Topology

Paul Martin

Based partly on some lovely notes by Ben Sharp,
Josh Cork, Derek Harland and others

2021

Contents

Chapter 1

Introduction

“Continuity” is a very useful notion in human thought. But what does it mean, exactly? In ‘human thought’, the term represents a bundle of useful ideas - we will talk about some of them below - but let us assume for a moment that this bundle of useful ideas is agreed. Our question is about how to make these ideas unified and precise, and thus even more useful. And Topology is the name given to the mathematical approach to this problem. *A topology* is the minimal extra structure with which we must equip a set (such as physical space) so that the idea of “continuity” makes sense.

The following two sections are introductory descriptions of topology, from different perspectives. (The two sections are similar — almost the same. What does it mean to say that they are almost the same? That itself is a question in topology.)

1.1 Introduction: a generalist’s viewpoint

Are you like me? If you are, then you wake up on Wednesday morning feeling that you are pretty much the same person you were when you went to bed on Tuesday night. This notion of same-ness is part of our sense of self. While sleep may have interupted the continuity of consciousness, there is a deeper feeling of continuity of the self that survives sleep. The subject of ‘Topology’ is about studying and using the general notion of continuity. It underpins our sense of self, but it is also very useful in countless other ways, as we will see.

A set X becomes a topological space if it has ‘a topology’. A topology is a collection of subsets of X that must include X and $\emptyset$ and be closed under finite intersections and all unions (see Definition 7.1). To have a reasonable notion of continuity for functions f on a space X to a space Y, we simply require the existence of a topology on each.

(This remarkable simplicity seems just to be a piece of good luck. It is worth admitting that. Just so that we can grasp the intellectual starting point for this branch of study. But

because of this good luck,...)
...We can study many profound and useful organisational tools for life at quite a good level of mathematical rigour.

Having said that, our definition allows for many different topologies on a set. And each one can give rise to a different version of continuity. Depending on the application, some are more useful than others, as we will see.

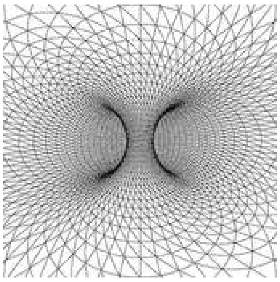
We will determine (rigorous and abstract) properties of topological spaces which will help us to classify them in useful ways; and then to organise specific sets of topological spaces according to their classification. Specifically, we will introduce the organisational notions of connectedness, compactness and 'shape'.

When it comes down to detail, different people may have different notions of sameness (think of the different possible notions associated to "she bought the same newspaper every day"). A Topologist's notion of sameness equates any two spaces if (and only if) one can be continuously mapped to the other, bijectively, and such that the inverse map is also continuous. We will explain and show how to use all these terms.

Why is this useful? (Apart from in understanding one's sense of self.)
Well, its a big scary world. Some tiger-sized things will eat you. Some won't. We need a reliable and practical way to detect the dangerous ones. (Choosing the tiger example in a country with no wild tigers is to choose a lighthearted example, but hopefully the transferability of the point will be clear.) If a given tiger eats you (or someone) then we can mark that tiger down as dangerous. But do you want to give other tigers the benefit of the doubt, or lump them together, just to be safe? Topology is the maths behind the how and what of this 'lumping together'.

Tigers have quite complicated shapes, so let's start with something simpler. Topology sees equivalence between the boundary of a cube, and the boundary of a solid ball (a sphere) — there is a continuous bijection from one to the other whose inverse is also continuous. (An example of such a map?: Imagine putting the cube inside the sphere, and picture light rays radiating from a point inside the cube!) Similarly, the surface of a ring donut is equivalent to the surface of a coffee mug, from the perspective of 'holdability' topology (both have a 'handle').

Can one continuously, and bijectively map the surface of a solid ball to that of a donut with a map whose inverse is continuous? We can rigorously prove that the answer is "NO!"



One can sometimes think of topological manifolds (special kinds of topological spaces) as being made out of some flexible material (flesh, fabric or rubber, say) so that a lot of bending, stretching and shrinking is allowed without changing the underlying topological structure.

Another place where we use topology heavily (again usually subconsciously) is in reading and writing. And indeed in communication generally. The notion that the words BUBBBLEGUM and NARNIA can be found in this picture:

involves a lot of 'experiments' with topological spaces and continuous functions. There are only finitely many words in a dictionary (certainly only finitely many that we need to know), but very many more different ways of writing each letter and each word.

One of our assumptions in these notes is that you can read them! (Passing from formal reading to understanding is another story, but we have to — indeed can — assume that you can read... or you would not have got this far.) If you can read then you already understand and are familiar with a lot of complex concepts and operations. Shortly it will be helpful to us to draw some of these operations (such as arranging objects into ordered sequences) to more explicit attention.

1.1.1 Overview

There are several useful kinds of relationships explained by topology. Useful because they explain how we classify unmanageably large collections of things into useful smaller groupings. One kind is a natural extension of the idea of cardinality (organising sets by size) to sets that are spaces: 'homeomorphism'. A second kind combines topology with everybody's favourite example of a mathematical structure — the real line, and uses it to relate different images of one space in another space: 'homotopy'. Finally, classification is all very well in principle, but how do we do it in practice? How do we tell which class each element belongs to? What can be useful here is a (computable) function on the set being classified that takes all elements of the same class to the same image point — an 'invariant'.

After studying foundational definitions in section 2-7; we investigate homeomorphism in section 9-10; invariants in section 11; homotopy in section 12.

The tools and ideas of topology are used wherever organisation and classification are useful, and hence across all realms of thought. In the course we will prove some powerful results in different fields. E.g. the fundamental theorem of algebra, and a 'ham sandwich' theorem (which proves that any sandwich made from bread, butter, and ham can always be sliced (with a single cut) into two parts, so that each part consists of equal quantities of the three separate ingredients). We can also prove that at each moment in time, there exist antipodal points on the surface of the earth which have the same temperature *and* pressure (assuming that temperature and pressure vary continuously in space).

In the realm of risk assessment we mentioned before — the risk posed by various tiger-sized things — what we need is a classification scheme for such things: a way of quickly identifying them into a grouping of established risk level. For tiger-like things in particular we tend to do this by looking at the surface (assuming the appropriate size of course). The exact shape is not helpful because tigers articulate their bodies when they run, but can we classify things like tigers, say, according to some common properties of their surfaces? In fact we can classify all surfaces up to the kind of articulations and movements that tigers can do. (In practice most hunted animals do their risk assessment classifications essentially subconsciously, rather than with maths research. But it is empowering, and transferable, to know how this works.)

We can also see topology as a bridge between the major sub-disciplines of modern mathematics, sometimes called algebra, analysis, geometry and logic. As we go through, we will see many examples of this bridge in action. Indeed the way it connects the disciplines shines a light on the nature of the disciplines themselves, and the reasons for their existence!

Thanks. I thank Ben for letting me have his lovely notes to use as a starting point (and note that Ben in turn thanked Derek and Josh). I also thank Paula, Joao and Fiona for many useful conversations.

1.2 Introduction: a pure maths viewpoint

(I have borrowed this beautiful short essay directly from Ben's notes for comparison.)

At its heart, topology is concerned with spaces upon which it is possible to discuss/define continuous functions (a topological space), and is geared towards rigorously classifying all such spaces. A space X is a topological space if it has a topology. A topology is a collection of special subsets of X that must satisfy three requirements (see Definition 7.1) - these special subsets are usually called **open subsets**. Remarkably, in order to have a reasonable notion of continuity for functions f on a space X to a space Y, we only require the existence of a topology on X and Y.

We will determine rigorous and abstract properties of topological spaces - topological invariants - which will help us distinguish between topological spaces. Specifically, we will introduce notions of connectedness (what does it mean for a topological space to be connected?), compactness (perhaps the most important concept to pure mathematicians) and 'shape' (or more precisely, homotopy). Topologists equate any two spaces if one can be continuously mapped to the other, bijectively, and whose inverse is also continuous[1]. If no such map exists then the spaces are topologically different.

A topologist sees no difference between the boundary of a cube, and the boundary of a football (a sphere); there is a continuous bijection from one to the other whose inverse is also continuous; what does this map look like?[2] Similarly, the surface of a donut is no different to the surface of a coffee mug, from the perspective of topology (can you imagine why?).

Question: can one continuously, and bijectively map the surface of a football to that of a donut with a map whose inverse is continuous? We'll be able to rigorously prove that the answer is "NO!" by the end of the course. Your intuition should tell you that this would be impossible without tearing one or the other surface.

You can sometimes think of topological spaces (more precisely topological manifolds) as being made out of rubber, so that any bending, stretching or shrinking is allowed without changing the underlying topological structure. However the rubber is so strong that an 'infinite amount' of stretching may survive this process[3], but the following are **not** allowed: tearing of the rubber; folding so hard that two regions become merged; or squeezing so hard that you 'lose dimensions'.

The abstract tools/ideas of topology are used heavily across all subfields of mathematics. We will not have time to go into the more algebraic side of things (via homology and cohomology), however we will introduce homotopy groups and use these to distinguish between different topological spaces. By the end of the course we will also be able to prove some powerful results in different fields: e.g. the fundamental theorem of algebra, and the ham sandwich theorem[4]. One more thing we'll be able to prove by the end of the course: at any moment in time, there exist antipodal points on the surface of the earth which have the same temperature *and* pressure[5]. To give you an idea of the power of topology, see if you can prove this before reading the notes...

[1]such a map is called a *homeomorphism*

[2]Put the cube inside the sphere and think of light rays emanating from a point inside!

[3]e.g. the continuous function $f : (0, 1) \to (1, \infty)$, $f(x) = \frac{1}{x}$ continuously stretches out a bounded interval to an unbounded one: you can check that the inverse exists and is also continuous

[4]which proves that any sandwich made from bread, butter, and ham can always be sliced (with a single cut) into two parts, so that each part consists of equal quantities of the three separate ingredients

[5]We are making the assumption that temperature and pressure vary continuously in space here

Chapter 2

Preliminaries

We assume familiarity with set theory ideas from earlier, but we will review some of them here in Chapter 2 (and see also appendix Section A).

2.1 Some reminders on sets

We assume here that you are reasonably happy with the idea of a collection of "objects". This is a bit vague and potentially troublesome. But it *is* very useful, and we have to start somewhere. We will use the term 'set' for a collection of objects.

Suppose that we (you and I) both have in mind a set. Let's call it S. To say that we both have it is to say that we agree on what the "elements" are — the objects that are collected in S. Thus if we both have in mind an object x (say), we can agree if the statement 'x is in S' (written $x \in S$) is true or false (if false then we write $x \notin S$).

What might constitute a good "object"? In practice this is anything that we can agree is a good object. Just to get things started with a minimum of trouble, we can say that a set itself can be an object. Let us also say that there is one formal set, call it $\emptyset$, that does not contain any objects — thus postponing the general issue of what an object is by avoiding it. Thus the statement '$x \in \emptyset$' is false for every object x.

Putting these two ideas together, we have another set: the set containing only the set $\emptyset$.

If we have given a name to an object, like $\emptyset$, or X perhaps, then we can write the set containing only that object as $\{X\}$. The only concrete example of this that we have so far is $\{\emptyset\}$. For this at least we can say $\emptyset \in \{\emptyset\}$ and $x \notin \{\emptyset\}$ for all other objects x.

We say that two sets are equal if they contain the same elements; and otherwise they are unequal. Thus $\emptyset \neq \{\emptyset\}$.

Suppose that x and y and z represent objects, somehow agreed between us. One way of writing that x and y and z are in S (that $x, y, z \in S$) is $S = \{x, y, z, ...\}$. Another way is $S = \{y, x, z, ...\}$. If x, y, z are the only elements in S then we can write $S = \{x, y, z\}$. The extension of this notation to more (or fewer) elements can be guessed. (For the moment the question of precisely what the objects x, y, z here are remains mysterious.)

And then, using this notation, another set with un-mysterious objects is $\{\emptyset, \{\emptyset\}\}$. Notice that this is not equal as a set either to $\emptyset$ or to $\{\emptyset\}$. And notice that we can 'iterate' this construction: the set containing all the sets we have so far as elements is a new set; and now we can make another new set by adding this new set as a new element.

With such unappealing constructions of new sets, and hence new objects, we can at least delay the discussion of more interesting (but maybe not clearly defined) objects. We do now have many objects available — just by iterating the construction of adding a new set to a set of sets.

One more device before we really get started. Suppose that a and b represent objects (not even necessarily distinct). An *ordered pair*, denoted (a, b), is a set $\{\{a\}, \{a, b\}\}$.
(Caveat: this notation (a, b) can be used in other contexts as well, to represent other things. So, to be safe, if we do mean it to denote an ordered pair then we will say so explicitly.)

Because of the way we write (and talk; and think) it sometimes looks like there is order in expressions like $\{a, b\}$ already. But note from above that $\{a, b\} = \{b, a\}$ so there is not. However note that $(a, b) \neq (b, a)$ (unless $a = b$) — this is a good exercise to prove.

Some further reading:
Beginning Finite Mathematics (Schaum's Outline Series), S Lipschutz et al.
Discrete Mathematics, J K Truss.
Sets, Logic and Categories, P J Cameron (Springer).
Algebra Volume 1, P M Cohn.

2.2 Elementary set theory notations and constructions

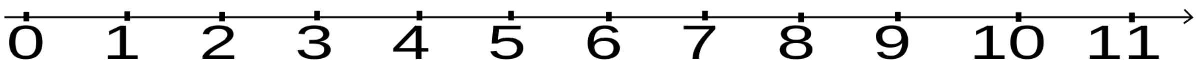
Notation: Let

$$\underline{n} \ := \ \{1, 2, .., n\}$$

Similarly here $\underline{n}' \ := \ \{1', 2', .., n'\}$, $\underline{n}'' \ := \ \{1'', 2'', .., n''\}$ and so on.

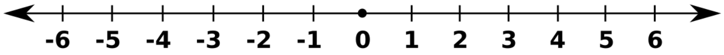
This familiar, unchallenging-seeming notation represents a huge store of knowledge and power derived from set theory: categories, cardinality and computation! The number 1 represents the class of sets in bijection with a certain 'cardinal', $\{\emptyset\}$. And so on. Thus

$$\mathbb{N}_0 : ...$$



Disjoint union yields an algebraic structure on $\mathbb{N}_0 = \{0, 1, 2, ...\}$. There is also an order structure. We order $i > j$ if there exist $a \in i$ and $b \in j$ such that $a \supset b$. The idea of lost kittens gives us negative numbers:

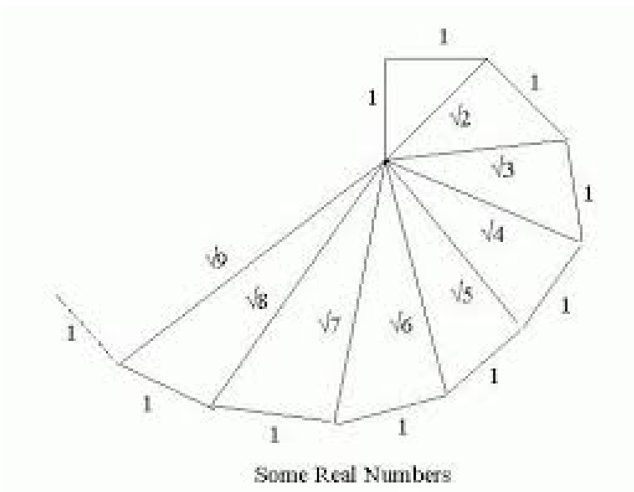
$$\mathbb{Z} : .....$$



One milk shared 3 ways gives us the idea for rational numbers.

$$\mathbb{Q} : ......$$

And we can go on. For example to real numbers. See later.



