, \dots, x_n) = (x_0, \dots, x_n, 0)$. This satisfies $i(-x) = -i(x)$, so it induces an inclusion $S^n/\{\pm 1\} \to S^{n+1}/\{\pm 1\}$, or in other words $\mathbb{R}P^n \to \mathbb{R}P^{n+1}$; we again call this i. It will be enough to check that the map $i_*\colon H_1(\mathbb{R}P^n) \to H_1(\mathbb{R}P^{n+1})$ is an isomorphism for $n \geq 2$.

Recall that $S^{n+1} = \{x \in \mathbb{R}^{n+2} \mid \|x\|^2 = 1\}$. We can identify $\mathbb{R}^{n+2}$ with $\mathbb{R}^{n+1} \times \mathbb{R}$; then

$$S^{n+1} = \{(y, z) \in \mathbb{R}^{n+1} \times \mathbb{R} \mid \|y\|^2 + z^2 = 1\}.$$

The subspace where $z = 0$ can be identified with S^n. Now let U_+ and U_- be the subspaces where $z > 0$ and $z < 0$ respectively, and put $\widetilde{U} = U_+ \cup U_-$. We also put

$$\widetilde{V} = \{(y, z) \in S^{n+1} \mid y \neq 0\} = \{(y, z) \in S^{n+1} \mid -1 < z < 1\} = S^{n+2} \setminus \{(0, 1), (0, -1)\}.$$

Now let $\pi\colon S^{n+1} \to S^{n+1}/\{\pm 1\} = \mathbb{R}P^{n+1}$ be the quotient projection (so $\pi(x) = \pi(x')$ iff $x' = \pm x$). Put $U = \pi(\widetilde{U})$ and $V = \pi(\widetilde{V})$. These sets are easily seen to be open with respect to the quotient topology on $\mathbb{R}P^{n+1}$.

Let U' be the open ball of radius one in $\mathbb{R}^{n+1}$. We have maps

$$U' \xrightarrow{f} U_+ \xrightarrow{\pi} U,$$

where $f(y) = (y, \sqrt{1 - \|y\|^2})$, and one can check that both of these are homeomorphisms. It follows that U is contractible. The same maps also give homeomorphisms

$$U' \setminus \{0\} \xrightarrow{f} U_+ \setminus \{(0, 1)\} \xrightarrow{\pi} U \cap V.$$

It follows that $U \cap V$ is homotopy equivalent to S^n.

Next, let $j\colon S^n \to \widetilde{V}$ be the inclusion, and define $r\colon \widetilde{V} \to S^n$ by $r(y, z) = y/\|y\|$ (which is valid because $y \neq 0$ when $(y, z) \in \widetilde{V}$). This satisfies $r \circ j = \mathrm{id}\colon S^n \to S^n$. We also define $h\colon [0, 1] \times \widetilde{V} \to \widetilde{V}$ by

$$h(t, (y, z)) = (y, tz)/\sqrt{\|y\|^2 + t^2 z^2};$$

this gives a homotopy between $j \circ r$ and the identity. These maps satisfy $j(-x) = -j(x)$ and $r(-x) = -r(x)$ and $h(t, -x) = -h(t, x)$ so they induce maps $j\colon \mathbb{R}P^n \to V$ and $r\colon V \to \mathbb{R}P^n$ and $h\colon [0, 1] \times V \to V$. Using this we see that j is a homotopy equivalence, so $H_1(V) = H_1(\mathbb{R}P^n)$.

The spaces U, V, $U \cap V$ and $U \cup V = \mathbb{R}P^{n+1}$ are all path connected, so we have a truncated Mayer-Vietoris sequence

$$H_1(U \cap V) \to H_1(U) \oplus H_1(V) \to H_1(\mathbb{R}P^{n+1}) \to 0.$$

Here $U \cap V$ is homotopy equivalent to S^n with $n \geq 2$ so $H_1(U \cap V) = 0$. The space U is contractible so $H_1(U) = 0$. The space V is homotopy equivalent to $\mathbb{R}P^n$, so we can assume inductively that $H_1(V) = \mathbb{Z}/2$. Exactness of the sequence now implies that $H_1(\mathbb{R}P^{n+1}) = \mathbb{Z}/2$ as required.

Now suppose that $k > n + 1$. We have a Mayer-Vietoris sequence

$$H_k(U) \oplus H_k(V) \to H_k(\mathbb{R}P^{n+1}) \to H_{k-1}(U \cap V),$$

or equivalently

$$H_k(\mathbb{R}P^n) \to H_k(\mathbb{R}P^{n+1}) \to H_{k-1}(S^n).$$

As $k > n + 1$ we have $H_{k-1}(S^n) = 0$. Also, we can assume inductively that $H_j(\mathbb{R}P^n) = 0$ for all $j > n$, so in particular $H_k(\mathbb{R}P^n) = 0$. It follows by exactness that $H_k(\mathbb{R}P^{n+1}) = 0$ as claimed. □

Remark 20.10. The full story is as follows. For $m \geq 0$ we have

$$H_k(\mathbb{R}P^{2m+1}) = \begin{cases} \mathbb{Z} & \text{if } k = 0 \text{ or } k = 2m+1 \\ \mathbb{Z}/2 & \text{if } 0 < k < 2m \text{ and } k \text{ is odd} \\ 0 & \text{otherwise.} \end{cases}$$

$$H_k(\mathbb{R}P^{2m+2}) = \begin{cases} \mathbb{Z} & \text{if } k = 0 \\ \mathbb{Z}/2 & \text{if } 0 < k < 2m+2 \text{ and } k \text{ is odd} \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently, let $P(n)$ be the following chain complex:

$$\cdots \leftarrow 0 \leftarrow 0 \leftarrow \mathbb{Z} \xleftarrow{0} \mathbb{Z} \xleftarrow{2} \mathbb{Z} \xleftarrow{0} \mathbb{Z} \xleftarrow{2} \cdots \leftarrow \mathbb{Z} \leftarrow 0 \leftarrow 0 \leftarrow \cdots .$$

There are copies of $\mathbb{Z}$ in degrees 0 to n inclusive, and the differentials alternate between zero and multiplication by 2. Then $H_*(\mathbb{R}P^n) \simeq H_*(P(n))$ for all n. This can be proved using the same Mayer-Vietoris sequence as mentioned above, but some extra work is needed to determine the maps in that sequence.

21. The Jordan Curve Theorem

Video (Statement of Theorem 21.1 and Remark 21.2)

The main aim of this section will be to prove the following theorem.

Theorem 21.1. *Suppose that $X \subseteq S^n$ and that X is homeomorphic to S^k for some $k \leq n$.*

(a) *If $k = n$ then X is just equal to S^n and so $S^n \setminus X$ is empty and $H_*(S^n \setminus X) = 0$.*
(b) *If $k = n-1$ then $H_0(S^n \setminus X) \simeq \mathbb{Z}^2$ and $H_m(S^n \setminus X) = 0$ for all $m > 0$.*
(c) *If $k < n-1$ then $H_0(S^n \setminus X) \simeq H_{n-k-1}(S^n \setminus X) \simeq \mathbb{Z}$ and all other homology groups are trivial.*

Remark 21.2. We can combine cases (b) and (c) in Theorem 21.1 by saying that $H_*(S^n \setminus X) \simeq H_*(S^{n-1-k})$. For the most obvious case of the theorem, we can express $\mathbb{R}^{n+1}$ as $\mathbb{R}^{k+1} \oplus \mathbb{R}^{n-k}$, so

$$S^n = \{(y, z) \mid \|y\|^2 + \|z\|^2 = 1\}.$$

The space $X = \{(y, 0) \mid \|y\| = 1\} \subseteq S^n$ is then homeomorphic to S^k, with

$$S^n \setminus X = \{(y, z) \mid \|y\|^2 + \|z^2\| = 1, \ z \neq 0\}.$$

We can define maps $S^{n-k-1} \xrightarrow{i} S^n \setminus X \xrightarrow{r} S^{n-k-1}$ by $i(z) = (0, z)$ and $r(y, z) = z/\|z\|$. We then have $r \circ i = \text{id}$, and we have a homotopy between $i \circ r$ and the identity given by $h(t, y, z) = (ty, z)/\|(ty, z)\|$. Thus, in this case $S^n \setminus X$ is actually homotopy equivalent to S^{n-k-1}. However, that is not true in general. For example, let X be a knotted circle in $\mathbb{R}^3$. We can identify $\mathbb{R}^3$ with $S^3 \setminus \{\text{point}\}$ by stereographic projection, and thus think of X as a subspace of S^3. The theorem tells us that in this case $S^3 \setminus X$ has the same homology as S^1, but it can be shown that $S^3 \setminus X$ is not homotopy equivalent to S^1.

We will also deduce the following result, for which the case $n = 2$ is called the Jordan Curve Theorem:

Theorem 21.3. *Suppose that $n \geq 2$, and let $f \colon S^{n-1} \to \mathbb{R}^n$ be an injective continuous map. Then the complementary set $\mathbb{R}^n \setminus f(S^{n-1})$ has precisely two path components, one bounded and the other unbounded.*

It is easy to see that the Jordan Curve Theorem is true for maps $u \colon S^1 \to \mathbb{R}^2$ that are reasonably simple. However, it is hard to prove in the general case, where u may wiggle in an extremely complicated way and can also be fractal or otherwise badly behaved.

We will work up to Theorem 21.1 by first considering $S^n \setminus X$ in cases where X is homeomorphic to B^k (or equivalently, $[0, 1]^k$) rather than S^k.