(**2.2.1**) Think about the Newton–Raphson method for finding roots of, say, $f(x) = x^2 - 2$. Our first guess can be $r_0 = 1$, say. Then the next guess is $r_1 = r_0 - \frac{f(r_0)}{f'(r_0)}$ and so on. Thus every approximation is rational. But the approximations get better and better (this is not obvious, but true). And of course $\sqrt{2}$ is not rational.

2.2.1 Power sets

(**2.2.2**) For S a set, let $\mathtt{P}(S)$ denote the *power set*, the set of subsets of S.

(We may consider $\mathtt{P}(S)$ to be partially ordered by inclusion. As such it has the structure of a lattice — see 4.2.12 below.)

Let $\mathtt{P}_n(S) \subset \mathtt{P}(S)$ be the subset of elements of order n.

(**2.2.3**) Example: The power set $\mathtt{P}(\mathbb{N})$ is a very interesting set! We can place it in correspondence with the set of all possible assignments of an element from $\{0,1\}$ to $\mathbb{N}$. Let b be such an assignment (so $b(7) = 0$ or 1, and so on: let β be the corresponding set, then $7 \in \beta$ gives $b(7) = 1$ and $7 \notin \beta$ gives $b(7) = 0$ and so on). Then we can further assign a number to each of these assignments:

$$n(b) = \sum_{i \in \mathbb{N}} \frac{b(i)}{2^i}$$

(this is an infinite sum so there is a question about convergence, but the denominator is such that we are ok here). For example if a is the assignment $a(i) = 1$ for all i then

$$n(a) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + ...$$

2.2.2 Relations

(**2.2.4**) For S, T sets, let $U_{S,T}$ denote the set of relations on S to T. That is,

$$U_{S,T} = \mathtt{P}(S \times T).$$

Let $U_S = U_{S,S}$. Even in this case we may consider the left-hand 'input' set to be distinct from the right-hand 'output' set — elements are distinguished by their position in the ordered pair.

(**2.2.5**) A relation on S to T is *'simple'* if no element of the left-hand set S appears more than once as the left-hand component of a pair.

(**2.2.6**) A relation on S to T is a *'function'* if every element of the left-hand set S appears once as the left-hand component of a pair.

(**2.2.7**) Notation: Given a relation $\rho \subseteq A \times B$ we may write $x \rho y$ as shorthand for $(x, y) \in \rho$.

(**2.2.8**) Having established ρ used as above on a set S to itself, we may write (S, ρ) for the 'relational structure'.

Example 2.1. (a) Let S be a non-empty set, and $s \in S$. Then $\rho_s = \{(s, s)\}$ is a relation on S to itself.
(b) A relation on set S to itself is given by $\rho_{id} = \{(a, b) \mid a, b \in S,\ a = b\}$.
(c) Can you give another (infinite) relation on the set $\mathbb{N}$? How about on $\mathbb{R}$? (We have assumed that we know these sets as sets, but if you want to use any structure on them you must first introduce it.)

Discuss: The axiom of choice.

Example 2.2. Let S be a set, and $\mathtt{P}(S)$ the power set. Then $(\mathtt{P}(S), \subseteq)$ gives a relation on $\mathtt{P}(S)$.

(**2.2.9**) Given a relation $\rho \subseteq A \times B$ then for each $a \in A$ we have a subset of B given by $\{b \in B | a \rho b\}$. In general the subset might be empty or contain many elements.

Notationally, given a symbol ρ say denoting a relation $\rho \subseteq A \times B$ we may write

$$x\rho- \ := \ \{b \in B \mid x \rho b\},$$

the subset of B such that $x\rho b$ for $b \in x\rho-$.

(**2.2.10**) From (2.2.6), given $\rho \subseteq A \times B$, if $x\rho-$ contains a single element for every $x \in A$ then ρ is a *function* from A to B.

In this case we write $\rho(x)$ for the single element of $x\rho-$. If $C \subseteq A$ we may write $\rho(C)$ for $\bigcup_{x \in C} x\rho-$. That is, $\rho(C) = \{b \in B | \exists x \in C \; s.t. \; x\rho b\} = \{\rho(x) | x \in C\}$.

(**2.2.11**) The *opposite* of a relation $\rho \subseteq A \times B$ is

$$\rho^\circ \; := \; \{(y, x) \mid (x, y) \in \rho\}$$

which is a relation $\rho^\circ \subseteq B \times A$.

2.2.3 Relations on a set to itself - special properties

Here we focus on relations from a set X (say) to itself, i.e. on elements $\sim$ of $U_X = \mathsf{P}(X \times X)$. (As in Example 2.1.)

(**2.2.12**) A relation $\sim$ on a set X is:
a) **reflexive:** if for all $a \in X$, $a \sim a$ (i.e. $(a, a) \in \sim$).
b) **symmetric:** if for all $a, b \in X$, if $a \sim b$, then $b \sim a$.
b) **antisymmetric:** if for all $a \neq b \in X$, if $a \sim b$, then $b \not\sim a$ (i.e. $(b, a) \notin \sim$).
c) **transitive:** if for all $a, b, c \in X$ if $a \sim b$ and $b \sim c$ then $a \sim c$.

(**2.2.13**) An **partial order** on a set X is a relation $\sim$ on X that is reflexive, antisymmetric and transitive — and then $(X, \sim)$ is called a *poset*.

(**2.2.14**) An **equivalence relation** on a set X is a relation on X that is reflexive, symmetric and transitive.

Let $\sim$ be an equivalence relation on X. Then for $x \in X$ the equivalence class of x is $[x] = \{y \in X | y \sim x\}$.

Notation: We write $\mathsf{E}(S)$ for the set of equivalence relations on a set S.

(**2.2.15**) **Definition**. Given an equivalence relation $\sim$ on set S, a transversal of $\sim$ is a subset of S containing exactly one element of each equivalence class $[s]$.

2.2.4 Functions

(**2.2.16**) For $(a, b) \in S \times T$, set $\pi_1(a, b) = a$ (the 'projection' onto the left-hand component, sometimes written π_S).
For $\rho \in U_{S,T}$ let $\mathrm{dom}(\rho) \; := \; \pi_1(\rho)$. Let

$$T^S \; = \; \hom(S, T) \; \subset \; U_{S,T}$$

be the subset of simple relations with $\mathrm{dom}(\rho) = S$, or (again from (2.2.6)) *functions*.

For example

$$\underline{2}^{\underline{2}} \; = \{\{(1,1),(2,1)\},\{(1,1),(2,2)\},\{(1,2),(2,1)\},\{(1,2),(2,2)\}\} \tag{2.1}$$

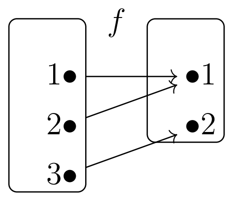
(**2.2.17**) For I a set and S_i a non-empty set for each $i \in I$, a choice function is a function that takes an element $i \in I$ as input and returns an element from S_i. (The axiom of choice says that such a function exists for any such family of sets.) Then $\prod_{i \in I} S_i$ is the set of all such functions.

2.2.5 Composition of functions

(**2.2.18**) It will be useful to have in mind the *mapping diagram* realisation of finite functions such as in (2.1). For example

$$f = \{(1,1),(2,1),(3,2)\} \in \underline{2}^{\underline{3}}$$

is



(**2.2.19**) If T, S finite it will be clear that any total order on each of T and S puts T^S in bijection with $\underline{|T|}^{\underline{|S|}}$. We may represent the elements of T^S as S-ordered lists of elements from T. Thus

$$\underline{2}^{\underline{2}} = \{11, 12, 21, 22\}, \qquad \underline{2}^{\underline{3}} = \{111, 112, 121, 122, 211, 212, 221, 222\}$$

(for example $22(1) = 2$, since the first entry in 22 is the image of 1).

(**2.2.20**) Composition of functions defines a map

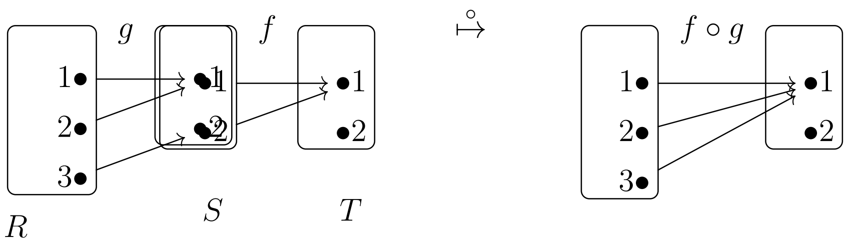
$$\hom(S,T) \times \hom(R,S) \rightarrow \hom(R,T) \tag{2.2}$$

$$(f,g) \mapsto f \circ g \tag{2.3}$$

where as usual $(f \circ g)(x) = f(g(x))$.

For example (in the notation of (2.2.19)) $11 \circ 22 = 11$ (since $11(22(1)) = 11(2) = 1$; and so on).

The *mapping diagram realisation* of composition is to first juxtapose the two functions so that the two instances of the set S coincide, then define a direct path from R to T for each path of length 2 so formed:



(2.4)

(**2.2.21**) Observe that $(\hom(S,S), \circ)$ has an identity. Also some elements are invertible. If f is invertible in $(\hom(S,S), \circ)$ we write f^{-1} for the inverse.

(**2.2.22**) Even if f is not invertible, for $B \subseteq S$ we may define $f^{-1}(B) = \{x \in S | f(x) \in B\}$.

(**2.2.23**) Let $f \in \hom(S,T)$. We also define $f^{-}(S) = f_{-}^{-1}(S) = \{f^{-1}(\{t\}) | t \in T\}$ (note the argument).

Note that this gives a partition of S.

Example: Consider f and g from (2.4). Here $g^{-1}(1) = \{1,2\}$ so $g^{-}(R) = \{\{1,2\},\{3\}\}$.

(**2.2.24**) If the image $f(S)$ of a map $f : S \to T$ is of finite order we shall say that f has order $|f(S)|$ (otherwise it has infinite order).

For $R \xrightarrow{f} S \xrightarrow{g} T$ (finite) we have the *bottleneck principle*

$$|(g \circ f)(R)| \;\leq\; \min(|g(S)|, |f(R)|)$$

To see this note that evidently $g(S) \supseteq g \circ f(R)$, from which the first inequality follows; meanwhile $|f_{-}^{-1}(R)| = |f(R)|$ for any $f \in \hom(R, -)$, leading to the second inequality.

Recall that a set with a closed binary operation is called a *magma*. If the operation is associative this structure is called a *semigroup*. If the operation also has an identity element then the structure is a *monoid*. And if every element has an inverse then it is a *group*.

A (two-sided) *ideal* in a monoid is a subset that is closed under the action of the monoid from both sides.

(**2.2.25**) PROPOSITION. *(i) For S a set, $S^S = \hom(S, S)$ is a monoid under composition of functions.*
(ii) For each $d \in \mathbb{N}$ then set $\hom^d(S, S) \;:= \{f \in S^S \mid |f(S)| < d\}$ is an ideal (hence a sub-semigroup) of S^S.

Proof. (i) Hint: $(f \circ (g \circ h))(x) = f(g(h(x))) = ((f \circ g) \circ h)(x)$
Exercise: explain this argument in terms of mapping diagrams.
(ii) Consider $g \circ f$, say, in $\hom(S, S)$. Evidently $g(S) \supseteq g \circ f(S)$. Thus $\hom^d(S, S) \circ f \subseteq \hom^d(S, S)$ for all f. Meanwhile $f(s) = f(t)$ implies $g \circ f(s) = g \circ f(t)$ so the partition $p = f^{-1}(S)$ of S implied by f cannot be refined in passing to the partition implied by $g \circ f$. Of course $|f_{-}^{-1}(S)| = |f(S)|$ for any f. Thus $g \circ \hom^d(S, S) \subseteq \hom^d(S, S)$ for all g. □

(**2.2.26**) A *composition* of n (not to be confused with composition of functions) is a finite sequence $\lambda = (\lambda_1, \lambda_2, ...)$ in $\mathbb{N}_0$ that sums to n. We write $\lambda \vDash n$.

We define the *shape* of an element f of $\underline{m}^{\underline{n}}$ as the composition of n given by

$$\lambda(f)_i = |f^{-1}(\{i\})|$$

Example: for $111432525 \in \underline{6}^{\underline{9}}$ we have $\lambda(111432525) = (3, 2, 1, 1, 2, 0)$. In particular $f^{-1}(\{1\}) = \{1, 2, 3\}$ so $\lambda(f)_1 = 3$.

If $\lambda \vDash n$ we write $|\lambda| = n$.

2.3 Setting the scene

2.3.1 'Intuitive' continuity

How can we tell if a function $f : A \to B$ is continuous?

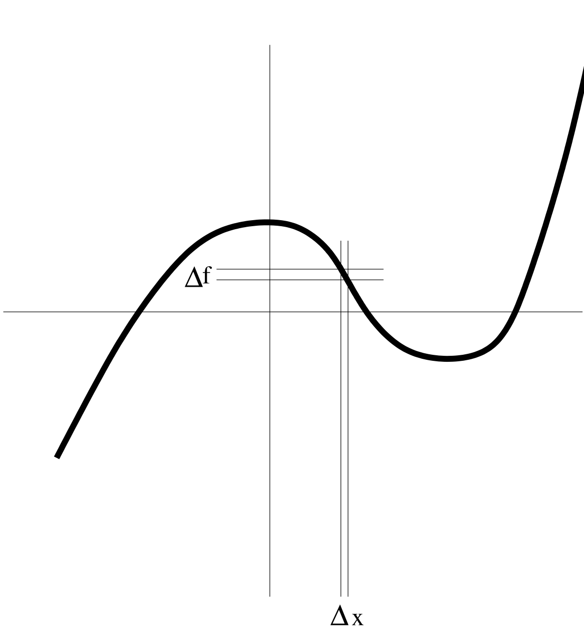
"Compare $f(x)$ with $f(x+\Delta x)$. As Δx gets small, $\Delta f = f(x+\Delta x) - f(x)$ gets small."

Several aspects of this heuristic will benefit from closer inspection if we want to use on $f : A \to B$ generally, say.

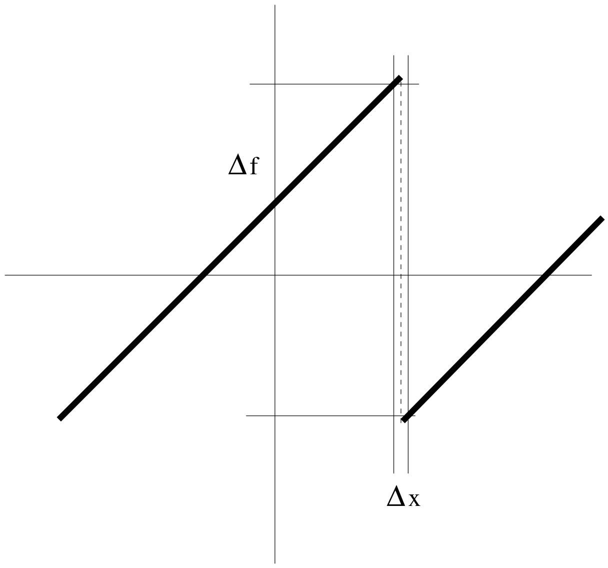
1. $x \in A$, and maybe $\Delta x \in A$ too, but what is $x + \Delta x$?
2. What is 'gets small' in A?
3. What is $f(x+\Delta x) - f(x)$ in B?
4. What is 'gets small' in B?

The familiar setting for examples is $f : \mathbb{R} \to \mathbb{R}$. Here all the questions have answers, and we can picture specific examples.

Continuous:



Not continuous:



We can also interpret Δx as giving a small 'ball' around x in our domain (in this case a 1d ball, so just an interval $(x-\delta, x+\delta)$; but if the domain was $\mathbb{R}^3$ say, it would be a real ball) — lets call it $B_\delta(x)$ (it is small if δ is small). We then ask about the size of ball that we would need in the codomain to contain $f(B_\delta(x))$. For continuity we are saying that it should be possible to make this containing ball 'correspondingly' small.

For some reason (habit) analysts always call the measure of smallness on the codomain side ϵ, along with using δ on the domain side.

More formally we say that
Function $f : \mathbb{R} \to \mathbb{R}$ is 'continuous at $x \in \mathbb{R}$' iff for all $\epsilon > 0$ there exists $\delta > 0$ such that

(I) if $y \in \mathbb{R}$ is such that $|x-y| < \delta$, then $|f(x) - f(y)| < \epsilon$.

(II) $d(x,y) < \delta$ implies $d(f(x), f(y)) < \epsilon$. (using $d(x,y) = |x-y|$)

(III) $f(B_\delta(x)) \subseteq B_\epsilon(f(x))$.

(IV) $B_\delta(x) \subseteq f^{-1}(B_\epsilon(f(x)))$.
At the last re-write here we have used one of the set theory identities, applying f^{-1} to both sides.

(**2.3.1**) **Definition**. A function $f : (X,d) \to (Y,d')$ between metric spaces (as in §6) is **metric-continuous** at $x \in X$ iff for all $\epsilon > 0$ there exists $\delta > 0$ such that

$$B_\delta(x) \subseteq f^{-1}(B'_\epsilon(f(x))).$$

(here B' flags that we are using the metric d' — later we'll just write B).

We say that $f : X \to Y$ is metric-continuous if and only if it is metric-continuous at x for all $x \in X$.

(**2.3.2**) N.B.: This is the usual definition of continuity when X, Y are the vector spaces $\mathbb{R}^n, \mathbb{R}^m$ equipped with the standard metrics.

(**2.3.3**) Let (X,d) be a metric space (as in §6). Define τ_d as the set of 'd-open' subsets U of X: subsets such that $x \in U$ implies there exists $\epsilon > 0$ such that $B_\epsilon(x) \subseteq U$.

(**2.3.4**) Example: the set of finite unions of open intervals of $\mathbb{R}$ has this property. It is an expression of the weird feature of open intervals that (in a certain sense) they do not have a boundary. Specifically if we think of the subset of positive real numbers, there is no smallest one. We can say that 0 is a sharp boundary of the subset... but it is not in the subset; and for every small ϵ we think of, close to zero, there is of course one that is even smaller.

Theorem 2.3. *A function $f : (X, d) \to (Y, d')$ is metric-continuous if and only if for all d'-open $U \subset Y$, $f^{-1}(U) \subset X$ is d-open.*

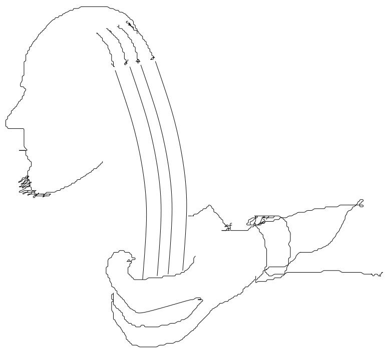
Proof. Suppose that for all open $U \subset Y$, $f^{-1}(U) \subset X$ is open. Then for all $\epsilon > 0$ we know that $f^{-1}(B_\epsilon(f(x))) \subseteq X$ is open, since $B_\epsilon(f(x))$ is open in Y. By definition (since $x \in f^{-1}(B_\epsilon(f(x))))$ this implies there exists $\delta > 0$ so that $B_\delta(x) \subseteq f^{-1}(B_\epsilon(f(x)))$, thus f is metric-continuous.

For the converse suppose that f is metric continuous. Let $U \subseteq Y$ be open so we want to prove that $f^{-1}(U)$ is open. Pick $x \in f^{-1}(U)$ giving $f(x) \in U$. Since U is open, there exists $\epsilon > 0$ so that $B_\epsilon(f(x)) \subseteq U$ and since f is metric continuous, there exists $\delta > 0$ so that $B_\delta(x) \subseteq f^{-1}(B_\epsilon(f(x))) \subseteq f^{-1}(U)$ so we are done (remember — or see later — that, since X is a metric space, $f^{-1}(U)$ is open if and only if, for all $x \in f^{-1}(U)$ there exists $\delta > 0$ so that $B_\delta(x) \subseteq f^{-1}(U)$). □

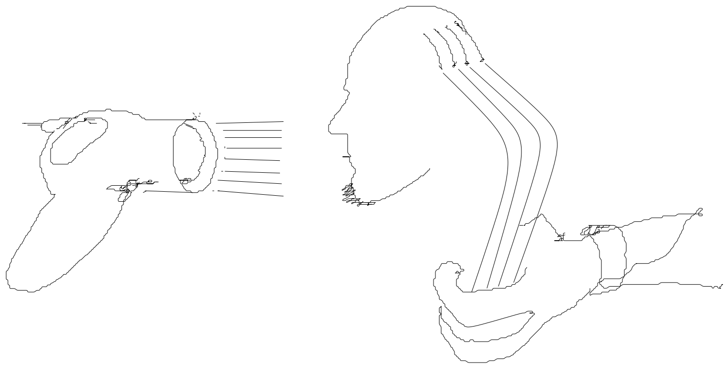
What have we done here?! It is not quite clear just yet, but in fact we have explained why that definition of topology that we started with works!! To really see this though... keep reading.

2.3.2 And why is it useful?...

... And why is it useful? Physics, Biology, Chemistry, Computing, Sociology, etc. ... See later for more on this.



Meet Tony. He is a hairdresser with a lovely ponytail. He is interested in organising the different configurations into which he can get his ponytail. We'll come back to him shortly.



Chapter 3

Topology: first pass

Next we give a relatively brief overview of topics to be addressed in these notes. Then more details (examples and so on) will be given later.

Some people say that maths organises thought by introducing and codifying structures along three main lines. In this view, the lines are:
∘ ordered structures (less than)
∘ algebraic structures (times)
∘ topological structures (near).

In this view, a 'structure' is a pair consisting of a set S and a collection of sets built on S. For example a *group* is an algebraic structure and is a set together with a closed binary operation. Taken together with our Introduction above, these remarks set us up in two ways. Firstly they indicate how topology fits into organisational thought, and secondly they tell us that we can characterise it as a 'structure'. Now read on.

3.1 Overview

See e.g. Armstrong [Arm79], Mendelson [Men62], Hartshorne [Har77].

(**3.1.1**) A *sigma-algebra* over a set S is a subset Σ of the power set $\mathtt{P}(S)$ which includes S and $\emptyset$ and is closed under countable unions, and complementation in S.

Any subset S' of $\mathtt{P}(S)$ defines a sigma-algebra — the smallest sigma-algebra generated by S'. For example $\{\{1\}\} \subset \mathtt{P}(\{1,2,3\})$ generates $\Sigma = \{\emptyset, \{1\}, \{2,3\}, \{1,2,3\}\}$.

(**3.1.2**) A *topological space* is a set S together with a subset T of the power set $\mathtt{P}(S)$ which includes S and $\emptyset$ and is closed under unions and finite intersections.

The set T is called a *topology* on S. The elements of T are called the *open sets* of this topology. A set is *closed* if it is the complement in S of an open set.

(**3.1.3**) Example. /Proposition. Consider the set $\mathbb{R}^n$ — say $\mathbb{R}^2$ to be specific — together with the set τ_d of subsets 'generated' by unions and finite intersections of the set of open balls

$$B_\epsilon(x) = \{y \in \mathbb{R}^n \ : \ d(x,y) < \epsilon\}$$

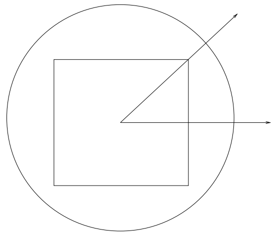
where d is the usual distance function or 'metric'. This is a topological space. (We will prove this in §3.1.3.)
The topology τ_d is called the metric topology with respect to metric d (again, see later for more examples). Indeed noting that our definition of $B_\epsilon(x)$ and hence τ_d works for any metric/distance-function d, we will see later (e.g. in §6.1) that every such τ_d is a topology. (And, more to the point, that some of them give rise to very useful notions of continuity.)

In particular this makes the set of $n \times n$ real matrices $M_n(\mathbb{R})$ a topological space — as a topological space it is $\mathbb{R}^{n^2}$. The subgroup $GL_n(\mathbb{R})$ of invertible matrices may be considered as a topological space by restriction. Note that $GL_n(\mathbb{R})$ is open and not closed (its complement is not open) in $M_n(\mathbb{R})$, but it is open and closed in the restricted topology.

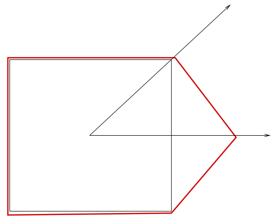
(**3.1.4**) A function between topological spaces is *continuous* if the inverse image of every open set is open.

(**3.1.5**) Two spaces are *homeomorphic* if there is a bijection between them, continuous in both directions.

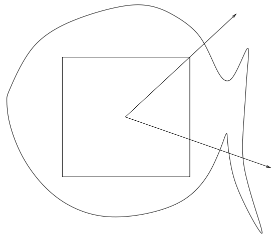
(**3.1.6**) Schematic examples using geometry to make a homeomorphism: In each case a light-ray from the origin places a point in the inner object in correspondence with a point in the outer object. We map a square to a circle...



And here a square to an irregular pentagon:

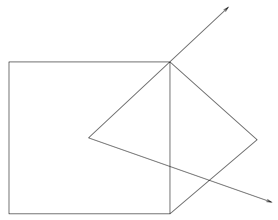


(**3.1.7**) Non-example showing that the great power of geometry requires responsible handling (the 'fish-edge' and box-edge spaces are homeomorphic, but not by any construction indicated by this figure):

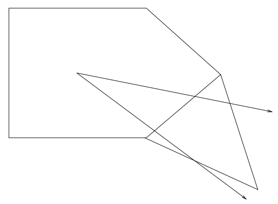


Geometry has a notion of convexity, and the fish shape is not convex.

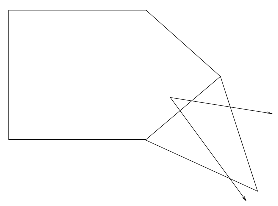
We can still use geometric constructions, but we just need to be a bit more patient. Consider the following:



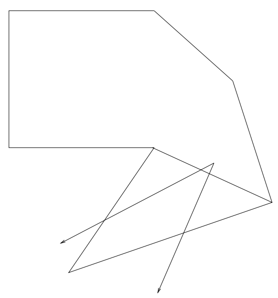
Following the light rays maps us from the square to the irregular pentagon. And we can continue to the irregular hexagon without (quite) hitting the fish-fin problem:



But we could also define a map by saying that it is the identity map everywhere except on the edge that we 'extrude', and moving the source of the 'light-ray':



Then we can continue, for example to this significantly non-convex irregular heptagon:



(As we will check later) The composition of homeomorphisms is a homeomorphism. So we have shown that all these simple closed polygonal shapes are homeomorphic.

(Tony's hair is not made of polygons... Or is it?!)

3.1.1 The world of topological spaces

(**3.1.8**) We write $\mathsf{Top}((X,\tau),(X',\tau'))$ (or just $\mathsf{Top}(X,X')$) for the set of continuous functions from topological space (X,τ) to topological space (X',τ').

Note that each $\mathsf{Top}(X,X') \subseteq hom(X,X')$. In this sense composition of functions restricts to a composition $\mathsf{Top}(Y,Z) \times \mathsf{Top}(X,Y) \to hom(X,Z)$.
Prop. The image is contained in $\mathsf{Top}(X,Z)$, so composition of functions gives a composition

$$\mathsf{Top}(Y,Z) \times \mathsf{Top}(X,Y) \to \mathsf{Top}(X,Z)$$

(**3.1.9**) The homeomorphisms are a subset in $\mathsf{Top}((X,\tau),(X',\tau'))$ (possibly empty). We can write it as $\mathsf{Top}^h((X,\tau),(X',\tau'))$.

(**3.1.10**) It will be convenient sometimes to use the notation Top for the class of all topological spaces — just so we can write $(X,\tau) \in \text{Top}$ as a quick way to say that (X,τ) is a topological space. The class Top is rather large. For X a set, we can write Top_X for the subset of Top of topological spaces whose underlying set is X.

3.1.2 Subspace topology

(**3.1.11**) Given a topological space (S,T), the restriction of T to $S' \subset S$

$$T' = \{t \cap S' | t \in T\}$$

is a topology on S'. (We will check this later or, better, you can check it as an Exercise.) The topology on the subspace is called the *subspace topology*.

Example. We have mentioned the (Euclidean distance) metric topology τ_d on $\mathbb{R}^2$. A subset such as $[0,1]^2 \subset \mathbb{R}^2$ inherits a topology from τ_d on $\mathbb{R}^2$.
It is interesting to note though that $[0,1]^2$ is closed, and not open, in $\mathbb{R}^2$; but it is both closed and open in its own subspace topology.

A subset S' of a topological space (S,T) is *irreducible* if $S' = S_1 \cup S_2$ with S_1 closed implies S_2 not closed.

3.1.3 Set theory aspects

There are several different useful formulations of the definition of topological space. We can set them up with some 'set arithmetic'.