Video (Definition 21.4 to Corollary 21.13)

Definition 21.4. We say that a space X is *acyclic* if $H_0(X) = \mathbb{Z}$ and $H_i(X) = 0$ for all $i > 0$.

Remark 21.5. We have seen previously that all contractible spaces are acyclic. Conversely, most commonly occurring acyclic spaces are contractible, but there are some exceptions.

Lemma 21.6. *Let X be a space. Let $1 = \{0\}$ denote the one-point space, so we have a constant map $p\colon X \to 1$. Then X is acyclic iff the map $p_*\colon H_*(X) \to H_*(1)$ is an isomorphism.*

Proof. If p_* is an isomorphism then $H_*(X)$ is isomorphic to $H_*(1)$ i.e. $H_0(X) \simeq \mathbb{Z}$ and $H_i(X) \simeq 0$ for $i > 0$, which means that X is acyclic. Suppose instead we start withthe assumption that X is acyclic. For $i > 0$ the groups $H_i(X)$ and $H_i(1)$ are both zero, so the map $p_*\colon H_i(X) \to H_i(1)$ is automatically an isomorphism. Next, we know that $H_0(X) \simeq \mathbb{Z}\{\pi_0(X)\}$. As X is acyclic we have $H_0(X) \simeq \mathbb{Z}$ so $|\pi_0(X)| = 1$ so X is nonempty and path connected. We can therefore choose $a \in X$ and we have $H_0(X) = \mathbb{Z}.[a]$. We also have $p(a) = 0$ and $H_0(1) = \mathbb{Z}.[0]$ so $p_*\colon H_0(X) \to H_0(1)$ is an isomorphism. □

Theorem 21.7. *Let X be a subset of S^n that is homeomorphic to $[0,1]^k$ for some $k \leq n$. Then $S^n \setminus X$ is acyclic.*

Before proving the theorem we will prove a simpler lemma. This will not directly contribute to the theorem, but will introduce some relevant ideas.

Lemma 21.8. *Suppose that Y and Z are closed subsets of S^n such that the sets $S^n \setminus Y$, $S^n \setminus Z$ and $S^n \setminus (Y \cap Z)$ are all acyclic. Then the set $S^n \setminus (Y \cup Z)$ is also acyclic.*

Proof. Put $V = S^n \setminus Y$ and $W = S^n \setminus Z$. These are open subsets of S^n with $V \cup W = S^n \setminus (Y \cap Z)$ and $V \cap W = S^n \setminus (Y \cup Z)$. Thus, our assumptions are that V, W and $V \cup W$ are acyclic, and we need to prove that $V \cap W$ is acyclic. We have a Mayer-Vietoris sequence

$$H_{m+1}(V \cup W) \xrightarrow{\delta} H_m(V \cap W) \xrightarrow{\alpha} H_m(V) \oplus H_m(W) \to H_m(V \cup W)$$

If $m > 0$ then the first and third terms are zero so $H_m(V \cap W) = 0$ as expected. If $m = 0$ we instead have an exact sequence

$$0 \to H_0(V \cap W) \to \mathbb{Z} \oplus \mathbb{Z} \to \mathbb{Z} \to 0$$

and it is not hard to deduce that $H_0(V \cap W) \simeq \mathbb{Z}$. □

In Theorem 21.7, we assume that X is homeomorphic to $[0,1]^k$. This means that we can choose an injective continuous map $f\colon [0,1]^k \to S^n$ with $f([0,1]^k) = X$. We will make some constructions in that context.

Definition 21.9. Let $f\colon [0,1]^k \to S^n$ be an injective continuous map. By a *slice* we mean a set of the form $f([0,1]^{k-1} \times [a,b])$ with $0 \leq a \leq b \leq 1$. The *width* of such a slice is $b - a$. A *coslice* is the complement in S^n of a slice; we define the width of a coslice to be the same as the width of the corresponding slice.

Remark 21.10. We could attempt to prove Theorem 21.7 as follows. We argue by induction on k, the case $k = 0$ being easy. For $k > 0$, we divide X into a large number of very thin slices. Any slice of width 0 has acyclic complement, by the induction hypothesis. Our slices have very small width, so it seems reasonable to assume that they will also have acyclic complement. The intersection of two adjacent slices is a slice of width 0, and so has acyclic complement. Using Lemma 21.8, we deduce that the union of two adjacent slices again has acyclic complement. By repeating this procedure, we see that the union of any block of adjacent slices has acyclic complement. In particular, the full set X has acyclic complement, as required.

The problem with this approach is that our "reasonable assumption" is hard to justify directly. Our actual proof of Theorem 21.7 will use many of the same ideas, but arranged in a slightly different way.

Lemma 21.11. *Suppose that Y and Z are closed subsets of S^n such that the set $S^n \setminus (Y \cap Z)$ is acyclic. Note that we have inclusions*

$$S^n \setminus Y \xleftarrow{i} S^n \setminus (Y \cup Z) \xrightarrow{j} S^n \setminus Z.$$

Suppose that $u \in H_m(S^n \setminus (Y \cup Z))$ is nonzero; then either $i_*(u)$ is nonzero in $H_m(S^n \setminus Y)$, or $j_*(u)$ is nonzero in $H_m(S^n \setminus Z)$.

Proof. We use the same notation and the same Mayer-Vietoris sequence as in Lemma 21.8:

$$H_{m+1}(V \cup W) \xrightarrow{\delta} H_m(V \cap W) \xrightarrow{\alpha} H_m(V) \oplus H_m(W) \to H_m(V \cup W)$$

We again have $H_{m+1}(V \cup W) = 0$, so α is injective, so $\alpha(u) \neq 0$. However, $\alpha(u)$ is just $(i_*(u), -j_*(u))$, so either $i_*(u)$ or $j_*(u)$ must be nonzero. □

Lemma 21.12. *Let U be a topological space, and let $U_0, U_1, U_2, \dots$ be a sequence of open sets with $U_0 \subseteq U_1 \subseteq U_2 \subseteq \cdots$ and $U = \bigcup_{i=0}^{\infty} U_i$. Then any chain $w \in C_p(U)$ is contained in $C_p(U_j)$ for some j.*

Proof. We can express w as $m_1 w_1 + \cdots + m_r w_r$ for some integers m_i and continuous maps $w_i \colon \Delta_p \to U$. Note that the sets $w_i^{-1}(U_j)$ are open in Δ_p, and the union of all these sets is $w_i^{-1}(U) = \Delta_p$. As Δ_p is compact, it must be covered by a finite subcollection of the sets $w_i^{-1}(U_j)$. As these sets are nested inside each other, this just means that there exists j_i with $w_i^{-1}(U_{j_i}) = \Delta_p$. Equivalently, we have $w_i(\Delta_p) \subseteq U_{j_i}$ or $w_i \in C_p(U_{j_i})$. Thus, if we put $j = \max(j_1, \dots, j_r)$ then we have $w_i \in C_p(U_j)$ for all i, and so $w \in C_p(U_j)$. □

Corollary 21.13. *In the context of Lemma 21.12, suppose we have an element $u \in H_m(U_0)$ which maps to zero in $H_m(U)$. Then u already maps to zero in $H_m(U_j)$ for some j.*

Proof. Choose a cycle $z \in Z_m(U_0)$ such that $u = [z]$. As u maps to zero in $H_m(U)$, there must be a chain $w \in C_{m+1}(U)$ with $z = \partial(w)$ in $C_m(U)$. By the lemma, the element w lies in $C_m(U_j)$ for some j. We can therefore interpret the equation $\partial(w) = z$ as an equation in $C_m(U_j)$, showing that $[z] = 0$ in $H_m(U_j)$ as required. □

Proof of Theorem 21.7. Video

If $k = 0$ then X is just a single point, so U is homeomorphic to $\mathbb{R}^n$ by stereographic projection. This means that U is contractible and therefore acyclic. For $k > 0$ we will argue by induction. Choose a homeomorphism $f \colon [0,1]^k \to X$.

Suppose (for a contradiction) that we have a nonzero element $u \in H_m(S^n \setminus X)$ for some $m \geq 0$. If $m = 0$ we also assume that $p_*(u) = 0$ in $H_0(1) = \mathbb{Z}$. We will define a sequence of slices $X(j)$ of width 2^{-j} such that u has nonzero image in $H_m(S^n \setminus X(j))$ for all j.

We start with $X(0) = X$. Suppose we have already defined $X(j)$. We can write $X(j)$ as $X(j) = X(j)_+ \cup X(j)_-$, where $X(j)_+$ and $X(j)_-$ are slices of width 2^{-j-1}, and the set $X(j)_0 = X(j)_+ \cap X(j)_-$ has width 0 and so is homeomorphic to $[0,1]^{k-1}$. Our induction hypothesis says that $X(j)_0$ has acyclic complement, so Lemma 21.11 tells us that u must have nonzero image in $H_m(S^n \setminus X(j)_+)$ or in $H_m(S^n \setminus X(j)_-)$. We choose $X(j+1) = X(j)_+$ or $X(j+1) = X(j)_-$ as appropriate to ensure that u has nonzero image in $H_m(S^n \setminus X(j+1))$.