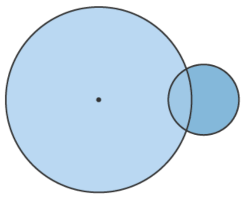
(**3.1.12**) The 'universe' of a set of sets β is the union of all the sets. Denote it by U_β.

For example, if $\beta = \{\{a, b\}, \{1, 2, 3\}\}$ then $\mathsf{U}_\beta = \{a, b, 1, 2, 3\}$.

Just pushing the notation around we have, for any set of sets β, that $\beta \in \mathsf{P}(\mathsf{P}(\mathsf{U}_\beta))$, which is the same as to say $\beta \subseteq \mathsf{P}(\mathsf{U}_\beta)$. So a topology is a pair of form $(\mathsf{U}_\beta, \beta)$, where β must satisfy some further conditions.

(**3.1.13**) A set β of sets is
finite-intersection-closed if $A, B \in \beta$ implies $A \cap B \in \beta$.
intersection-closed if closed under all intersections.
finite-union-closed if $A, B \in \beta$ implies $A \cup B \in \beta$.
union-closed if closed under all unions.
weak-intersection-closed (WIC) if for $A, B \in \beta$ then $x \in A \cap B$ implies $\exists\ C \in \beta$ such that $C \ni x$, $C \subseteq A \cap B$.

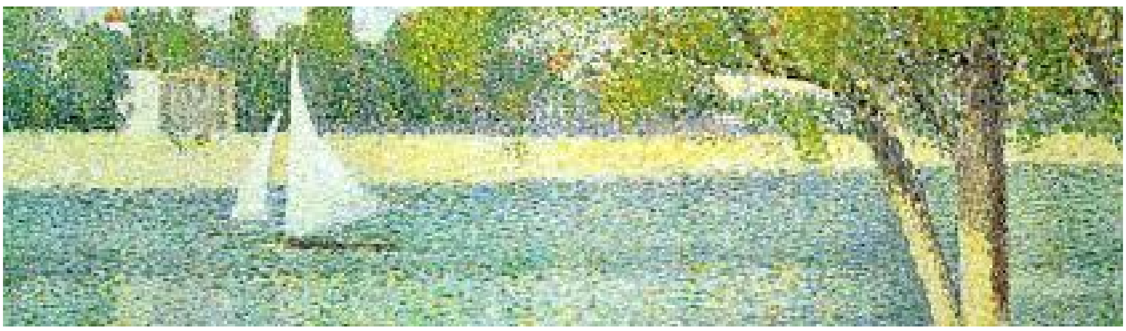
(**3.1.14**) Examples:
1. The set of open intervals in $\mathbb{R}$ is finite-intersection-closed, and so also WIC, but not intersection-closed. If two open intervals intersect then they intersect in an open interval, not in a closed one and not in a point.
2. The set of open balls in $\mathbb{R}^2$ is WIC but not intersection-closed. Consider the two open balls, $B_r(x)$ and $B_{r'}(x')$ say, here:



Call them A, B. The intersection is a 'lens' shape, not a ball. But can you see that, because of the open-ness, every point in $A \cap B$ has a (possibly very small, but finite) ball around it that is entirely inside $A \cap B$?

Can you describe an infinite union of open balls that coindices with our 'lens' $A \cap B$? (There is no finite union of balls that does so, although we can get arbitrarily close! Your computer screen resolution probably demonstrates that we can get close enough to fool the eye.)

Hint: no point in $A \cap B$ is 'on the edge', so each point $x \in A \cap B$ has on open ball around it all of which is in $A \cap B$. There is even a largest radius r_x so that the open ball $B_{r_x}(x) \subseteq A \cap B$. Now consider the union of these finite balls over all x.



(**3.1.15**) The *union-closure* of a set of sets β is the set of all unions. (This includes the unions

over all indexed subsets of β, so for example it includes the 'union' of a single element of β, and even no elements of β, as well as the unions over countable sets of elements and unions over uncountable sets of elements.)
Denote the union-closure of β by $\mathsf{C}_\cup\beta$. (See also (7.1).)
Example: Consider the set $S^{(1)}$ of single-element subsets of a set S. Then

$$\mathsf{C}_\cup S^{(1)} = \mathsf{P}(S).$$

The finite-union-closure, which we can denote by $\mathsf{C}^f_\cup\beta$, is the smallest subset of $\mathsf{P}(\mathsf{U}_\beta)$ that contains β and is closed under pairwise union.
It is the same as $\mathsf{C}_\cup\beta$ if β is finite, but can be much smaller in general.
Exercise: What can you say about $\mathsf{C}^f_\cup\mathbb{R}^{(1)}$? (Hint: Suppose that L is a finite subset of $\mathbb{N}$ in $\mathsf{C}^f_\cup\mathbb{R}^{(1)}$. Since it is finite, there is an element, n_L say, of $\mathbb{N}$ that is not in it. But then $L \cup \{n_L\}$ is larger than L, and in $\mathsf{C}^f_\cup\mathbb{R}^{(1)}$. Thus $\mathbb{N}$ is in $\mathsf{C}^f_\cup\mathbb{R}^{(1)}$?)

(**3.1.16**) (I) The union-closure of a set of sets β is a topology iff β is WIC.
(II) If β is WIC, then $(\mathsf{U}_\beta, \mathsf{C}_\cup\beta)$ is a topological space.

(**3.1.17**) Exercise: prove it.
Hints — (I) if part: Here we assume that β is WIC and try to verify the topology axioms. In particular one topology axiom to show is the finite intersection closure (observe that the other axioms follow more or less by construction here). If β is WIC then first consider $A \cap B$ for $A, B \in \beta$ (later we will also need to consider $A, B \in \mathsf{C}_\cup\beta$). We have that $x \in A \cap B$ implies that there exists $C \in \beta$ with $x \in C$ and $C \subseteq A \cap B$. Now for each such x let C_x be a specific choice of such a $C \in \beta$ (using the axiom of choice) — it will not matter which one. Now consider

$$\bigcup_{x \in A\cap B} C_x \quad \in \; \mathsf{C}_\cup\beta.$$

Note that this contains $A \cap B$ by construction; but also that it is a union of subsets of $A \cap B$ so it is contained in $A \cap B$, so it is $A \cap B$. In other words we have shown that $A \cap B$ lies in the union closure of β.
But now suppose V and W are in the union closure and consider $V \cap W$. For $y \in V \cap W$ then V is a union from β so there is a subset V_y of V that contains y and is in β; and similarly there is a subset W_y of W that contains y and is in β. Thus $y \in V_y \cap W_y$. But we have already shown that $V_y \cap W_y$ lies in the union closure of β. Thus $\cup_{y \in V \cap W} V_y \cap W_y$ also lies in the union closure. This is $V \cap W$ by the two-way inclusion argument. So we are done.

(**3.1.18**) A WIC set of sets β is called a **basis** for a topology on U_β. The topology is that 'generated' by β.

(**3.1.19**) The metric topology on a metric space (X, d) (see e.g. §6) is generated by the set $\{B_r(x) \mid x \in X, r > 0\}$ of open balls.
Exercise: Prove it!

(**3.1.20**) A topological space is 'second countable' if it can be generated by a countable basis.

(**3.1.21**) Example: $\mathbb{R}^n$ (with Euclidean metric topology) is second countable. The set of all balls of rational radius centered at the points in $\mathbb{Q}^n$ is a basis.
Exercise: Prove it!

3.1.4 Neighbourhoods (and an aside on smallest neighbourhoods)

Fix $(X, \tau) \in \text{Top}$. Then for $x \in X$ the set of neighbourhoods of x is

$$\mathcal{V}_x := \{N \in \tau \mid x \in N\}$$

(**3.1.22**) Note that if (X, τ) is finite then for each $x \in X$ there is a unique smallest $N \in \mathcal{V}_x$. We denote it N_x. We can give (X, τ) by giving the function

$$N_- : X \to \tau$$

$x \mapsto N_x$ — say by giving the list of N_xs in some natural order.

(**3.1.23**) For example for $X = \{a, b\}$ the topologies are given by:

$$(ab, ab), (a, ab), (ab, b), (a, b)$$

— here we streamline by writing just ab for $\{a, b\}$ and so on. Note that the middle two are homeomorphic, so there are three homeomorphism classes altogether. (The class of the middle two is sometimes called the *Sierpinski class* of topologies.)

See §3.6 for some exercises in this direction.

(**3.1.24**) Note that the N_- function is not defined in general. There is no smallest neighbourhood of a point $x \in \mathbb{R}$ for example.
We will come back to this shortly.

3.1.5 Limit-closure of a subset of a space

(**3.1.25**) Fix $(X, \tau) \in \text{Top}$. Then for $A \subset X$, the set of limit points (or 'limit-closure') of A in X is

$$\overline{A} := \{x \in X \mid N \cap A \neq \emptyset \ \forall \ N \in \mathcal{V}_x\}$$

(**3.1.26**) E.g. consider $\mathbb{I} = ([0, 1], \tau_d)$ and let $A = [0, 1/2) \subset \mathbb{I}$. Then $\overline{A} = [0, 1/2]$.

(**3.1.27**) A space (X, τ) is *connected* if $A \cup B = X$ implies that either $\overline{A} \cap B \neq \emptyset$ or $A \cap \overline{B} \neq \emptyset$.

(**3.1.28**) Caveat. Just for convenience we assume here that you know about the construction of $\mathbb{R}$! See later for details.

(**3.1.29**) E.g. for $S \subset \mathbb{R}$ recall that $\sup S \in \mathbb{R}$ is the least upper bound — which exists *by construction of* $\mathbb{R}$.
(Whereas for example the subset $\{x \in \mathbb{Q} | \ x < \sqrt{2}\}$ has no l.u.b. in $\mathbb{Q}$.)

Then

$$\sup S \in \overline{S} \tag{3.1}$$

(**3.1.30**) Theorem. $\mathbb{R}$ is connected.

Proof. Suppose

$$\mathbb{R} = A \cup B$$

is a proper partition. Without loss of generality we can have $a \in A$ and $b \in B$ with $a < b$. Let

$$X = \{x \in A \mid x < b\}.$$

Let l be the least upper bound of X in $\mathbb{R}$ (which exists — a special property of $\mathbb{R}$! — see also later):

$$l = \sup X$$

This could be b or less than b. (Note, b is in B not A, but l must be in the closure of A, not *necessarily* in A itself.)
If $l = b$ then $\overline{A} \cap B \ni b$.
If $l < b$ then $(l, b) \subset B$ so $l \in \overline{B}$ so $l \in A \cap \overline{B}$. □

(**3.1.31**) An *interval* in $\mathbb{R}$ is a subset J such that $a < b \in J$ implies $(a, b) \subset J$.

(**3.1.32**) Theorem. Every interval in $\mathbb{R}$ is connected.

Proof. We can use essentially the same argument as for $\mathbb{R}$, considering a proper partition of interval $J = A_J \cup B_J$, choosing $a \in A_J$ and $b \in B_J$ with $a < b$, defining $X = \{x \in A_J \mid x < b\}$; $l = \sup X$, with $l \in J \subset \mathbb{R}$ by the interval property, and so on. We can extend $J = A_J \cup B_J$ to a partition $\mathbb{R} = A \cup B$ as above, such that $A_J = A \cap J$ and $B_J = B \cap J$. The argument then proceeds roughly as before. □

3.2 Basic ‘low-dimensional’ topology

Here $\mathbb{R}$ has the Euclidean metric topology and the interval $[0,1]$ the corresponding subset topology.

‘Low-dimensional’ refers to aspects of topology narrowed from general topology in a couple of ways that are ‘important’ in a humanistic sense. Perhaps the most notable of these is the involvement of $[0,1]$ in constructions. The claim is that the space $[0,1]$ captures some crucial part of human experience.

In Chapter 5 we look a little at what makes the space $\mathbb{I} = [0,1]$ special from a mathematical point of view. One thing would be that there are lots of interesting functions in $\mathsf{Top}(\mathbb{I},\mathbb{I})$, and even in $\mathsf{Top}^h(\mathbb{I},\mathbb{I})$. For example

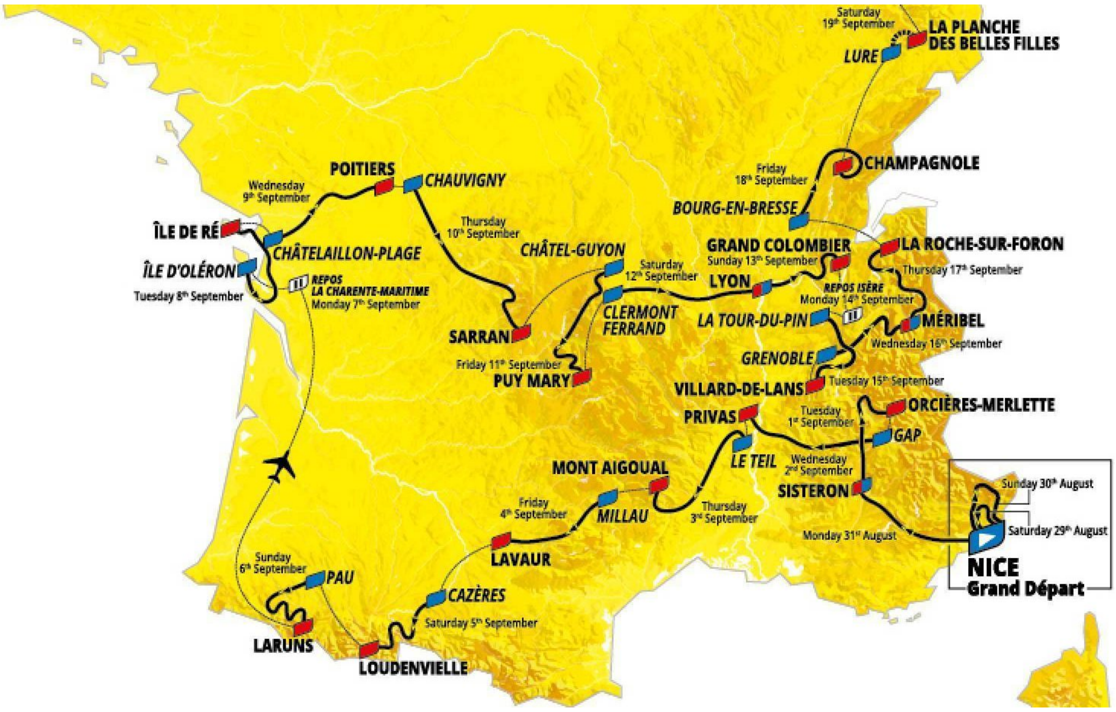
$$f_{flip}(s) = 1 - s \tag{3.2}$$

This is familiar. But note that it is a very sophisticated way of giving a function — using algebra. Also it uses symmetry or self-similarity properties of $\mathbb{I}$ — what we might call a ‘flip’ symmetry. (We will return to this many times. See for example §9.3.)

Before we get into low-dimensional topology *per se*, we set the scene with a little geometry and algebra.

3.2.1 Preliminaries for paths, pictures and plumbing

Let us introduce some notation for journeys and chaining journey stages together. Such chains of journeys are common. But they contain a lot of topological ideas.



For example, informally we could partition the set of all journeys according to where they start. And end. Let us introduce some general notation for this.

(**3.2.1**) A function $G \xrightarrow{s} X \in hom(G,X)$ induces a partition

$$G = \cup_{x \in X} G(x)$$

where

$$G(x) \;=\; s^{-1}(x) \;=\; \{g \in G \mid s(g) = x\}$$

A pair of functions $G \overset{s}{\underset{t}{\rightrightarrows}} X$ induces

$$G = \cup_{x,y \in X} G(x,y)$$

similarly.

(**3.2.2**) Remark. We can partition the stages of the *Tour de France*, or the journeys of the individual cyclists, according to where they begin and end.
We can partition the class of all functions according to the domain and codomain.
We can partition the set of all real matrices according to the number of rows and the number of columns.

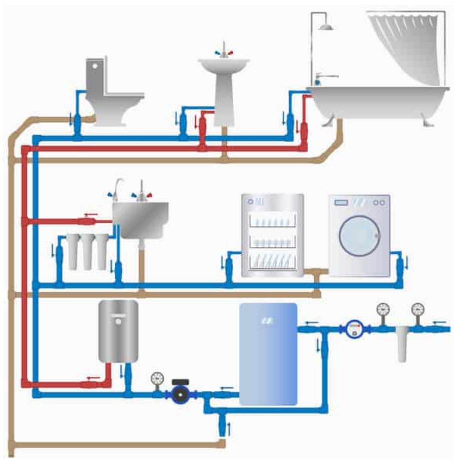
(**3.2.3**) A pair of functions $G \overset{s}{\underset{t}{\rightrightarrows}} X$ is sometimes called a *graph*. The set G is the set of 'edges' and X is the set of 'vertices'. So here $g \in G$ is seen as the label for an edge that passes from vertex $s(g)$ (the start of g) to vertex $t(g)$ (the terminus of g).

(**3.2.4**) A *magmoid* is a pair $G \overset{s}{\underset{t}{\rightrightarrows}} X$ with, for each triple $x, y, z \in X$, a function

$$* : G(x,y) \times G(y,z) \to G(x,z)$$

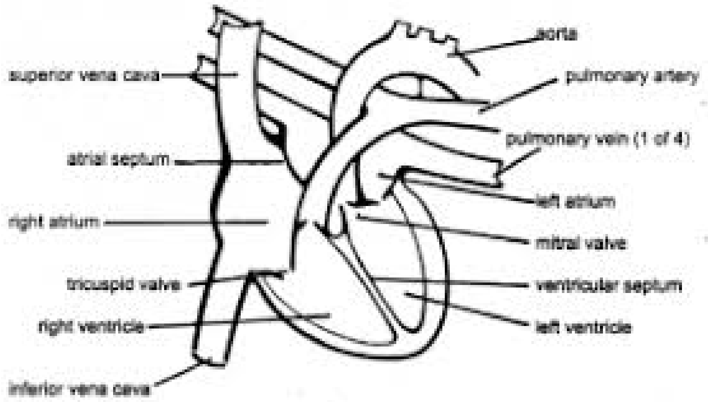
(**3.2.5**) Remark. Note that if X has only a single element, x say, then we only have $G(x,x)$, with a closed binary operation (so an example of a magmoid would be any group).

(**3.2.6**) The requirement for fitting two journeys together was that one began where the other ended. For other things that can in principle be fitted together, the condition might be different.



(**3.2.7**) If we have two lengths of 1cm bore copper pipe, we can weld them together to get a longer length. If we include the pipes 'of length zero' then this operation even has an identity element. So it is almost a group. It is not a group because (like $(\mathbb{R}_{\geq 0}, +)$) it does not have inverses. It is a 'plumbing monoid'.



Now suppose we buy some T-pieces and some pipe-caps for our plumbing installation. What happens to our plumbing monoid now? (There are several good answers to this question.)

(**3.2.8**) A *category* is a magmoid $(G \overset{s}{\underset{t}{\rightrightarrows}} X , *)$ with associative composition (so that each $(G(x,x),*)$ is a semigroup); and an identity element in each $(G(x,x),*)$ (so that each $(G(x,x),*)$ is actually a monoid). Furthermore the identity in each $(G(x,x),*)$, denoted id_x say, acts trivially on the left on each $(G(x,y),*)$; and on the right on each $(G(y,x),*)$.

(**3.2.9**) Example: Let Mat denote the set of all real matrices. Let $s : Mat \to \mathbb{N}$ and $t : Mat \to \mathbb{N}$ be the functions giving the number of rows and columns respectively. Then matrix multiplication makes Mat into a magmoid, and because it is associative, with 'units', this magmoid is a category.

In particular here

$$id_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(**3.2.10**) Remark. Foundational mathematics allows the underlying 'set' G for a category to be a class rather than a set (to bypass Russell's paradox). Thus G could be the class of all functions and X the class of all sets, and s and t the maps to domain and codomain respectively. Then hom from (2.2.20) gives an example of a category.

In general the class X is called the class of objects; and the class G is called the class of morphisms or arrows.

3.2.2 Briefly down the rabbit hole (groupoids)

(**3.2.11**) Since 'category' is a mathematical structure, we should (strictly speaking) next say what maps between categories look like — the category equivalent of group homomorphism. For now let us just note that category homomorphisms are called 'functors'; and (in a moment) give examples. A functor from category $(G \overset{s}{\underset{t}{\rightrightarrows}} X , *, id)$ to category $(H \overset{s'}{\underset{t'}{\rightrightarrows}} Y , \circ, I)$ (say) consists of a map $F : X \to Y$ and a map $F : G \to H$ subject to conditions that make the maps 'commute' with the source and target maps.

A subcategory of a category $(G \overset{s}{\underset{t}{\rightrightarrows}} X , *, id)$ is a subclass X' of X and a subclass G' of G that is defined on X' and is closed under $*$ and is a category with the id obtained merely by suitably restricting the original id.

A subcategory is called 'wide' if $X' = X$; and called 'full' if for every pair $x, y \in X'$ then $G'(x, y) = G(x, y)$.

Example: let $\hom^i$ denote the wide subcategory of hom (from (2.2.20)) consisting of bijections.

Example: let Mat^i denote the wide subcategory of Mat of invertible matrices.

(**3.2.12**) A *groupoid* is a category in which each $g \in G$ has an inverse.

(**3.2.13**) Example: $\hom^i$ is a groupoid.

(**3.2.14**) Recall that given a normal subgroup of a group G we may form the quotient group.

More generally, an equivalence relation on the underlying set of an algebraic structure is called a *congruence* if the class $[a * b] = [a' * b']$ whenever $a' \in [a]$ and $b' \in [b]$.

In fact there are so many parallels between constructions for categories and constructions for groups that it is worth having a little review of constructions for groups.

For example, for each group $(G, .)$ there is an 'opposite' group, with the same set but $a \circ b = b.a$. The identity set map gives an 'antihomomorphism' of a group to its opposite. Meanwhile the bijective set map given by $g \mapsto g^{-1}$ is an antihomomorphism of G to itself. ...

(**3.2.15**) The notion of opposite lifts to categories in the 'obvious' way.
('Obvious' in inverted commas because there really is only one way that makes any sense; but it is neither trivial nor un-eventful to pass through the looking glass in this way.)

(**3.2.16**) The notion of antihomomorphism lifts to functors — which variants are then called *contravariant* functors. For example in $\hom^i$ the map $f \mapsto f^{-1}$ gives a contravariant functor:

$$(-)^{-1} : \hom^i \to \hom^i$$

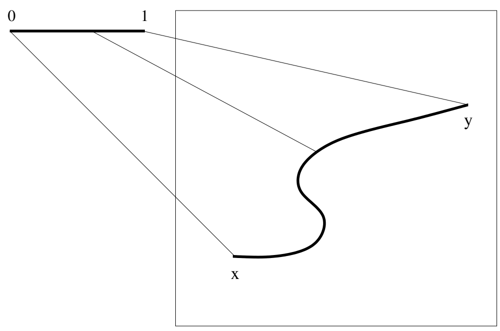
3.2.3 Paths and journeys

(**3.2.17**) Here $\mathbb{I}^n$ means $[0,1]^n$ with the subspace version of the Euclidean metric topology (as for example in (3.1.3), via (3.1.11)).

(**3.2.18**) Let (X,τ) be a topological space. An element $\alpha \in \mathsf{Top}(\mathbb{I},(X,\tau))$ is called a path in (X,τ).

An element in $\cup_l \mathsf{Top}([0,l],(X,\tau))$ is called a journey in (X,τ).

(**3.2.19**) Schematic example of a path α in $X = \mathbb{I}^2$:



We could think of this $\mathbb{I}^2$ as a model of France (we should cut out some holes to represent lakes and buildings and what have you), and the image of the path as a model for the route of a stage in the *Tour de France*.

The set $\hom(\mathbb{I},X) \supset \mathsf{Top}(\mathbb{I},X)$ is much bigger. It contains all the functions with jumps as well, and a 'generic' one of those is pretty-much impossible to draw even schematically. But the set $\mathsf{Top}(\mathbb{I},X)$ is still huge.

Then we can consider how much of this information is relevant for the *Tour*. The space and time metrics are important, but the route is not actually specified down to the millimeters (and certainly not the pace) ... what is more important is which side of the various lakes and buildings the cyclists must pass on. We'll come on to that. But first let's organise the key data of the begin and end points of the stage.

(**3.2.20**) We can partition the set of elements $\alpha \in \mathsf{Top}(\mathbb{I}, (X, \tau))$ according to $\alpha(0)$ and $\alpha(1)$. Let

$$\mathsf{P}_{(X,\tau)}(x, y) = \{\alpha \in \mathsf{Top}(\mathbb{I}, (X, \tau)) | \alpha(0) = x,\ \alpha(1) = y\}$$

$$\mathsf{P}'_{(X,\tau)}(x, y) = \bigcup_l \{\alpha \in \mathsf{Top}([0, l], (X, \tau)) | \alpha(0) = x,\ \alpha(l) = y\}$$

Note that we can stop a journey part way, and thus obtain a restricted function that is in fact a shorter journey (since continuity of the original journey ensures continuity). Indeed any interval of the domain yields a journey.

Note that there is a natural function

$$* : \mathsf{P}'_{(X,\tau)}(x, y) \times \mathsf{P}'_{(X,\tau)}(y, z) \to \mathsf{P}'_{(X,\tau)}(x, z)$$

by doing one journey then the other. (It is an exercise to prove continuity. See 3.2.24 *et seq* for a trail of hints.)

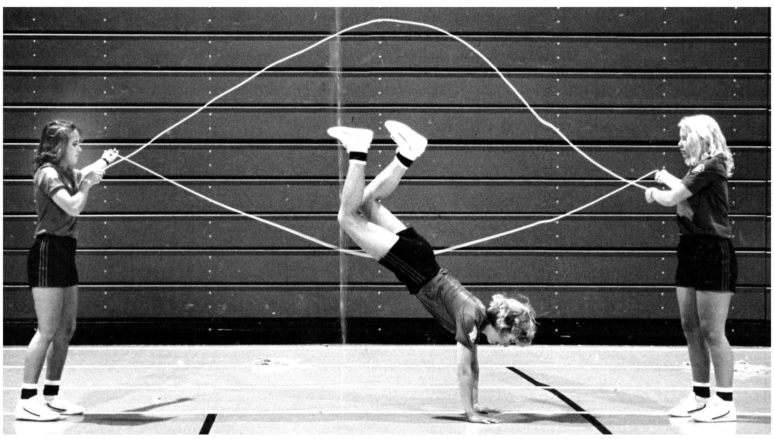
Note that this composition is associative and (allowing duration $l = 0$) unital. Thus $\mathsf{P}'_{(X,\tau)}$ is a category.

Note that $\mathsf{P}'_{(X,\tau)}$ is not a groupoid, and $\mathsf{P}'_{(X,\tau)}(x, x)$ is not a group, because we don't have inverses. There are 'reverse' trips, but composition with the reverse is a longer trip.

3.2.4 Notions of equivalence of paths

(**3.2.21**) To understand why $\alpha \in \mathsf{P}_{(X,\tau)}(x, y)$ is a 'path in X from x to y' it is perhaps helpful to have a picture. This raises several nice questions. One is the question of the relationship between pictures and spaces, and between pictures and $\mathbb{R}$ and $\mathbb{I}$ in particular.

A picture is a representation of something (possible an abstract thing; possibly a 'real world' thing) in the real world — albeit squashed onto a sheet of paper or other physical surface. Constructions in mathematics are not in the business of 'being' real physical things. But they do sometimes attempt to model physical things.



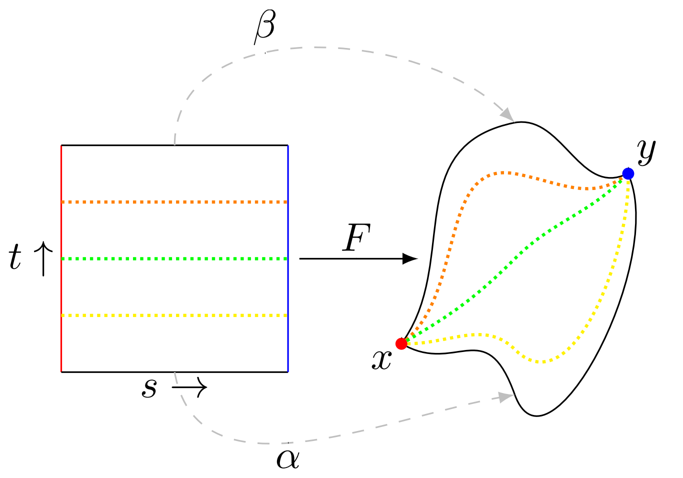
Another nice question is how your brain is able to extract meaning from a picture. The picture above in particular is a very complex representation of an even more complex real situation. How is your brain analysing it?...

(With the above in mind)... Next let us try to organise the set of paths into classes of equivalent paths with respect to a useful and natural notion of equivalence.

(**3.2.22**) **Definition**. L et (X,τ) be a topological space. Let $x,y \in X$ and let $\alpha,\beta \in \mathsf{P}_{(X,\tau)}(x,y)$ be paths from x to y. Then α is **path homotopic** to β (written $\alpha \simeq \beta$) if there exists $F \in \mathsf{Top}(\mathbb{I}^2,(X,\tau))$ such that

$$\begin{aligned} F(s,0) &= \alpha(s) \quad \forall s \in [0,1] \\ F(s,1) &= \beta(s) \quad \forall s \in [0,1] \\ F(0,t) &= x \quad \forall t \in [0,1] \\ F(1,t) &= y \quad \forall t \in [0,1]. \end{aligned}$$

Such a function F is a **path homotopy** from α to β.



Note well that F is continuous. This means that it is continuous in s and also in t. This means that for each fixed t it gives a path from x to y. And as t varies we get a kind of path of paths, from α to β. Intuitively at least this means that there can't be any 'gaps; in X in the part swept out by the path of paths. And if there is a gap between some α and β then there is no homotopy between them. Later we will see that this intuition is right.