By construction we have $X(j) = f([0,1]^{k-1} \times [a_j, a_j + 2^{-j}])$ for some sequence (a_j) with $a_{j+1} \in \{a_j, a_j + 2^{-j-1}\}$. It follows that the numbers a_j converge to a limit a_∞, and that the set $X(\infty) = \bigcap_j X(j)$ is just $f([0,1]^{k-1} \times \{a_\infty\})$. This is homeomorphic to $[0,1]^{k-1}$ and so has acyclic complement by our induction hypothesis. In particular, the element u must map to zero in $H_m(S^n \setminus X(\infty))$. (In the case $m = 0$, we are using the assumption $p_*(u) = 0$ here.) However, we can regard $S^n \setminus X(\infty)$ as the union of the nested open sets $S^n \setminus X(j)$, so u must map to zero in $H_m(S^n \setminus X(j))$ for some j, which contradicts our construction of $X(j)$.

This contradiction shows that no element u as described above can exist. In other words, for $m > 0$ we have $H_m(S^n \setminus X) = 0$, and also the kernel of the map $p_* \colon H_0(S^n \setminus X) \to \mathbb{Z}$ is zero, so p_* is injective. On the other hand, X is contractible but S^n is not, so X cannot be equal to S^n, so we can choose a point $a \in S^n \setminus X$, and then $p_*[a] = 1$. This shows that p_* is also surjective, so it is an isomorphism as required. □

Proof of Theorem 21.1. Video

Suppose that $X \subseteq S^n$ is homeomorphic to S^k for some $k \leq n$. If $k = 0$ then this just means that X consists of two points. If $n = 0$ this clearly means that $X = S^0 = \{1, -1\}$ as claimed. If $n > 0$ then we recall that $S^n \setminus \{\text{point}\}$ is homeomorphic to $\mathbb{R}^n$, so removing two points gives $\mathbb{R}^n \setminus \{\text{point}\}$ which is homotopy equivalent to S^{n-1} and so has the same homology as S^{n-1}, as claimed.

We now suppose that $k>0$, and argue by induction on k. We can write S^k as the union of two hemispheres, with intersection S^{k-1}. Correspondingly, we can write X as $Y\cup Z$, where Y and Z are homeomorphic to B^k (or $[0,1]^k$) and $Y\cap Z$ is homeomorphic to S^{k-1}. Theorem 21.7 tells us that the sets $V=S^n\setminus Y$ and $W=S^n\setminus Z$ are acyclic. The induction hypothesis tells us that the space $V\cup W=S^n\setminus(Y\cap Z)$ has the same homology as $S^{n-(k-1)-1}=S^{n-k}$. We need to show that the space $V\cap W=S^n\setminus X$ has the same homology as S^{n-1-k} (or that $V\cap W=\emptyset$ if $k=n$). For $m>0$ we note that $H_m(V)=H_m(W)=H_{m+1}(V)=H_{m+1}(W)=0$ so the Mayer-Vietoris sequence

$$H_{m+1}(V)\oplus H_{m+1}(W)\to H_{m+1}(V\cup W)\xrightarrow{\delta} H_m(V\cap W)\to H_m(V)\oplus H_m(W)$$

shows that $\delta\colon H_{m+1}(V\cup W)\to H_m(V\cap W)$ is an isomorphism. This means that $H_m(V\cap W)=0$ for all $m>0$ with $m\neq n-1-k$, but that if $n-1-k>0$ (or equivalently $k<n-1$) then $H_{n-1-k}(V\cap W)\simeq\mathbb{Z}$.

This just leaves a few exceptional cases to consider. First suppose that $k<n-1$, so $n-k>1$, so $H_1(V\cup W)=H_1(S^{n-k})=0$. We then have a Mayer-Vietoris sequence

$$H_1(V\cup W)\to H_0(V\cap W)\to H_0(V)\oplus H_0(W)=\mathbb{Z}^2\to H_0(V\cup W)=\mathbb{Z}\to 0,$$

and it follows easily that $H_0(V\cap W)\simeq\mathbb{Z}$ as required.

Now suppose instead that $k=n-1$. In this case we must show that $V\cap W$ has the same homology as S^0, or in other words that $H_0(V\cap W)\simeq\mathbb{Z}^2$ and $H_m(V\cap W)=0$ for $m>0$. The case $m>0$ is covered by our main discussion above. The induction hypothesis tells us that $V\cup W$ has the same homology as S^1, so the Mayer-Vietoris sequence

$$H_1(V)\oplus H_1(W)\to H_1(V\cup W)\to H_0(V\cap W)\to H_0(V)\oplus H_0(W)\to H_0(V\cup W)\to 0$$

becomes

$$0\to\mathbb{Z}\xrightarrow{\delta} H_0(V\cap W)\xrightarrow{\alpha}\mathbb{Z}^2\xrightarrow{\beta}\mathbb{Z}\to 0.$$

We can describe H_0 as the free abelian group generated by π_0. Using this we see that β is essentially the addition map $(n,m)\mapsto n+m$, with kernel generated by $(1,-1)$. We can choose an element $u\in H_0(V\cap W)$ with $\alpha(u)=(1,-1)$, then it is not hard to deduce that $\{\delta(1),u\}$ is a basis for $H_0(V\cap W)$. This means that $H_0(V\cap W)\simeq\mathbb{Z}^2$, or in other words that $V\cap W$ has precisely two path components.

Finally suppose that $k=n$. Here the induction hypothesis shows that $V\cup W$ has the same homology as S^0, so in particular it has two path components. We must show that $V\cap W=\emptyset$. The Mayer-Vietoris sequence

$$H_1(V\cup W)\to H_0(V\cap W)\to H_0(V)\oplus H_0(W)\to H_0(V\cup W)\to 0$$

becomes

$$0\to H_0(V\cap W)\xrightarrow{\alpha}\mathbb{Z}^2\xrightarrow{\beta}\mathbb{Z}^2\to 0.$$

Choose $a\in V$ and $b\in W$, so $H_0(V)=\mathbb{Z}.[a]$ and $H_0(W)=\mathbb{Z}.[b]$. The sequence shows that β is surjective, but that is only possible if a lies in one path component of $V\cup W$ and b lies in the other. That implies that β is actually an isomorphism, and then exactness shows that $H_0(V\cap W)=0$. However, we know that $H_0(V\cap W)$ is the free abelian group on $\pi_0(V\cap W)$, so $\pi_0(V\cap W)=\emptyset$, so $V\cap W=\emptyset$ as claimed. □

Lemma 21.14. Video

If U is an open subset of $\mathbb{R}^n$, then every path component of U is also an open subset of $\mathbb{R}^n$. Similarly, if U is an open subset of S^n, then every path component of U is also an open subset of S^n.

Proof. First let U be open in S^n, and suppose we have a point $a\in U$, with path component A say. Suppose that $b\in A$, so there is a path u from a to b in U. As U is open, we can find a radius $\epsilon>0$ such that $OB(b,\epsilon)\cap S^n\subseteq U$. By reducing ϵ if necessary, we can assume that $\epsilon<1$. For any $c\in OB(b,\epsilon)\cap S^n$ we have a linear path v from b to c in $\mathbb{R}^n$. As $\|b-c\|<\epsilon<1$ we see that this does not pass through 0, so we can define $w(t)=v(t)/\|v(t)\|$; this gives a path from b to c in S^n. This stays within $OB(b,\epsilon)$ so it stays within U. This means we have a path $u*w$ from a to c in U, so $c\in A$. This proves that $OB(b,\epsilon)\subseteq A$. As b was arbitrary, this proves that A is open as claimed.

The argument for $\mathbb{R}^n$ is similar but easier. □

Proof of Theorem 21.3. Video

Let $f\colon S^{n-1} \to \mathbb{R}^n$ be an injective continuous map, where $n \geq 2$. First note that $f(S^{n-1})$ is compact, so it is bounded and closed in $\mathbb{R}^n$. It will be harmless to multiply f by a small positive constant, so we can assume that $\|f(x)\| < 1$ for all $x \in S^{n-1}$.

We now recall a few more details about stereographic projection. We identify $\mathbb{R}^{n+1}$ with $\mathbb{R}^n \times \mathbb{R}$, so $S^n = \{(y,z) \mid \|y\|^2 + z^2 = 1\}$. We put $a = (0,1)$ and $A_+ = \{(y,z) \in S^n \mid z > 0\}$. We have a homeomorphism $g\colon \mathbb{R}^n \to S^n \setminus \{a\}$ given by $g(u) = (u, \|u\|^2 - 1)/(\|u\|^2 + 1)$. This also gives a homeomorphism $\mathbb{R}^n \setminus f(S^{n-1}) \to U \setminus \{a\}$, where $U = S^n \setminus g(f(S^{n-1}))$. Theorem 21.1 tells us that U has the same homology as S^0, so in particular $H_0(U) \simeq \mathbb{Z}^2$, so U has precisely two path components. Let A be the path component containing a, and let B be the other path component, so U is the disjoint union of A and B. For $m > 0$ we therefore have $0 = H_m(U) = H_m(A) \oplus H_m(B)$, so $H_m(A) = H_m(B) = 0$. This shows that both A and B are acyclic.

Because $\|f(x)\| < 1$ for all $x \in S^{n-1}$ we see that the last coordinate of $g(f(x))$ is always negative. Thus, the whole upper hemisphere A_+ of S^n is contained in $S^n \setminus g(f(S^n))$. Moreover, A_+ is clearly path connected and contains a so $A_+ \subseteq A$. It follows that B is contained in the lower hemisphere, and so the corresponding subset of $\mathbb{R}^n$ is bounded.