This is along the lines that we are looking for. In the Tour de France everyone has some freedom in exactly which path they take from x to y, but they must all pass the same side of the lake.

(**3.2.23**) **Lemma.** The relation on $\mathsf{P}_{(X,\tau)}(x,y)$ given by path homotopy is an equivalence relation.
Proof. See later.

(**3.2.24**) **Definition**. Let (X,τ) be a topological space. Let $x,y,z \in X$. Let $\alpha \in \mathsf{P}_{(X,\tau)}(x,y)$, $\beta \in \mathsf{P}_{(X,\tau)}(y,z)$. The **meld** of α and β is the function $\alpha * \beta : [0,1] \to X$ defined by

$$\alpha * \beta(s) = \begin{cases} \alpha(2s) & s \in [0,\frac{1}{2}] \\ \beta(2s-1) & s \in [\frac{1}{2},1]. \end{cases}$$

Note that $\alpha * \beta \in \mathsf{P}_{(X,\tau)}(x,z)$, since at $s = 1/2$ we have $\alpha(1) = y = \beta(0)$. (We will give a more careful proof later, using a 'glue lemma' 7.35.)
The **reverse** of α is the function $\overline{\alpha} : [0,1] \to X$ defined by

$$\overline{\alpha}(s) = \alpha(1-s).$$

Note that $\overline{\alpha} \in \mathsf{P}_{(X,\tau)}(y,x)$, since the change of argument is continuous. (Again see later for a careful proof.)
For any point $x \in X$, the **constant path at** x is the function $e_x : [0,1] \to X$ defined by

$$e_x(s) = x \quad \forall s \in [0,1].$$

Note that $e_x \in \mathsf{P}_{(X,\tau)}(x,x)$.

Exercise: check that $\alpha * \beta$, $\overline{\alpha}$ and e_x are paths.

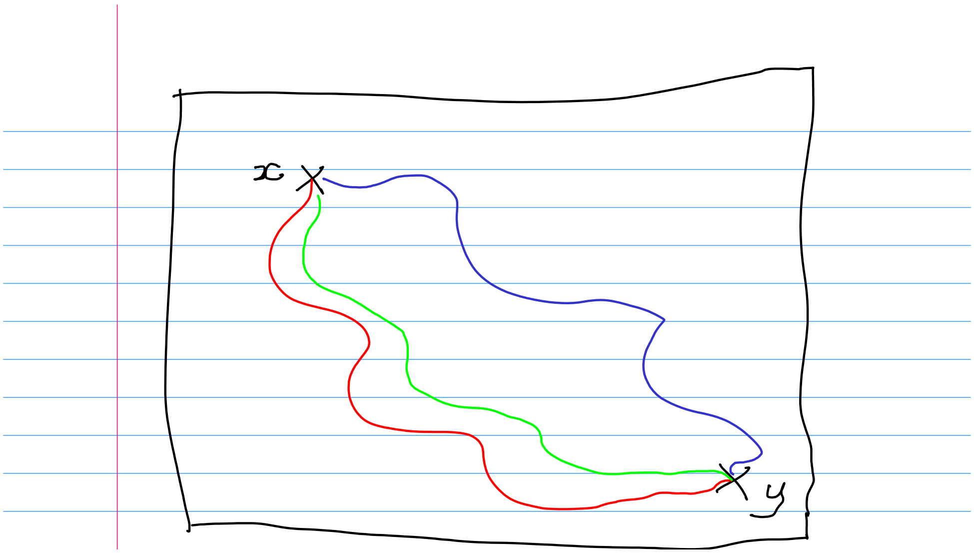
(**3.2.25**) **Lemma.** Path homotopy yields a congruence on our journey category $\mathsf{P}'_{(X,\tau)}$, making a groupoid — called the *fundamental groupoid* of (X,τ).
Proof. See later.

(**3.2.26**) Why?! ... (More interesting than the how is the why.) $\mathsf{P}'_{(X,\tau)}$ is very big. And really contains more information than we can handle or want. Every fine details of every journey. No-one needs this much information. Most tiny variations in the route are not interesting to us. Just as we do not care that a tiger has a tuft of fur slightly out of place while it is eating us.

Path homotopy gives a way of brushing over the high level of detail.

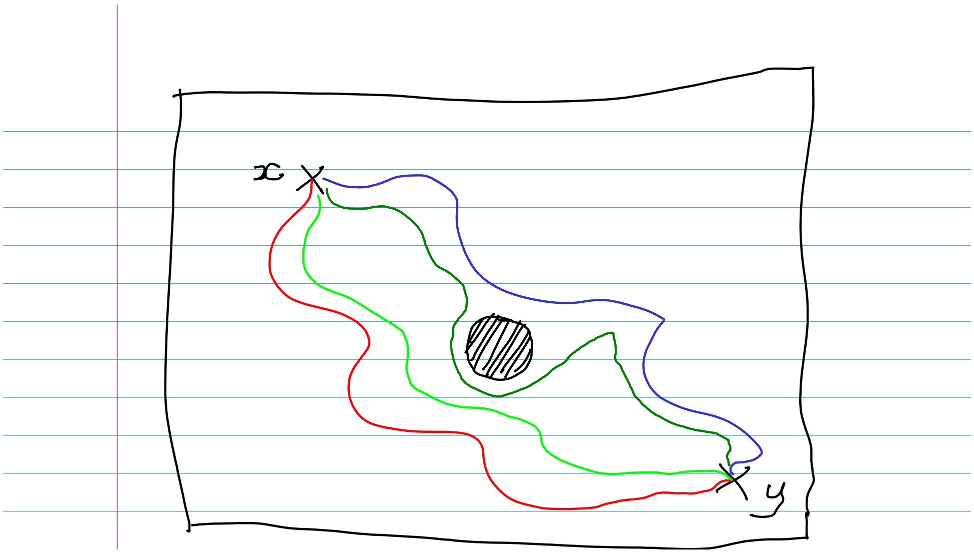
So a big question is: what information does it keep? (See later!)

Consider the example of the plane $\mathbb{R}^2$ with no 'scenery' in it, and paths from some point x to y. Roughly speaking, we might not care to distinguish between the similar red and green paths in $\mathbb{R}^2$ as sketched here:



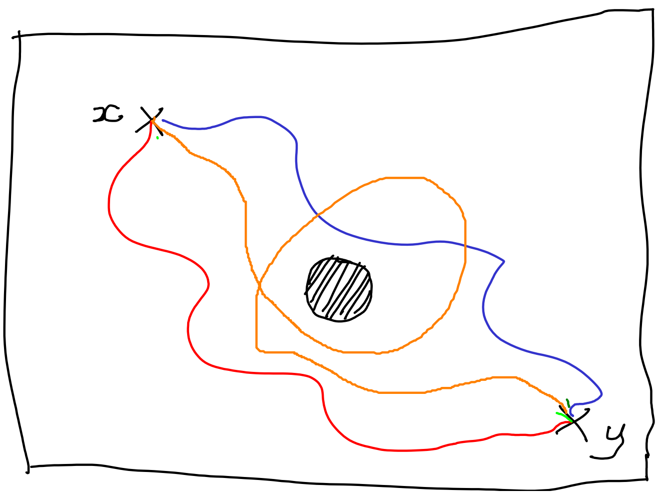
And intuitively (and in fact) they are equivalent under our path equivalence.
And then of course there are a sequence of such small changes taking us all the way over to the blue.
So in fact every path from x to y is equivalent in $\mathbb{R}^2$. We can make this very precise. We can smoothly linearly interpolate between any two such paths.

But now suppose there is some ‘scenery’ — a lake (a hole):



Now the continuum of paths from red to blue is disrupted. Now intuitively (and in fact) the lower three paths are still equivalent to each other, but they are not equivalent to the blue path. The lake has made an obstruction. The red class is different to the blue class of paths. Here it will be clear that a ‘smooth linear interpolation’ between green and blue paths no longer works. The formal interpolating paths will include paths that are impossible because they pass through the hole (the forbidden ‘lake’).
But how do we know that there is *no* way to interpolate? How do we prove it? (Hopefully it is intuitively ‘clear’. But how can we upgrade the intuition to maths?)

And, granted for a moment that the intuition is correct, how much information are we now retaining. We don’t distinguish red and green. But then how many different classes of paths are there? We’ve seen two. Are there more?



In fact all three of the path classes given by the three paths above are distinct.
The orange one perhaps gives a clue for a good way to codify the problem. We could ask about paths where we go around the lake and then actually go back to x. In $\mathbb{R}^2$ every loop path x-to-x can be smoothly interpolated to a very small loop and ultimately to the constant path. So there is only one class of paths from x to x.
But once we put the hole in, this changes. Paths winding around the lake clockwise (or anti-clockwise) multiple times cannot be smoothly unwound.

As we will see, even working with path-equivalence, we still have different path classes, corresponding to this winding number. So the classes of paths are in correspondence with $\mathbb{Z}$.
We can even compose (classes of) paths from x-to-x to make more paths from x-to-x. And we find that the group formed from the classes in this way is isomorphic to the group $(\mathbb{Z}, +)$.

(**3.2.27**) So how about the formal proof question? How can we prove these intuitive claims? The intuition probably feels pretty strong — you can't ignore a big lake. But suppose that instead of excluding a big disk like that we excluded a much smaller disk? Or suppose we even just excluded a single point? What happens then?

The first useful step in formalising the problem is to note that all of the issues that are arising are arising from parts of the path close to the hole. There are some simplifying changes that we could make to the problem that would not change the key aspects. For example we would move x close to the hole. And the wide open uninterrupted spaces over on the far left and right and above and below are not really relevant. The problem at hand would not change if we narrowed the band of space to a thin corridor around the hole.

...What we are saying here is that it would essentially be the same problem if we narrowed to a circular region around the hole. This kind of simplification is sometimes called a...
deformation retract.

(**3.2.28**) The other ingredient we are going to need in our formalism is a generalisation of the 'fundamental group' of classes of paths from x to x in X. We have given a construction, for each space X and point $x \in X$ of a fundamental group $\pi_1(X, x)$ of classes of paths. Suppose we have a continuous function $h : X \to Y$. What can we say about $\pi_1(Y, h(x))$?

Well firstly, there is an image of any path α in X under the map h. So this gives a path in Y:

$$h \circ \alpha : [0, 1] \to Y$$

This passes to a well-defined construction of an element (call it $h_*([\alpha])$) of $\pi_1(Y, h(x))$ from an element $[\alpha]$ of $\pi_1(X, x)$ (one checks that different choices of path within the class do not effect the outcome class):

$$h_*([\alpha]) = [h \circ \alpha]$$

Even better news is that

$$h_*([\alpha] * [\beta]) = h_*([\alpha]) * h_*([\beta])$$

(note that the $*$ is being used in three different ways here - so care is needed).

(**3.2.29**) Better news yet is the following.
Theorem.
(a) If $f : X \to Y$ and $g : Y \to Z$ are continuous functions then $(g_* \circ f_*) = (g \circ f)_*$.
(b) If $id_X : X \to X$ denotes the identity function then $(id_X)_* : \pi_1(X, x) \to \pi_1(X, x)$ is the identity function.

(**3.2.30**) This tells us that π_1 is a topological invariant.

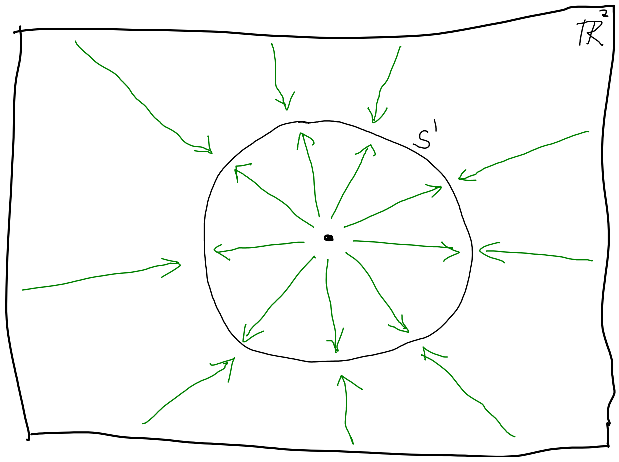
(**3.2.31**) And we can do even better still. We just need to formalise the deformation retract mentioned above.
Let X be a space and $A \subset X$. The subspace A is a *strong deformation retract* of X if there is a continuous map

$$H : X \times \mathbb{I} \to X$$

such that $H(x, 0) = x$; $H(x, 1) \in A$ and $H(a, t) = a$ for $a \in A$.
This is the codification of the contraction we mentioned before, for example, of $\mathbb{R}^2\setminus$ hole to a narrow circle around the hole.

(**3.2.32**) Example. $\{0\}$ is a strong DR of $\mathbb{I}$.
Example. S^1 is a SDR of $\mathbb{R}^2 \setminus \{0\}$.

Here is a schematic:



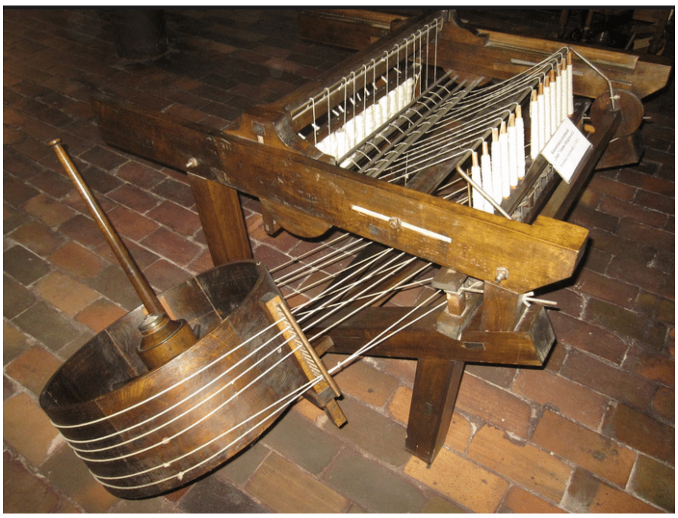
Note that this map would not be defined at 0 since there there is symmetry and it is not clear which direction to set off in to go towards the circle. But with this point removed, for every other point it is well-defined how to move — to retract — continuously towards the circle.

(**3.2.33**) Theorem.
Let A be a SDR of X, and let $a \in A$. Then the inclusion map $i : A \to X$ yields an isomorphism

$$i_* : \pi_1(A, a) \to \pi_1(X, a)$$

(**3.2.34**) Example. $\pi_1(\mathbb{R}^2 \setminus \{0\}) \cong \pi_1(S^1)$.
(We omit the a here since it is the same for any a.)

(**3.2.35**) So now we know about the key properties of $\mathbb{R}^2$ with a hole... so long as we can work out $\pi_1(S^1)$.



3.2.5 Zen hairdryer (motions of embedded sets)

(**3.2.36**) What we have so far is a kind of person-centric model of a 'journey' — we record the embedding of the traveller's self in X at each instant of time.

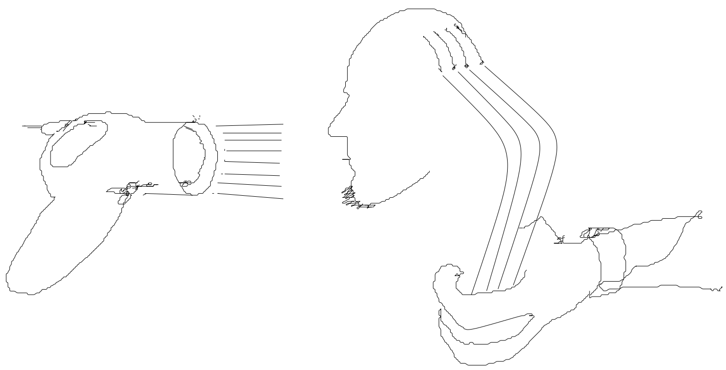
What happens if we try to see this as part of a bigger picture. The traveller is just a construct emerging from the arrangement of 'molecules' in X viewed holistically. Can we give a description of the changing configuration of X itself that has the journey as a particular manifestation? — This is going in the other direction in terms of keeping more information than we want (or is it!? we can impose equivalences on the ambient space motions too).

Because our (X, τ) has a metric topology (let's say d is the metric, so $\tau = \tau_d$) we can consider the set $\mathsf{Top}(X, X)$ as a topological space itself, and by restriction also $\mathsf{Top}^h(X, X)$. (In essence, two homeomorphisms f, f' are near to each other if the movements of every individual point $f(x), f'(x)$ are near to each other. More explicitly we can define an 'elastic' metric on $\mathsf{Top}(X, X)$ by

$$d'(f, f') = sup_{x \in X} d(f(x), f'(x))$$

and use the corresponding metric topology. See §6.1, and cf. in particular Def.6.5, later.) That means that we can consider paths in $\mathsf{Top}^h(X, X)$. Specifically we can consider elements of $\mathsf{Top}(\mathbb{I}, (\mathsf{Top}^h(X, X), \tau_\circ))$, where $\tau_\circ$ is the topology on $\mathsf{Top}(X, X)$ indicated above.

Now the whole 'universe' is evolving along some path, and the journey of the traveller we started with is just a special bit of that.



We will return to these points later.

3.3 More algebraic aspects of topology

This section can be skipped (or skimmed) on first reading. We assume a little ring theory notation. And we will leave the key (but more complex) examples until later. References include Hartshorne [Har77], on Algebraic Geometry; and Borel [Bor69] on Linear Algebraic Groups. (The section should be read as 'non-examinable but interesting'.)

(**3.3.1**) Let k be a field. A polynomial $p \in k[x_1, ..., x_r]$ determines a map from k^r to k by evaluation. For $P = \{p_i\}_i \subset k[x_1, ..., x_r]$ define

$$Z(P) = Z(\{p_i\}_i) = \{x \in k^r \ : \ p_i(x) = 0 \ \forall \ i\}$$

An *affine algebraic set* is any such set, in case k algebraically closed. An *affine variety* is any such set, that cannot be written as the union of two proper such subsets. (See for example, Hartshorne [Har77, I.1].)

(**3.3.2**) Example. (I) $Z(x_1x_2 - 1) = Z(\{p(x_1, x_2) = x_1x_2 - 1\})$ is a variety in k^2. Its points $(x_1, x_2) = (\alpha, \beta)$ may be given by a free choice of α (say) from $k^\times$, with β then determined. (Note that this latter characterisation looks like an open subset of k (specifically the complement of $Z(x)$), but the original formulation makes it clear that it is closed in k^2.)

(II) The fundamental theorem of algebra says that if $k = \mathbb{C}$ and $p \in \mathbb{C}[x] \setminus \mathbb{C}$ (a non-constant polynomial) then $Z(\{p\}) \neq \emptyset$, i.e. every non-constant complex polynomial has a root.

(**3.3.3**) The set of affine varieties in k^r satisfy the axioms for closed sets in a topology. This is called the *Zariski topology.* The Zariski topology on an affine variety is simply the corresponding subspace topology.

The set $I(P) \in k[x_1, ..., x_r]$ of all functions vanishing on $Z(P)$ is the ideal in $k[x_1, ..., x_r]$ generated by P. We call

$$k_P = k[x_1, ..., x_r]/I(P)$$

the *coordinate ring* of $Z(P)$.

(**3.3.4**) Let Z be an affine variety in k^r and $f : Z \to k$. We say f is *regular* at $z \in Z$ if there is an open set containing z, and $p_1, p_2 \in k[x_1, ..., x_r]$, such that f agrees with p_1/p_2 on this set.

(**3.3.5**) A morphism of varieties is a Zariski continuous map $f : Z \to Z'$ such that if V is open in Z' and $g : V \to k$ is regular then $g \circ f : f^{-1}(V) \to k$ is regular.

(**3.3.6**) Given affine varieties X, Y then $X \times Y$ may be made in to an algebraic variety (affine algebraic set — see e.g. [Bor69]) in the obvious way.

(**3.3.7**) An *algebraic group* G is a group that is an affine variety such that inversion is a morphism of algebraic varieties; and multiplication is a morphism of algebraic varieties from $G \times G$ to G.

3.4 More new from old: Quotients

Recall that a topology on a set X is an element of $\mathtt{PP}(X)$ satisfying some special conditions (T0-3) (from T0: $\emptyset \in \tau$ to T3: closure under union).

(**3.4.1**) **Lemma.** Let (X, τ) be a topological space and Y a set. We have a function

$$\tau_{-} : hom(X, Y) \to \mathtt{PP}(Y)$$

given for $f : X \to Y$ a function by

$$\tau_f = \{S \subseteq Y \mid f^{-1}(S) \in \tau\}$$

This τ_f is a topology on Y.
Proof. (T0,1): note $f^{-1}(Y) = X$ and $f^{-1}(\emptyset) = \emptyset$.
(T2): Suppose $U, V \in \tau_f$. Then

$$f^{-1}(U \cap V) = f^{-1}(U) \cap f^{-1}(V)$$

and the RHS is in τ because it is a topology.
(T3): Suppose $\{S_\lambda\}_{\lambda \in \Lambda}$ a collection in τ_f. Then

$$f^{-1}(\bigcup_{\lambda \in \Lambda} S_\lambda) = \bigcup_{\lambda \in \Lambda} f^{-1}(S_\lambda)$$

The RHS is a union in τ by construction, and hence in τ as required. □

(**3.4.2**) Given an equivalence relation $\sim$ on a set X, recall that $X/\sim$ denotes the set of classes.

(**3.4.3**) Let (X, τ) be a topological space. Let $\sim$ be an equivalence relation on X and let

$$f_\sim : X \to X/\sim$$

be the map $x \mapsto [x]$. The topology $\tau_{f_\sim}$ is the 'quotient topology' on $X/\sim$. The space $(X/\sim, \tau_{f_\sim})$ is the 'quotient space'.

3.5 Homeomorphism and 'shapes'

See Sec.?? for more examples of the following. Hereafter we may write $(X, \tau) \in \mathsf{Top}$ to mean that (X, τ) is a topological space.

(**3.5.1**) Let $(X, \tau) \in \mathsf{Top}$. Let $U, V \in \tau \setminus \{\emptyset\}$. If $\{U, V\}$ a partition of X then (X, τ) is 'disconnected'. If there is no such U, V then (X, τ) is 'connected'.

(**3.5.2**) Example: the real line is connected (this will perhaps not come as a surprise, but it does serve to illustrate the definition). There are of course lots of ways of partitioning $\mathbb{R}$ into two subsets, but then at least one of them is not open. For example, although $(-\infty, 0)$ is open, its complement in $\mathbb{R}$ is $[0, \infty)$ which is not open.
On the other hand the subspace $[0, 1] \cup [2, 3]$ is not connected ($[0, 1]$ is actually open in this subspace, as we will see later).

(**3.5.3**) Path-connected: Let $(X, \tau) \in \mathsf{Top}$. (X, τ) is 'path-connected' if for $x, y \in X$ there is a path from x to y. That is, there is an element $f \in \mathrm{Top}(\mathbb{I}, (X, \tau))$ with $f(0) = x$ and $f(1) = y$.

Example: With the usual topology, $\mathbb{R}$ is path-connected, but $[0, 1] \cup [2, 3]$ is not.

As you might expect, path-connected implies connected. But connected does not imply path-connected. See later.

(**3.5.4**) Compact: Let $(X, \tau) \in \mathsf{Top}$. Let $A \subset X$. An *open cover* of A is a family $\{U_\lambda\}_{\lambda \in \Lambda} \subset \tau$ such that

$$A \subseteq \bigcup_{\lambda \in \Lambda} U_\lambda$$

An open cover is *finite* if index set Λ is finite. If $\Lambda' \subset \Lambda$ also gives a cover then it is called a *subcover.*

Let $(X, \tau) \in \mathsf{Top}$. A subset A is a 'compact subset' if every open cover of A has a finite subcover. (X, τ) itself is compact if it is a compact subset.

(**3.5.5**) Example: the real line is not compact. An example of an open cover is $\{(-n, n)\}_{n \in \mathbb{N}}$ — all the elements are open and for each $x \in \mathbb{R}$ there is an n such that $x \in (-n, n)$. But if $S \subset \mathbb{N}$ is finite then it has a maximal element, m say, and note that m is not in $(-m, m)$ and hence not in the union over S.

(**3.5.6**) Linearity: Note that $[0, 1]$ is a topological space, but not a vector space. ... We will come back to this point later.

...OK, that should be enough glimpses into the matrix to attach our overview to. Soon now we will get into *details*!

3.6 Aside: Examples and exercises on the sets Top_X

Recall that in the N_x notation we can give the set Top_X of all topologies on $X = \{a, b\}$ in the order (a, b) as: (ab, ab), (a, ab), (ab, a), (a, b). And the classes under the permutation action on $\{a, b\}$ are: $[(ab, ab)]$, $[(a, ab)]$, $[(a, b)]$, with the second class being a permutation group orbit of two topologies.

(**3.6.1**) In the N_x notation, the topologies on $\{a, b, c\}$ in the order (a, b, c) are given by:

(abc, abc, abc),	orbit of 1 topologies of order 2
$(ab, ab, abc), ...,$	orbit of 3 topologies of order 3
$(a, abc, abc), ...,$	orbit of 3 topologies of order 3
$(a, ab, abc), ...,$	orbit of 6 topologies of order 4
$(a, bc, bc), ...,$	orbit of 3 topologies of order 4
$(a, ab, ac), ...,$	orbit of 3 topologies of order 5
$(a, b, abc), ...,$	orbit of 3 topologies of order 5
$(a, b, ac), ...,$	orbit of 6 topologies of order 6
(a, b, c)	orbit of 1 topologies of order 8

where 'orbit' refers to the permutation of *abc*.

(**3.6.2**) Exercise. How do we know we've got them all?
(Remark. If we define a relation by $a \sim b$ if $a \in N_b$ then note that this is reflexive and transitive but not necessarily symmetric.)
(Remark. As noted above, all the topologies on a finite set X are elements of $\mathtt{P}(\mathtt{P}(X))$. For each element of this set we can in principle check closure under (here necessarily finite) union and intersection.)

(**3.6.3**) It is relatively easy to enumerate the topologies of order 3 on a finite set X. We simply have to choose a non-empty proper subset Y of X. The closure conditions will be satisfied automatically, because $\emptyset \subset Y \subset X$.

(**3.6.4**) For the topologies of order 4 on a finite set X we can proceed as follows. First we may choose a proper subset Y as above. The second subset, Y' say, must satisfy certain conditions, so as to avoid that the union and intersection closures give something new. If we take $Y \subset Y'$ then we have closure under both. Otherwise we will get something new unless the intersection is $\emptyset$ and the union is X; i.e. unless Y, Y' is a partition of X.

Let us consider the inclusion-type cases $Y \subset Y'$ — altogether

$$\emptyset \subset Y \subset Y' \subset X$$

Up to homeomorphism the choice of the first subset is characterised by its order y say, with $y \in \{1, 2, ..., n-2\}$. For $y = i$ there are then choices for the order of Y', $y' \in \{i+1, ..., n-1\}$. So altogether there are $\frac{n(n-1)}{2}$ classes where $n = |X| - 1$.

For the partition-type cases we simply choose one subset (the other one is then determined) - if we choose from sizes up to $|X|/2$ then we will avoid double counting.

(**3.6.5**) Exercise. What about order 5?

(**3.6.6**) Note that if $(N_1, N_2, ..., N_{n-1})$ gives a topology on $\{1, 2, ..., n-1\}$ then

$$(N_1, N_2, ..., N_{n-1}, 12..n)$$

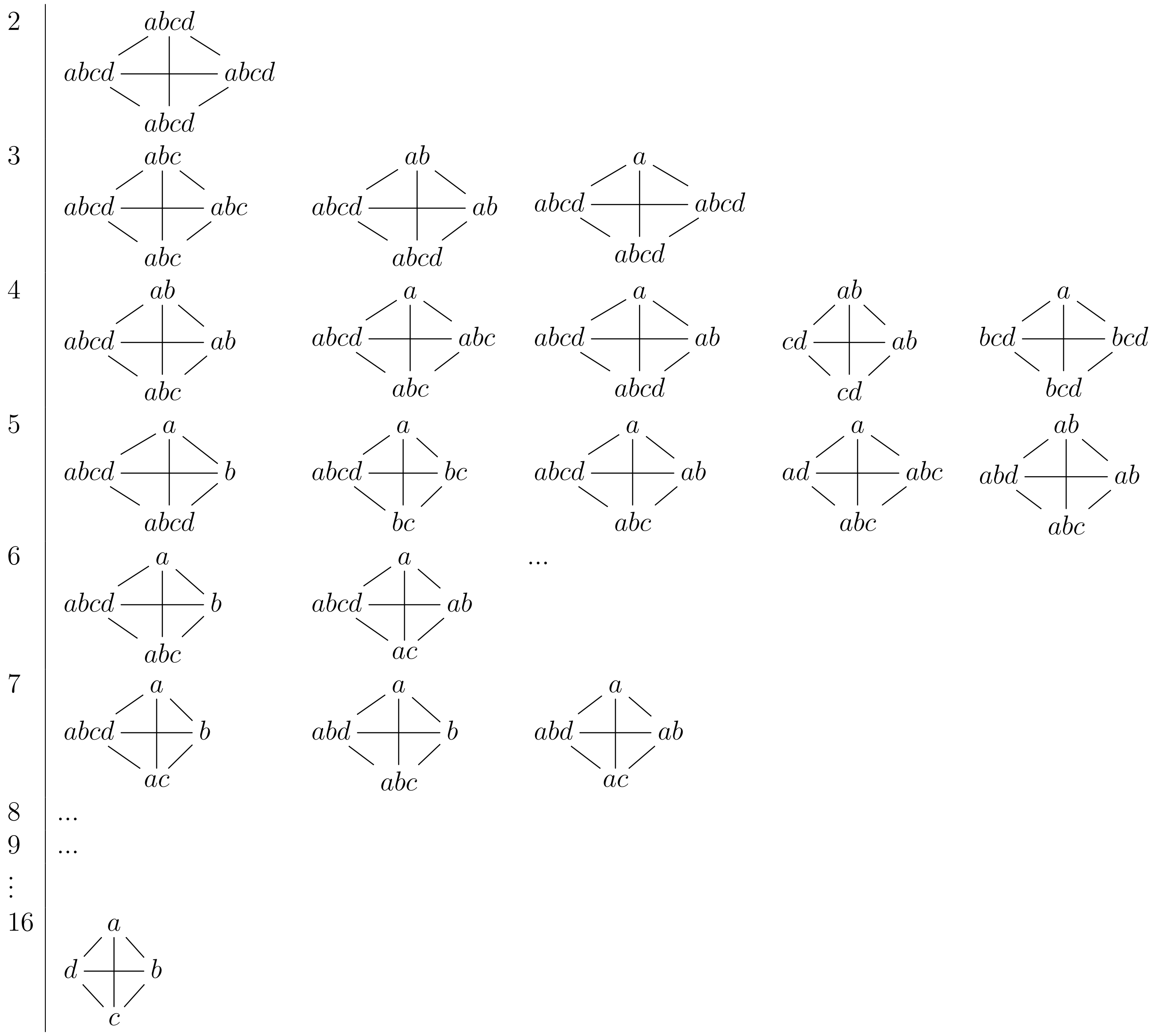
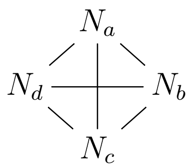


Figure 3.1: Topologies on $\{a, b, c, d\}$ arranged by order $|\tau|$ (incomplete).

gives a topology on $12..n$.