We next claim that the set $A' = A \setminus \{a\}$ is path connected. We will prove this using the Mayer-Vietoris sequence

$$H_1(A_+ \cup A') \to H_0(A_+ \cap A') \to H_0(A_+) \oplus H_0(A') \to H_0(A_+ \cup A') \to 0.$$

Here $A_+ \cup A'$ is just A, and A is acyclic. Thus, the first and fourth groups above are 0 and $\mathbb{Z}$. The space $A_+ \cap A'$ is the same as $A_+ \setminus \{a\}$, which is homeomorphic to $(0,1) \times S^{n-1}$ and so homotopy equivalent to S^{n-1}. In particular, as we are assuming that $n \geq 2$, we know that $A_+ \cap A'$ is connected. Thus, the second group in our sequence is $\mathbb{Z}$. It is also clear that A_+ is contractible and so $H_0(A_+) = \mathbb{Z}$. We therefore have an exact sequence

$$0 \to \mathbb{Z} \to \mathbb{Z} \oplus H_0(A') \to \mathbb{Z} \to 0.$$

This is only consistent if $H_0(A') \simeq \mathbb{Z}$, which means that A' is path connected.

We now see that $U \setminus \{a\}$ is the disjoint union of path connected sets A' and B, so these are the path components. It is clear that $g^{-1}(B)$ is bounded and $g^{-1}(A')$ is unbounded. □

Video (Proposition 21.15 and Corollary 21.16)

Proposition 21.15. *Let $f\colon B^n \to S^n$ be continuous and injective. Then $f(OB^n)$ is open in S^n.*

Proof. Put $U = S^n \setminus f(S^{n-1})$. As $f(S^{n-1})$ is compact, it is closed in S^n, so U is open. Theorem 21.1 tells us that $H_*(U) \simeq H_*(S^0)$, so U has two path components. Let V be the path component containing $f(0)$, and let W be the other one. The sets V and W are open in S^n by Lemma 21.14. Now put $V' = f(OB^n)$ and $W' = S^n \setminus f(B^n)$. Using the injectivity of f we see that V' and W' are disjoint and $U = V' \cup W'$. Given $x, y \in OB^n$ we have a path $t \mapsto f((1-t)x + ty)$ from $f(x)$ to $f(y)$ in the set $f(OB^n) = V'$; this shows that V' is path connected. As $f(0) \in V'$ we see that $V' \subseteq V$. Next, Theorem 21.7 tells us that W' is acyclic and therefore also path connected. If there was a path in U joining some point in V' to some point in W', then we could conclude that the whole space $U = V' \cup W'$ was path connected. However, we know that U has two path components, so no such path can exist.

Now consider a point $x \in V$. As V is the path component of $f(0)$, we can find a path $u\colon [0,1] \to U$ with $u(0) = f(0)$ and $u(1) = x$. Here $u(0) \in V'$ and no path in U can cross from V' to W' so the point $u(1) = x$ must also lie in V'. This proves that $V = V' = f(OB^n)$. We have already remarked that V is open, and it follows that $f(OB^n)$ is open, as claimed. □

Corollary 21.16. *Let U be an open subset of $\mathbb{R}^n$, and let $f\colon U \to \mathbb{R}^n$ be a continuous injective map. Then $f(U)$ is also open.*

Proof. As usual we can identify $\mathbb{R}^n$ with the complement of a point in S^n, which is an open subset of S^n. It will therefore be enough to show that $f(U)$ is open in S^n. Consider a point $a \in U$. As U is open, we can choose $\epsilon > 0$ such that $OB(a,\epsilon)$ is contained in U. We then have a continuous injective map $g_a\colon B^n \to S^n$ given by $g_a(x) = f(a + \epsilon x/2)$. The proposition tells us that the set $V_a = g_a(OB^n)$ is open, and it is clear that $f(a) \in V_a = f(OB(a,\epsilon/2)) \subseteq f(U)$. It follows that $f(U)$ is the union of all these open sets V_a, and thus that $f(U)$ is open. $\square$

22. Covering maps

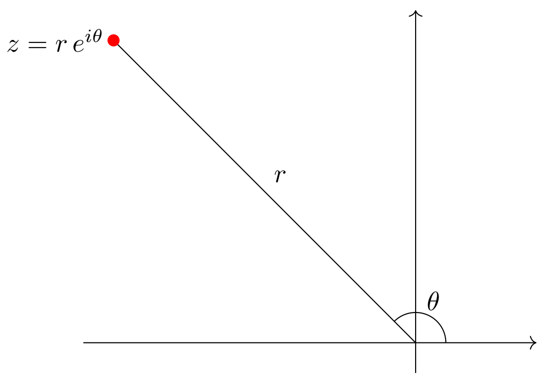
Consider a continuous map $p\colon X \to Y$. For each point $y \in Y$, we have a subset $p^{-1}\{y\} \subseteq X$, which we call the *fibre* of p over y. We next define what it means for p to be a *covering map*. The key points are that

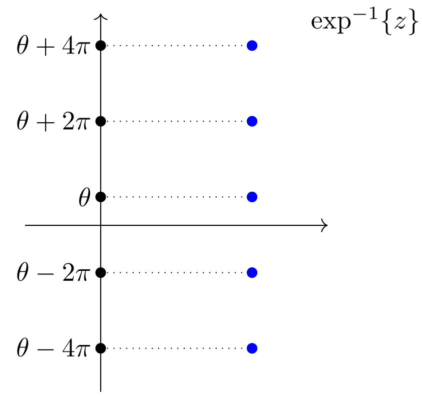
- All the fibres $p^{-1}\{y\}$ must be discrete subsets of X;
- The fibre $p^{-1}\{y\}$ must depend continuously on y, in an appropriate sense.

It is not easy to formulate the second condition directly, so the formal definition looks rather different from this informal discussion.

Example 22.1. Consider the exponential map $\exp\colon \mathbb{C} \to \mathbb{C} \setminus \{0\}$. The fibres are

$$\exp^{-1}\{r\,e^{i\theta}\} = \{\log(r) + i\theta + 2n\pi i \mid n \in \mathbb{Z}\}$$





Each fibre is a discrete set, which suggests that the map should be a covering. We will check that this is true once we have given the proper definition.

Definition 22.2. Let $p\colon X \to Y$ be a continuous map of spaces. Consider an open subset $V \subseteq Y$. We say that V is *trivially covered* by p if there is a discrete space F and a map $f\colon p^{-1}(V) \to F$ such that the combined map $\langle p, f\rangle\colon p^{-1}(V) \to V \times F$ is a homeomorphism. We say that p is a *covering map* (or that X is a *covering space* of Y) if for each point $y \in Y$ there is an open set V that contains y and is trivially covered.

Remark 22.3. Suppose we have open subsets $V' \subseteq V \subseteq Y$ and that V is trivially covered, as witnessed by a map $f\colon p^{-1}(V) \to F$. We then note that $p^{-1}(V') \subseteq p^{-1}(V)$, se we can restrict f to get a map $f'\colon p^{-1}(V') \to F$. It is not hard to check that the combined map $\langle p, f'\rangle\colon p^{-1}(V') \to V' \times F$ is a homeomorphism, so V' is also trivially covered.

Example 22.4. Take $X = \mathbb{R}\times\mathbb{Z}$ and $Y = \mathbb{R}$, and let $p\colon X \to Y$ be the projection map, given by $p(x,n) = x$. We claim that the whole space Y is trivially covered. Indeed, we can take $F = \mathbb{Z}$ and define $f\colon X \to F$ by $f(x,n) = n$. Then the combined map $\langle p, f\rangle\colon X \to Y \times F$ is just the identity map $\mathbb{R}\times\mathbb{Z} \to \mathbb{R}\times\mathbb{Z}$, which is certainly a homeomorphism. From this it is clear that p is a covering map.

Example 22.5. Take X and Y to be the unit circle in $\mathbb{C}$, or in other words $X = Y = \{e^{i\theta} \mid \theta \in \mathbb{R}\}$. Define $p\colon X \to Y$ by $p(z) = z^2$, or equivalently $p(e^{i\theta}) = e^{2i\theta}$. Each fibre $p^{-1}\{y\}$ is a discrete set consisting of the two square roots of y, which are negatives of each other. This suggests that p should be a covering, which we can prove as follows. We first take $V_0 = Y \setminus \{-1\}$, so each element $y \in V_0$ can be expressed in a unique way as $y = e^{i\theta}$ with $-\pi < \theta < \pi$. The two square roots of y are then $x = e^{i\theta/2}$ (which has $\mathrm{Re}(x) > 0$) and $-x = -e^{i\theta/2} = e^{i(\theta/2+\pi)}$ (which has $\mathrm{Re}(-x) < 0$). We therefore have

$$p^{-1}(V_0) = \{z \in X \mid z^2 \neq -1\} = X \setminus \{i, -i\} = \{z \in X \mid \mathrm{Re}(z) \neq 0\},$$

and we can define $f_0\colon p^{-1}(V_0)\to\{1,-1\}$ by

$$f_0(z)=\begin{cases}1 & \text{if } \mathrm{Re}(z)>0\\ -1 & \text{if } \mathrm{Re}(z)<0.\end{cases}$$

We then find that the map $\langle p,f_0\rangle\colon p^{-1}(V_0)\to V_0\times\{1,-1\}$ is a homeomorphism, showing that V_0 is trivially covered. A similar approach can be used to check that the set $V_1=Y\setminus\{1\}$ is also trivially covered. Any element $y\in V_1$ can be expressed uniquely as $y=e^{i\theta}$ with $0<\theta<2\pi$, and the two square roots are then $x=e^{i\theta/2}$ (which has $\mathrm{Im}(x)>0$) and $-x$ (which has $\mathrm{Im}(-x)<0$). We therefore have

$$p^{-1}(V_1)=\{z\in X\mid z^2\neq 1\}=X\setminus\{1,-1\}=\{z\in X\mid \mathrm{Im}(z)\neq 0\},$$

and we can define $f_1\colon p^{-1}(V_1)\to\{1,-1\}$ by

$$f_1(z)=\begin{cases}1 & \text{if } \mathrm{Im}(z)>0\\ -1 & \text{if } \mathrm{Im}(z)<0.\end{cases}$$

We then find that the map $\langle p,f_1\rangle\colon p^{-1}(V_1)\to V_1\times\{1,-1\}$ is a homeomorphism, as required. As $Y=V_0\cup V_1$, this proves that p is a covering.

Recall that the real projective space $\mathbb{R}P^n$ is defined to be the quotient space $S^n/\sim$, where $x\sim y$ iff $y=\pm x$. We therefore have a quotient map $\pi\colon S^n\to\mathbb{R}P^n$, for which $\pi(x)=\pi(y)$ iff $x=\pm y$. We give $\mathbb{R}P^n$ the quotient topology, which means that a subset $V\subseteq\mathbb{R}P^n$ is open iff $\pi^{-1}(V)$ is open in S^n.

Proposition 22.6. *The map $\pi\colon S^n\to\mathbb{R}P^n$ is a covering map.*

Proof. For each $a\in S^n$ we put $U_a=\{x\in S^n\mid x.a>0\}$ and $V_a=\pi(U_a)\subseteq\mathbb{R}P^n$. We then find that

$$\pi^{-1}(V_a)=\{x\mid \pi(x)\in\pi(U_a)\}=\{x\mid x\in U_a \text{ or } -x\in U_a\}=\{x\mid x.a\neq 0\}.$$

This first shows that $\pi^{-1}(V_a)$ is an open subset of S^n and so (by the definition of the quotient topology) that V_a is an open subset of $\mathbb{R}P^n$. It also allows us to define a continuous map $f_a\colon\pi^{-1}(V_a)\to\{1,-1\}$ by

$$f_a(x)=\begin{cases}1 & \text{if } x.a>0\\ -1 & \text{if } x.a<0.\end{cases}$$

It is easy to see that the combined map $\langle\pi,f_a\rangle\colon\pi^{-1}(V_a)\to V_a\times\{1,-1\}$ is a homeomorphism, so V_a is trivially covered. The sets V_a cover all of $\mathbb{R}P^n$ (because $\pi(a)\in V_a$) so π is a covering map as claimed. ☐

Proposition 22.7. *The map* $\exp\colon\mathbb{C}\to\mathbb{C}\setminus\{0\}$ *is a covering map, as is the map* $\exp\colon i\mathbb{R}\to S^1$ *(where we identify S^1 with $\{z\in\mathbb{C}:|z|=1\}$).*

Proof. This is closely related to the proof in Example 22.5. We put

$$\begin{aligned} V_0&=\mathbb{C}\setminus(-\infty,0] & V_1&=\mathbb{C}\setminus[0,\infty)\\ U_0&=\{x+iy\mid x\in\mathbb{R},-\pi<y<\pi\} & U_1&=\{x+iy\mid x\in\mathbb{R},0<y<2\pi\}\\ W_0&=\{x+iy\mid y \text{ is not an odd multiple of } \pi\} & W_1&=\{x+iy\mid y \text{ is not an even multiple of } \pi\}.\end{aligned}$$

If $y\in V_0$ then there is a unique choice of r and θ with $-\pi<\theta<\pi$ and $r>0$ and $y=r\,e^{i\theta}$. It follows that the number $x=\log(r)+i\theta$ lies in U_0 and has $\exp(x)=y$. This means that the restricted map $\exp\colon U_0\to V_0$ is bijective. Standard complex analysis shows that the inverse is also continuous, so we have a homeomorphism $\exp\colon U_0\to V_0$. Similarly, the restricted map $\exp\colon U_1\to V_1$ is also a homeomorphism. From this we see that

$$\begin{aligned}\exp^{-1}(V_0)&=\{x_0+2n\pi i\mid x_0\in U_0,\ n\in\mathbb{Z}\}=W_0\\ \exp^{-1}(V_1)&=\{x_1+2n\pi i\mid x_1\in U_1,\ n\in\mathbb{Z}\}=W_1.\end{aligned}$$

We can thus define continuous maps $f_i\colon W_i\to\mathbb{Z}$ (for $i=0,1$) by $f_i(x_i+2n\pi)=n$. Equivalently, $f_0(x+iy)$ is the closest integer to $y/(2\pi)$; this is well-defined and continuous on W_0 because points where y is an odd multiple of π have been removed from W_0. Similarly, $f_1(x+iy)$ is the closest integer to $(y-\pi)/(2\pi)$. We find that the maps $\langle\exp,p_0\rangle\colon W_0\to V_0\times\mathbb{Z}$ and $\langle\exp,p_1\rangle\colon W_1\to V_1\times\mathbb{Z}$ are homeomorphisms, so V_0 and V_1 are trivially covered. We also have $\mathbb{C}\setminus\{0\}=V_0\cup V_1$, so $\exp\colon\mathbb{C}\to\mathbb{C}\setminus\{0\}$ is a covering map. The proof for the restricted map $\exp\colon i\mathbb{R}\to S^1$ is essentially the same.

Interactive demo

$\square$

Definition 22.8. Let $p\colon X \to Y$ be a covering map.

(a) Consider a point $y \in Y$. A *lift* of y means a point $\widetilde{y} \in X$ with $p(\widetilde{y}) = y$.
(b) Consider a continuous path $u\colon [0,1] \to Y$. A *lift* of u means a continuous path $\widetilde{u}\colon [0,1] \to X$ such that $p \circ \widetilde{u} = u$. This means in particular that $\widetilde{u}(t)$ is a lift of $u(t)$ for all t.
(c) More generally, let T be any space and let $u\colon T \to Y$ be a continuous map. A *lift* of u is a continuous map $\widetilde{u}\colon T \to X$ with $p \circ \widetilde{u} = u$.
(d) For a map $u\colon T \to Y$ as in (c), we say that u is *small* if there is a trivially covered open set $V \subseteq Y$ such that $u(T) \subseteq V$.

Lemma 22.9. *Suppose we have a path-connected space T and a small continuous map $u\colon T \to Y$. Suppose we also have points $t_0 \in T$ and $x_0 \in X$ with $u(t_0) = p(x_0)$. Then there is a unique lift $\widetilde{u}\colon K \to X$ such that $\widetilde{u}(t_0) = x_0$.*

Proof. By assumption, we can choose an open subset $V \subseteq Y$ containing $u(K)$, and a homeomorphism $\langle p, f\rangle\colon p^{-1}(V) \to V \times F$ as in Definition 22.2. If $\widetilde{u}\colon T \to X$ is a lift of u, then we have $p(\widetilde{u}(t)) = u(t) \in V$, so $\widetilde{u}(t) \in p^{-1}(V)$ for all $t \in T$, so we have a well-defined and continuous composite $f \circ \widetilde{u}\colon T \to F$. As T is path connected and F is discrete, this must be constant. Thus, if $\widetilde{u}(t_0) = x_0$, then $f(\widetilde{u}(t)) = f(x_0)$ for all t. It follows that the only possibility is

$$\widetilde{u}(t) = \langle p, f\rangle^{-1}(u(t), f(x_0)).$$

$\square$

Proposition 22.10. *Let $p\colon X \to Y$ be a covering map. Let u be a path from a to b in Y, and let $\widetilde{a} \in X$ be a lift of a. Then there is a unique lift $\widetilde{u}$ of u such that $\widetilde{u}(0) = \widetilde{a}$.*

Proof. Because p is a covering map, we can find a family of trivially covered open sets $V_i \subseteq Y$ such that $Y = \bigcup_i V_i$. The preimages $u^{-1}(V_i)$ then form an open covering of $[0,1]$. Because $[0,1]$ is a compact metric space, this covering has a Lebesgue number $\epsilon > 0$ (by Proposition 8.31). Choose $n > 1/\epsilon$ and divide $[0,1]$ into subintervals $T_k = [(k-1)/n, k/n]$ for $k = 0, \dots, n$. By the Lebesgue number property, we can choose an index i_k such that $T_k \subseteq u^{-1}(V_{i_k})$, so $u(T_k) \subseteq V_{i_k}$. This means that the restriction of u to T_k is small, so Lemma 22.9 is applicable. We are given $\widetilde{a} \in X$ with $p(\widetilde{a}) = a = u(0)$. We put $x_0 = \widetilde{a}$, and use Lemma 22.9 to show that there is a unique map $\widetilde{u}_1\colon T_1 \to X$ with $p(\widetilde{u}_1(t)) = u(t)$ and $\widetilde{u}_1(0) = x_0$. We now define $x_1 = \widetilde{u}_1(1/n)$, so $p(x_1) = u(1/n)$. Applying Lemma 22.9 again, we see that there is a unique map $\widetilde{u}_2\colon T_2 \to X$ with $p(\widetilde{u}_2(t)) = u(t)$ and $\widetilde{u}_2(1/n) = x_1$. We put $x_2 = \widetilde{u}_2(2/n) \in X$, so $p(x_2) = u(2/n)$. We then repeat the process in the obvious way, to get a family of maps $\widetilde{u}_k\colon [(k-1)/n, k/n] \to X$ and points $x_k \in X$ with $\widetilde{u}_k(k/n) = x_k = \widetilde{u}_{k+1}(k/n)$. It follows that the maps $\widetilde{u}_k$ can be patched together to give a continuous map $\widetilde{u}\colon [0,1] \to X$ with $p \circ \widetilde{u} = u$ and $\widetilde{u}(0) = x_0 = \widetilde{a}$. The same kind of induction shows that this is unique. $\square$

Remark 22.11. Note that Proposition 22.10 does not say anything about $\widetilde{u}(1)$. We know that $p(\widetilde{u}(t)) = u(t)$ for all t, so in particular $p(\widetilde{u}(1)) = u(1) = b$, so $\widetilde{u}(1) \in p^{-1}\{b\}$. However, if we have two different paths $u, v\colon a \rightsquigarrow b$ in Y and we use the same starting point $\widetilde{a} \in p^{-1}\{a\}$ in both cases, then it can easily happen that the endpoints $\widetilde{u}(1), \widetilde{v}(1)$ are different elements of $p^{-1}\{b\}$. However, this cannot happen if there is a pinned homotopy between u and v, as we will show later.

Corollary 22.12. *Let $p\colon X \to Y$ be a covering map. Suppose we have a path-connected space T and a continuous map $u\colon T \to Y$. Suppose that $m, n\colon T \to X$ are continuous lifts of u, and that there is at least one point $t_0 \in T$ with $m(t_0) = n(t_0)$; then $m = n$.*

Proof. Consider a point $t \in T$; we must show that $m(t) = n(t)$. As T is path connected, we can choose a path v from t_0 to t in T. Now $m \circ v$ and $n \circ v$ are both lifts of the path $u \circ v\colon [0,1] \to Y$, and they satisfy $(m \circ v)(0) = m(t_0) = n(t_0) = (n \circ v)(0)$. Thus, the uniqueness clause in Proposition 22.10 tells us that $m \circ v = n \circ v$. In particular, we have $(m \circ v)(1) = (n \circ v)(1)$, or in other words $m(t) = n(t)$ as required. $\square$

Proposition 22.13. *Let T be a compact convex subset of $\mathbb{R}^N$, and suppose that $t_0 \in T$. Let $p: X \to Y$ be a covering map, and let $u: T \to Y$ be continuous. Suppose that $x_0 \in X$ with $p(x_0) = u(t_0)$. Then there is a unique continuous lift $\widetilde{u}: T \to X$ with $p \circ \widetilde{u} = u$ and $\widetilde{u}(t_0) = x_0$.*

Proof. Roughly speaking, the idea is as follows: to define $\widetilde{u}(t)$, we move a short distance from t towards t_0 to reach a point t', then $\widetilde{u}(t')$ will already be defined and we define $\widetilde{u}(t)$ to be the unique lift of $u(t)$ that is close to $\widetilde{u}(t')$. The rest of this proof should be seen as a more complete and rigorous version of this idea.

We first claim that there exists $\epsilon > 0$ such that for all $t \in T$, the restricted map $u: OB(t, \epsilon) \to Y$ is small. (Here and elsewhere in this proof, notation for balls should be interpreted relative to T, so $OB(t, \epsilon) = \{t' \in T \mid \|t - t'\| < \epsilon\}$.) Indeed, for each $t \in T$ we can choose a trivially covered open set $V_t \subseteq Y$ containing $u(t)$. The set $u^{-1}(V_t)$ is then open in T and contains t. This means that the sets $u^{-1}(V_t)$ form an open cover of the compact metric space T, so there is a Lebesgue number $\epsilon > 0$. This has the required property.

Next, for $j > 0$ we put $T_j = \{t \in T \mid \|t - t_0\| \leq j\epsilon/2\}$. We will prove by induction on j that there is a unique continuous map $\widetilde{u}_j: T_j \to X$ with $\widetilde{u}_j(t_0) = x_0$ and $p(\widetilde{u}_j(t)) = u(t)$ for all $t \in T_j$. To start with, the map $u: T_1 \to Y$ is small by our choice of ϵ, so Lemma 22.9 gives $\widetilde{u}_1$. Suppose we have already constructed $\widetilde{u}_j$. For each $a \in T_j$, we note that the map $u: OB(a, \epsilon) \to Y$ is small, so there is a unique continuous $v_a: OB(a, \epsilon) \to X$ lifting u with $v_a(a) = \widetilde{u}_j(a)$. Both $\widetilde{u}_j$ and v_a restrict to give lifts of u over the convex set $OB(a, \epsilon) \cap T_j$, and they agree at the point a, and the restricted map $u: OB(a, \epsilon) \cap T_j \to Y$ is small; it follows that v_a agrees with $\widetilde{u}_j$ on $OB(a, \epsilon) \cap T_j$.

Now suppose that $a, b \in T_j$ and that the set $U = OB(a, \epsilon) \cap OB(b, \epsilon)$ is nonempty. We then find that the point $c = (a + b)/2$ must lie in U and it also lies in T_j because T_j is convex. The maps v_a and v_b both agree with $\widetilde{u}_j$ at c, so they agree with each other. The restricted map $u: U \to Y$ is small, so we conclude that $v_a|_U = v_b|_U$. Because of this consistency property, we see that the maps v_a can be combined to give a map $v: \bigcup_{a \in T_j} OB(a, \epsilon) \to X$. As the sets $OB(a, \epsilon)$ are all open, an open patching argument shows that v is continuous. As each map v_a is a lift of u, we see that v is a lift of u. As v_a agrees with $\widetilde{u}_j$ on T_j, we see that v agrees with $\widetilde{u}_j$ on T_j.

It is also easy to see that T_{j+1} is contained in the domain of v, so we can define $\widetilde{u}_{j+1}$ to be the restriction of v to T_{j+1}. This is a continuous lift of u extending $\widetilde{u}_j$ and therefore satisfying $\widetilde{u}_{j+1}(t_0) = x_0$ as required.

As T is assumed to be compact, it must be bounded. We therefore have $T_j = T$ for sufficiently large j, and this completes the proof. □

23. Transfers, coefficients and homology of projective spaces

Definition 23.1. Let n be a nonnegative integer. We say that a map $p: X \to Y$ is an *n-sheeted covering* if it is a covering map, and $|p^{-1}\{y\}| = n$ for all $y \in Y$.

Example 23.2. For any $n > 0$ we have an n-sheeted covering $p: \mathbb{C}^\times \to \mathbb{C}^\times$ given by $p(z) = z^n$. This restricts to give an n-sheeted covering $S^1 \to S^1$.

Example 23.3. For any space Y and any discrete set F with $|F| = n$, the projection $Y \times F \to Y$ is an n-sheeted covering.

Example 23.4. For any $n > 0$, the projection $p: S^n \to \mathbb{R}P^n$ is a 2-sheeted covering.

Lemma 23.5. *Let $p: X \to Y$ be an n-sheeted covering, and let $u: \Delta_k \to Y$ be continuous. Then there are precisely n different continuous maps $\Delta_k \to X$ lifting u.*

Proof. By assumption, the set $F = p^{-1}\{u(e_0)\}$ has size n, say $F = \{x_1, \ldots, x_n\}$. Proposition 22.13 tells us that for each i there us a unique lift $\widetilde{u}_i: \Delta_k \to X$ with $\widetilde{u}_i(e_0) = x_i$. If $\widetilde{u}: \Delta_k \to X$ is an arbitrary lift of u, then $p(\widetilde{u}(e_0)) = u(e_0)$ so $\widetilde{u}(e_0) \in F$ so $\widetilde{u}(e_0) = x_i$ for some i, so $\widetilde{u} = \widetilde{u}_i$. □

Definition 23.6. Let $p: X \to Y$ be an n-sheeted covering. For any continuous map $u: \Delta_k \to Y$, we define $\tau(u)$ to be the sum of all the lifts of u, considered as an element of $C_k(X)$. More generally, given an element $u = m_1 u_1 + \cdots + m_r u_r \in C_k(Y)$, we define $\tau(u) = m_1\tau(u_1) + \cdots + m_r\tau(u_r) \in C_k(X)$. This defines a homomorphism $\tau: C_k(Y) \to C_k(X)$, which is called the *transfer*.

Example 23.7. Define $p\colon \mathbb{C}^\times \to \mathbb{C}^\times$ by $p(z) = z^3$, so this is a 3-sheeted covering. Define $u\colon \Delta_1 \to \mathbb{C}^\times$ by $u(1-t,t) = 8\exp(2\pi i t)$. Then $u(e_0) = 8$ so $p^{-1}\{u(e_0)\} = \{2, 2e^{2\pi i/3}, 2e^{4\pi i/3}\}$. Define $v_j\colon \Delta_1 \to \mathbb{C}^\times$ by $v_j(1-t,t) = 2\exp(2\pi i(t+j)/3)$ for $j = 0,1,2$. These are the three lifts of u, so $\tau(u) = v_0+v_1+v_2 \in C_1(\mathbb{C}^\times)$.

Proposition 23.8. *Let $p\colon X \to Y$ be an n-sheeted covering. Then the associated transfer map $\tau\colon C_*(Y) \to C_*(X)$ is a chain map, and satisfies $p_\#(\tau(u)) = nu$ for all $u \in C_k(Y)$.*

Proof. Consider a continuous map $u\colon \Delta_k \to Y$, and let $v_1,\dots,v_n$ be the continuous lifts of u, so $\tau(u) = \sum_{j=1}^n v_j$. This means that $\partial(\tau(u)) = \sum_{i=0}^k \sum_{j=1}^n (-1)^i (v_j \circ \delta_i)$. Now note that $p\circ(v_j\circ\delta_i) = u\circ\delta_i$, so $v_j\circ\delta_i$ is one of the lifts of $u\circ\delta_i$. If $v_j\circ\delta_i = v_{j'}\circ\delta_i$ then v_j and $v_{j'}$ agree at $\delta_i(e_0)$ so they must be the same so $j = j'$. This proves that the list $v_1\circ\delta_i,\dots,v_n\circ\delta_i$ is the complete list of lifts of $u\circ\delta_i$, so $\tau(u\circ\delta_i) = \sum_{j=1}^n (v_j\circ\delta_i)$. From this we get

$$\tau(\partial(u)) = \sum_{i=0}^k (-1)^i \tau(u\circ\delta_i) = \sum_{i=0}^k \sum_{j=1}^n (-1)^i (v_j\circ\delta_i) = \partial(\tau(u)).$$

This proves that τ is a chain map. As $p\circ v_j = u$ for all j we also have $p_\#(\tau(u)) = p_\#(\sum_{j=1}^n v_j) = \sum_{j=1}^n u = n\,u$. □

Remark 23.9. It follows that we have an induced map $\tau_*\colon H_*(Y) \to H_*(X)$, which satisfies $p_*(\tau_*(u)) = n\,u$ for all $u \in H_k(Y)$.

We would like to use the transfer to obtain homological information about $\mathbb{R}P^n$. For this, it is convenient to use a slightly different version of homology.

Definition 23.10. We define $C_k(X;\mathbb{Z}/2)$ to be the set of formal linear combinations $m_1u_1 + \dots + m_ru_r$ where each u_i is a continuous map $\Delta_k \to X$, but now the coefficients m_i lie in $\mathbb{Z}/2$ rather than $\mathbb{Z}$. We again make this a chain complex by defining $\partial(u) = \sum_{i=0}^k (u\circ\delta_i)$. (We have left out the sign $(-1)^i$ because it makes no difference mod 2.) We define $H_*(X;\mathbb{Z}/2)$ to be the homology of this chain complex.

Remark 23.11. If all the groups $H_i(X)$ are free abelian groups, one can check that $H_i(X;\mathbb{Z}/2) = H_i(X)/2H_i(X)$ for all i. If some groups $H_i(X)$ are not free abelian, then the relationship between $H_i(X)$ and $H_i(X;\mathbb{Z}/2)$ is a little more complicated. In particular, this applies when $X = \mathbb{R}P^n$, because we have already seen that $H_1(\mathbb{R}P^n) = \mathbb{Z}/2$ for $n > 1$, and this is not a free abelian group.

Remark 23.12. Any element $u \in C_k(X;\mathbb{Z}/2)$ can be expressed as a formal linear combination $m_1u_1 + \dots + m_ru_r$ with $m_i \in \mathbb{Z}/2$. If $u_i = u_j$ for some $i \neq j$ then we can combine the corresponding terms. We can then discard all terms with coefficient zero. As $\mathbb{Z}/2 = \{0,1\}$, and remaining terms must have coefficient 1. This means that u can be expressed as $u_1 + \dots + u_s$, where the elements u_i are distinct maps from Δ_k to X.

Remark 23.13. Essentially everything that we have done previously works in the same way with coefficients $\mathbb{Z}/2$. In particular, the groups $H_*(X;\mathbb{Z}/2)$ are functorial and homotopy invariant, and we have Mayer-Vietoris sequences and transfers. For $n > 0$ we have

$$H_k(S^n;\mathbb{Z}/2) = \begin{cases} \mathbb{Z}/2 & \text{if } k = 0 \text{ or } k = n \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 23.14. *Let $p\colon S^n \to \mathbb{R}P^n$ be the usual projection, which is a 2-sheeted covering. Then the sequence*

$$C_*(\mathbb{R}P^n;\mathbb{Z}/2) \xrightarrow{\tau} C_*(S^n;\mathbb{Z}/2) \xrightarrow{p_\#} C_*(\mathbb{R}P^n;\mathbb{Z}/2)$$

is a short exact sequence of chain complexes and chain maps. It therefore gives a long exact sequence of homology groups

$$H_i(S^n;\mathbb{Z}/2) \xrightarrow{p_*} H_i(\mathbb{R}P^n;\mathbb{Z}/2) \xrightarrow{\Delta} H_{i-1}(\mathbb{R}P^n;\mathbb{Z}/2) \xrightarrow{\tau_*} H_{i-1}(S^n;\mathbb{Z}/2) \xrightarrow{p_*} H_{i-1}(\mathbb{R}P^n;\mathbb{Z}/2)$$

Proof. First suppose that $u \in C_k(\mathbb{R}P^n;\mathbb{Z}/2)$. As in Remark 23.12, we can write $u = u_1 + \ldots + u_r$ for some list of distinct maps $u_i\colon \Delta_k \to \mathbb{R}P^n$. Let u_i' and u_i'' be the two lifts of u_i. Note that $p_\#(\sum_i u_i') = u$; this proves that $p_\#$ is surjective. Note also that $\tau(u) = \sum_i (u_i' + u_i'')$. If $i \neq j$ then

$$p\circ u_i' = p\circ u_i'' = u_i \neq u_j = p\circ u_j' = p\circ u_j'',$$

so neither u_i' nor u_i'' can be equal to u_j' or u_j''. Also, $u_i' \neq u_i''$ by construction. Thus, there can be no cancellation in our expression for $\tau(u)$, so $\tau(u) \neq 0$ except in the case where our original expression for u had no terms. This proves that τ is injective. We also have $p_\#(\tau(u)) = 2u$, which is zero as we are working modulo 2. This shows that $\mathrm{img}(\tau) \leq \ker(p_\#)$. Conversely, suppose we have an element $v \in \ker(p_\#)$. As in Remark 23.12, we can write $v = v_1 + \ldots + v_r$ for some distinct continuous maps $v_i \colon \Delta_k \to S^n$. Put $u_i = p \circ v_i \colon \Delta_k \to \mathbb{R}P^n$, so that $p_\#(v) = u_1 + \ldots + u_r$. We are assuming that $v \in \ker(p_\#)$, so the sum $u_1 + \ldots + u_r$ must cancel down to zero. After reordering the terms if necessary, we can assume that $r = 2r'$ for some r' and $u_{2i-1} = u_{2i}$ for $i = 1, \ldots, r'$. This means that v_{2i-1} and v_{2i} must be the two different lifts of u_{2i}, so $v = \tau(u_2 + u_4 + \ldots + u_{2r'})$. We conclude that $\mathrm{img}(\tau) = \ker(p_\#)$. This completes the proof that we have a short exact sequence of chain complexes and chain maps, and the Snake Lemma gives the claimed long exact sequence of homology groups. □

Theorem 23.15. *For any $n > 0$ we have*

$$H_k(\mathbb{R}P^n; \mathbb{Z}/2) = \begin{cases} \mathbb{Z}/2 & \text{if } 0 \leq k \leq n \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, the map $\tau_* \colon H_n(\mathbb{R}P^n; \mathbb{Z}/2) \to H_n(S^n; \mathbb{Z}/2)$ is an isomorphism, as are the maps $\Delta \colon H_i(\mathbb{R}P^n; \mathbb{Z}/2) \to H_{i-1}(\mathbb{R}P^n; \mathbb{Z}/2)$ for $1 \leq i \leq n$.

Proof. We proved in Proposition 20.9 that

$$H_i(\mathbb{R}P^n) = \begin{cases} \mathbb{Z} & \text{if } i = 0 \\ \mathbb{Z}/2 & \text{if } i = 1 \\ 0 & \text{if } i > n. \end{cases}$$

Essentially the same argument shows that

$$H_i(\mathbb{R}P^n; \mathbb{Z}/2) = \begin{cases} \mathbb{Z}/2 & \text{if } i = 0, 1 \\ 0 & \text{if } i > n. \end{cases}$$

Next, we have a long exact sequence

$$H_i(S^n; \mathbb{Z}/2) \xrightarrow{p_*} H_i(\mathbb{R}P^n; \mathbb{Z}/2) \xrightarrow{\Delta} H_{i-1}(\mathbb{R}P^n; \mathbb{Z}/2) \xrightarrow{\tau_*} H_{i-1}(S^n; \mathbb{Z}/2)$$

For $2 \leq i \leq n-1$ we have $H_i(S^n; \mathbb{Z}/2) = H_{i-1}(S^n; \mathbb{Z}/2) = 0$, so the map Δ is an isomorphism. It follows by induction on i that $H_i(\mathbb{R}P^n; \mathbb{Z}/2) = \mathbb{Z}/2$ for $0 \leq i \leq n-1$. Finally, we have an exact sequence

$$H_{n+1}(\mathbb{R}P^n; \mathbb{Z}/2) \xrightarrow{\Delta} H_n(\mathbb{R}P^n; \mathbb{Z}/2) \xrightarrow{\tau_*} H_n(S^n; \mathbb{Z}/2) \xrightarrow{p_*} H_n(\mathbb{R}P^n; \mathbb{Z}/2) \xrightarrow{\Delta} H_{n-1}(\mathbb{R}P^n; \mathbb{Z}/2) \xrightarrow{\tau_*} H_{n-1}(S^n; \mathbb{Z}/2).$$

After filling in the known groups, this becomes

$$0 \to H_n(\mathbb{R}P^n; \mathbb{Z}/2) \xrightarrow{\tau_*} \mathbb{Z}/2 \xrightarrow{p_*} H_n(\mathbb{R}P^n; \mathbb{Z}/2) \xrightarrow{\Delta} \mathbb{Z}/2 \to 0.$$

This shows that the first map τ_* is injective, so $H_n(\mathbb{R}P^n; \mathbb{Z}/2)$ is isomorphic to a subgroup of $\mathbb{Z}/2$, so it is either trivial or of order two. If it was trivial then the sequence could not be exact, so we must have $H_n(\mathbb{R}P^n; \mathbb{Z}/2) \simeq \mathbb{Z}/2$ as claimed. Given this, the only way the sequence can be exact is if τ_* and Δ are isomorphisms, and $p_* = 0$. □

24. Borsuk-Ulam and related results

Definition 24.1. A continuous map $f \colon S^n \to S^m$ is *odd* (or *antipodal*) if $f(-x) = -f(x)$ for all $x \in S^n$.

Example 24.2. If $n \leq m$ then it is easy to produce examples of odd continuous maps $f \colon S^n \to S^m$. Most obviously, we can just define

$$f(x_0, \ldots, x_n) = (x_0, \ldots, x_n, 0, \ldots, 0).$$

Remark 24.3. If $f \colon S^n \to S^m$ is odd, then we have a well-defined map $\overline{f} \colon \mathbb{R}P^n \to \mathbb{R}P^m$ given by $\overline{f}([x]) = [f(x)]$.

Proposition 24.4. *Suppose that $f \colon S^n \to S^n$ is continuous and odd. Then the induced map $f_* \colon H_n(S^n; \mathbb{Z}/2) \to H_n(S^n; \mathbb{Z}/2)$ is the identity.*

Proof. As f is odd we have a well-defined map $\overline{f}\colon \mathbb{R}P^n \to \mathbb{R}P^n$ given by $\overline{f}([x]) = [f(x)]$. This satisfies $p \circ f = \overline{f} \circ p\colon S^n \to \mathbb{R}P^n$, so the right-hand square below commutes:

$$\begin{tikzcd}
C_i(\mathbb{R}P^n;\mathbb{Z}/2) \arrow[r, "\tau"] \arrow[d, "\overline{f}_\#"'] & C_i(S^n;\mathbb{Z}/2) \arrow[r, "p_\#"] \arrow[d, "f_\#"'] & C_i(\mathbb{R}P^n;\mathbb{Z}/2) \arrow[d, "\overline{f}_\#"] \\
C_i(\mathbb{R}P^n;\mathbb{Z}/2) \arrow[r, "\tau"'] & C_i(S^n;\mathbb{Z}/2) \arrow[r, "p_\#"'] & C_i(\mathbb{R}P^n;\mathbb{Z}/2)
\end{tikzcd}$$

We claim that the left hand square commutes as well. To see this, define $\chi\colon S^n \to S^n$ by $\chi(x) = -x$, so $f \circ \chi = \chi \circ f$ and $p \circ \chi = p$. Consider a continuous map $u\colon \Delta_k \to \mathbb{R}P^n$. Choose a lift $v\colon \Delta_k \to S^n$. The other lift is then $\chi \circ v$, so $\tau(u) = v + (\chi \circ v)$, so

$$f_\#(\tau(u)) = (f \circ v) + (f \circ \chi \circ v) = (f \circ v) + (\chi \circ f \circ v).$$

Here $f \circ v$ and $\chi \circ f \circ v$ are the two lifts of the map $p \circ f \circ v = \overline{f} \circ p \circ v = \overline{f} \circ u$, so we see that $(f \circ v) + (\chi \circ f \circ v) = \tau(\overline{f}_\#(u))$. This shows that $f_\# \circ \tau = \tau \circ \overline{f}_\#$, so the left hand square commutes as claimed.

As the diagram commutes, we see that the maps f_* and $\overline{f}_*$ are compatible with the maps in the exact sequence obtained in Lemma 23.14. In particular, we have commutative squares

$$\begin{tikzcd}
H_i(\mathbb{R}P^n;\mathbb{Z}/2) \arrow[r, "\Delta"] \arrow[d, "\overline{f}_*"'] & H_{i-1}(\mathbb{R}P^n;\mathbb{Z}/2) \arrow[d, "\overline{f}_*"] \\
H_i(\mathbb{R}P^n;\mathbb{Z}/2) \arrow[r, "\Delta"] & H_{i-1}(\mathbb{R}P^n;\mathbb{Z}/2)
\end{tikzcd}$$

for $1 \leq i \leq n$. It is clear that $\overline{f}_*$ gives the identity on $H_0(\mathbb{R}P^n;\mathbb{Z}/2)$, and we have seen that all of the maps Δ are isomorphisms, so it follows inductively that $\overline{f}_*$ gives the identity on $H_i(\mathbb{R}P^n;\mathbb{Z}/2)$ for $1 \leq i \leq n$. We also have a commutative square

$$\begin{tikzcd}
H_n(\mathbb{R}P^n;\mathbb{Z}/2) \arrow[r, "\tau_*"] \arrow[d, "\overline{f}_*"'] & H_n(S^n;\mathbb{Z}/2) \arrow[d, "f_*"] \\
H_n(\mathbb{R}P^n;\mathbb{Z}/2) \arrow[r, "\tau_*"'] & H_n(S^n;\mathbb{Z}/2).
\end{tikzcd}$$

We have seen that τ_* is an isomorphism and $\overline{f}_*$ is the identity so f_* is also the identity, as claimed. □

Theorem 24.5 (Borsuk-Ulam)**.** *If $n > m$, then there are no odd continuous maps from S^n to S^m.*

Proof. Suppose that $f\colon S^n \to S^m$ is odd and continuous. Define $i\colon S^m \to S^n$ by

$$i(x_0, \dots, x_m) = (x_0, \dots, x_m, 0, \dots, 0).$$

It is clear that i is odd, so the composite $f \circ i\colon S^m \to S^m$ is odd, so the induced map $(f \circ i)_*\colon H_m(S^m;\mathbb{Z}/2) \to H_m(S^m;\mathbb{Z}/2)$ must be the identity by Proposition 24.4. In particular, $(f \circ i)_*$ is nonzero.

On the other hand, we can define $h\colon [0,1] \times S^m \to S^n$ by

$$h(t,x) = \cos(\pi t/2) i(x) + \sin(\pi t/2) e_{m+1}.$$

(Using the fact that $i(x)$ and e_{m+1} are orthogonal, we see that this does indeed lie in S^n.) This gives a homotopy between i and a constant map, which implies that $f \circ i$ is also homotopic to a constant map, so $(f \circ i)_* = 0$. This contradiction shows that no such map f can exist. □

Corollary 24.6. *Let $g\colon S^n \to \mathbb{R}^m$ be a continuous map, with $0 < m \leq n$. Then there is a point $x \in S^n$ with $g(x) = g(-x)$.*

Proof. Suppose (for a contradiction) that no such point exists, so $g(x) - g(-x)$ is always nonzero. We can then define $f\colon S^n \to S^{n-1}$ by $f(x) = (g(x) - g(-x))/\|g(x) - g(-x)\|$. It is easy to check that this is continuous and antipodal, which contradicts Theorem 24.5, as required. □

Example 24.7. There are two opposite points on the Earth's surface that have the same temperature and also the same atmospheric pressure, as we see by considering the map $f\colon S^2 \to \mathbb{R}^2$ given by

$$f(a) = (\text{ temperature at } a,\ \text{ pressure at } a).$$

Theorem 24.8 (Sandwich Slicing Theorem)**.** *Let A_1, A_2 and A_3 be three reasonable subsets of $\mathbb{R}^3$. Then there is a plane $P \subset \mathbb{R}^3$ such that for each set A_i, half of the volume lies on one side of P, and half of the volume lies on the other side.*

Example 24.9. We could have a sandwich, with the top slice of bread filling the set A_1, and cheese filling the set A_2, and the bottom slice of bread filling A_3. The theorem then says that we can make a single straight cut with a knife to share all three components equally.

Example 24.10. Suppose that the sets A_i are solid balls, with centres a_i in general position. Then there is a unique possible choice for P, namely the plane passing through a_1, a_2 and a_3.

We will not be very rigorous about what "reasonable" means, but we will make some comments here and in the body of the proof. To start with, each set A_i should be bounded (which implies that the volume is finite) and the volume should also not be zero.

Proof of Theorem 24.8. For any unit vector $u = (u_0, u_1, u_2, u_3) \in S^3$ and $v = (v_1, v_2, v_3) \in \mathbb{R}^3$ we put $m(u, v) = u_0 + u_1 v_1 + u_2 v_2 + u_3 v_3 \in \mathbb{R}$, so $m(-u, v) = -m(u, v)$. We then put

$$P(u) = \{v \in \mathbb{R}^3 \mid m(u, v) = 0\}$$
$$H(u) = \{v \in \mathbb{R}^3 \mid m(u, v) > 0\},$$

so $P(u)$ is a plane and $H(u)$ is the half-space on one side of that plane. Note that the plane $P(-u)$ is the same as $P(u)$, and $H(-u)$ is the half-space on the opposite side to $H(u)$.

We now define $g\colon S^3 \to \mathbb{R}^3$ by

$$g(u)_i = \mathrm{vol}(A_i \cap H(u)).$$

It is not too hard to check that this is continuous. By Corollary 24.6, there exists $u \in S^3$ with $g(-u) = g(u)$, which means that $\mathrm{vol}(A_i \cap H(u)) = \mathrm{vol}(A_i \cap H(-u))$ for $i = 1, 2, 3$. In other words, the plane $P(u)$ bisects each of the sets A_i. □