(**3.6.7**) Figure 3.1 shows the begining of rank-4 ($|X| = 4$) arranged in the format



Note that we need to check union and intersection at each edge (and so on, if these give something new). Here the rows are organised by increasing order of topology τ (indicated on the left in the figure).

(**3.6.8**) Exercise. How about $\{a, b, c, d, e\}$? (Cf. e.g. Sloane [**?**].)

3.7 Aside: Algebra, geometry and combinatorics

This section is about Tony and his hair. (Or rather it is about a big class of problems that we can use Tony's hair to illustrate.) We are interested in getting some control over the choice of configurations for his hair that Tony can make. If we look in sufficient detail, it seems that there might be infinitely many different possible configurations. Indeed there might be uncountably infinitely many.

This multitude presents a problem, if it is control that we seek. What does 'control' look like? In this context control might look like a list of styles or a recipe for generating a list of styles to choose from.

So what is a style? Roughly speaking it is a collection of different hair configurations that we consider, at least informally, to be the same. So when are two configurations the same? And when are they not?

And what is a configuration anyway?

A first notion of a configuration of Tony's hair might be the position of his hair at an instant in time. His hair is going to occupy a subset of the space in which Tony himself lives. Roughly speaking we might think of this living space as being a copy of $\mathbb{R}^3$ for now. Thus the configuration of the hair becomes represented by a subset of $\mathbb{R}^3$ (of course we might care about which molecule is in which position; and what the molecules are doing — let us not worry about that for now).

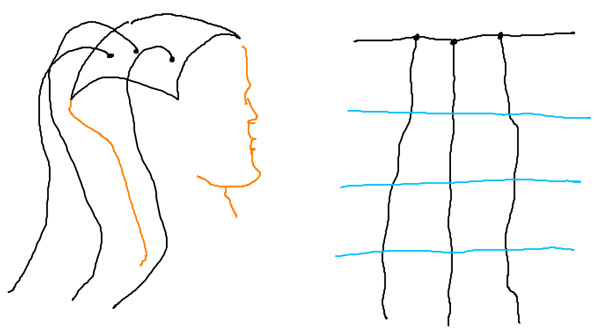
But there are lots of such subsets. And most of them are not any kind of configuration of hair. One way to think of a configuration treats the individual strands of hair as essentially linear, i.e. to treat each as given by a copy of a real interval embedded in $\mathbb{R}^3$:

$$\Psi : [0,1] \hookrightarrow \mathbb{R}^3$$

say. The hair does not have zero thickness like an embedded interval does, but we can perhaps think of a small ball around each point of the embedded line to build the thickness. ...So how do we describe an embedding?

To think about this we can first note that there are simpler problems with the same issue, which perhaps we can think about first. Instead of embedding a 1d thing in a 3d thing, we could embed a 0d or 1d thing in a 1 or 2d thing. Perhaps the simplest is to embed a 0d thing (a point) in $\mathbb{R}$. This amounts to choosing a point in $\mathbb{R}$. So here immediately we see that the set of configurations is uncountably infinite.

A notable feature of Tony's hair is that it grows (in a linear/stringy way) from follicles in his scalp.



His scalp has the shape of a surface, so we can think of this as belonging to a copy of $\mathbb{R}^2$ sitting inside $\mathbb{R}^3$. Indeed as Tony's hair grows, we might think of the tips as also residing in a surface (since for any finite set of points a surface can always be found such that the points all lie in the surface, this is so far a triviality; but let's keep going) — perhaps a surface Σ_t 'moving

away' from the scalp surface as the hair gets longer/bigger over time t.
The growth of the hair is not part of the instaneous configuration of the hair, it is just a convenient way of thinking of introducing a parameterisation that breaks the configuration down to a sequence of subsets of a sequence of embedded $\mathbb{R}^2$s (sometimes called a lamination, after the process of building a 3d thing — perhaps a boat or a piece of furniture — from 2d layers).
So the overall configuration can be thought of as a collection of configurations of infinitessimal sections of hair in copies of $\mathbb{R}^2$ (together with the data of how these copies of $\mathbb{R}^2$ are a lamination of $\mathbb{R}^3$). In other words we have a sequence (or at least an ordered set) of subsets of $\mathbb{R}^2$.
But of course the subset in a given 'layer' is not independent of the subset in a nearby layer. If we think of the follicles as a discrete, even finite, collection of points in the scalp $\mathbb{R}^2$, then we can label them and in so doing label the individual strands of hair. But if we need to treat the follicles as more complex than finite collections of points then it is not so clear how to do this. One approach is to use the realisation of $\mathbb{R}^3$ as laminated $\mathbb{R}^2$s as an underlying structure that holds the hair structure. If we do this we are, for the moment, adding *more* information rather than simplifying from configurations to styles. But the simplification can still come — slightly later.

3.7.1 Notation

Recall the following.

$\mathsf{Top}^h(M, N)$ is the subset of $\mathsf{Top}(M, N)$ of homeomorphisms.

If X is a compact space and Y a metric topological space with metric d then

$$d'(\alpha, \beta) = \sup_{x \in X} d(\alpha(x), \beta(x))$$

is a metric on $\mathsf{Top}(X, Y)$ and we write $\mathsf{TOP}(X, Y)$ for the corresponding topological space (and if we refer to $\mathsf{Top}(X, Y)$ as a topological space we will mean this topology).

For example if Y is a metric topological space then $\mathsf{TOP}(\mathbb{I}, Y)$ is a space of paths in Y.

Chapter 4

More preliminaries

It is safe to skip this Chapter on first reading.

4.1 Set partitions

(**4.1.1**) Let $\mathsf{E}_S \subset U_S$ denote the set of equivalence relations (reflexive, symmetric, transitive/RST relations) on set S. Let P_S denote the set of partitions of S. Note the natural bijection

$$\mathsf{E}_S \underset{\kappa}{\overset{\epsilon}{\leftrightarrows}} \mathsf{P}_S.$$

For $\rho \in U_S$ let $\bar{\rho} \in U_S$ be the smallest transitive relation containing ρ. The relation $\bar{\rho}$ is called the *transitive closure* of ρ.

(**4.1.2**) Let a, b be RS relations on any two finite sets. Then $a \cup b$ is an RS relation on the union. Let $ab := \overline{a \cup b}$ be the transitive closure of $a \cup b$.

Note that $\overline{a \cup b}$ is an equivalence relation on the union of the two finite sets. Note that

$$\overline{\bar{a} \cup b} = \overline{a \cup b} \tag{4.1}$$

If a, b partitions then $\epsilon(a), \epsilon(b)$ are RS (indeed RST), and we will understand by ab the partition given by $ab = \kappa(\epsilon(a)\epsilon(b))$.

(**4.1.3**) Proposition. *For a, b, c RS relations*

$$a(bc) = (ab)c$$

Proof.

$$(ab)c = \overline{(\overline{a \cup b}) \cup c} \overset{(4.1)}{=} \overline{(a \cup b) \cup c} = \overline{a \cup b \cup c} = \overline{a \cup \overline{b \cup c}} = a(bc)$$

□

(**4.1.4**) Let $\mathsf{P}_{n,m} = \mathsf{P}_{\underline{n} \cup \underline{m'}}$; and $\mathsf{P}_n = \mathsf{P}_{n,n}$. Let $\mathsf{E}_{n,m} = \mathsf{E}_{\underline{n} \cup \underline{m'}}$ similarly. For $a \in \mathsf{P}_{n,m}$ let a' be the partition of $\underline{n'} \cup \underline{m''}$ obtained by adding a prime to each object in every part.

For $a \in \mathsf{P}_{l,m}$, $b \in \mathsf{P}_{m,n}$ partitions (and hence $\epsilon(a)$, $\epsilon(b)$ equivalence relations) note that $\epsilon(a)\epsilon(b')$ is an equivalence relation on $\underline{l} \cup \underline{m'} \cup \underline{n''}$. Restricting to $\underline{l} \cup \underline{n''}$ this equivalence relation gives again a partition, call it $r(ab')$ (indeed if a, b are pair-partitions then so is $r(ab')$).

For $x \in \underline{l} \cup \underline{n''}$ let $u(x) \in \underline{l} \cup \underline{n'}$ be the image under the action of replacing double primes with single.

We may define a map

$$\circ : \mathsf{P}_{l,m} \times \mathsf{P}_{m,n} \to \mathsf{P}_{l,n}$$

by

$$a \circ b \;=\; u(r(ab')) \in \mathsf{P}_{l,n}$$

— the image under the obvious application of the u map.

(**4.1.5**) Proposition. *For each $n \in \mathbb{N}$ the map $\circ : (a,b) \mapsto u(r(ab'))$ defines an associative unital product on P_n, making it a monoid, with identity $1_n \;=\; \{\{1,1'\},\{2,2'\},...,\{n,n'\}\}$.*

Proof. To show associativity note that ab' encodes $a \circ b$ directly, except that it is encoded via the unprimed and double-primed 'vertices'. Thus $(ab')c''$ encodes $(a \circ b) \circ c$ via the unprimed and triple-primed vertices. Meanwhile $b'c''$ encodes $b \circ c$ via the primed and triple-primed vertices; thus $a(b'c'')$ encodes $a \circ (b \circ c)$ via the unprimed and triple-primed vertices. But by Prop.4.1.3 we have $a(b'c'') = (ab')c''$.

To show unital with identity 1_n: Exercise. ■

(**4.1.6**) A convenient pictorial realisation of such a set partition p, i.e. a realisation as a picture in the plane, is as follows. (See also **??**.)

Firstly, a digraph G (as in 4.2.16) on vertex set V determines a relation on V in the obvious way. In particular a graph determines a symmetric relation. Hence a graph G determines an equivalence relation on V (take the RT closure); and hence also an equivalence relation on (or partition of) any subset of V. Thus it is enough to realise a suitable graph G of p as a picture.

To depict such a G one draws a set of points for the vertices V, and specifies an injective map from the underlying set of p to V; and then draws a 'regular' collection of 'edges'. Here a picture edge is a piecewise smooth line between two vertices. A collection is *regular* if two lines never meet at points where they do not have distinct tangents. The collection consists of one picture edge for each vertex pair that are associated by an edge in G. ('Incidental' vertices in the picture are those *not* associated to the underlying set.)

Note that two elements from the underlying set are in the same part in p if there is a path between their vertices.

(**4.1.7**) For a partition in P_n one may arrange the underlying-set vertices naturally as two parallel rows of vertices (if there are incidental vertices these are drawn between the two rows). In this realisation the product $\circ$ may be computed, schematically, by concatenating the two pictures so as to identify certain vertices in pairs between two rows — one row from each picture (thus forming a 'middle' row).

(**4.1.8**) Let $J_S \subset \mathsf{P}_S$ be the set of pair-partitions of S. Let

$$J_{n,m} = J_{\underline{n} \cup \underline{m}'} \;\subset \mathsf{P}_{n,m}$$

Set $J_n = J_{n,n}$.

(**4.1.9**) Proposition. *The compostion $\circ$ restricts to make J_n a monoid.*

Proof. Exercise. ■

(**4.1.10**) A partition is *non-crossing* if there is such a pictorial realisation having the property that all lines are drawn in the interior of the interval defined by the two rows, and no two lines cross. Let T_n denote the subset of non-crossing pair partitions.

One easily checks that the product above restricts to make T_n a monoid. This is sometimes called the n-th *Temperley–Lieb monoid*.

(**4.1.11**) One could similarly imagine drawing a realisation of a partition on a cylinder. This leads us to a notion of cylinder-non-crossing pair partitions. There are several further subsets of the set of partitions that are characterised in terms of geometrical embeddings.
Exercise: Find some more submonoids of P_n.

4.1.1 Exercises on closed binary operations

(**4.1.12**) A closed binary operation on a (finite) set S (of degree n) may be given by a multiplication table — an element of $S^{S\times S}$. There are $|S^{S\times S}| = n^{(n^2)}$ of these. Note that an ordering of S induces an ordering on the set of closed binary operations (read order the entries in the multiplication table and dictionary order the ordered lists).

Define a natural notion of isomorphism of closed binary operations on S, and determine the number of isomorphism classes for $n = 2$. Is commutativity a class property? If so, how many of these classes are commutative?

Which of the following are semigroups/ monoids / groups?:

	a	b
a	aa	ab
b	ba	bb

:

	a	b
a	a	a
b	a	a

	a	b
a	a	a
b	a	b

	a	b
a	a	a
b	b	a

	a	b
a	a	a
b	b	b

	a	b
a	a	b
b	b	a

	a	b
a	b	a
b	a	a

(Hint: S,M,X ($b(ab) = b$),S,G,X ($(aa)b = a$).)

Explain the following statement: "For $n = 3$, most binary operations are not associative." (Hint: 113 of them are associative.)

4.2 Partial orders, lattices and graphs

4.2.1 Posets

General references on posets and lattices include Birkhoff [Bir48], and Burris and Sankappanavar [BS81, §1].

(**4.2.1**) A *relation* on a set S is a subset of $S \times S$ as in (2.2.2). Thus the intersection of any set of relations on S is certainly a relation. Indeed the intersection of any set of transitive relations is transitive.

The *transitive closure* of a relation ρ on S is the intersection of all transitive relations containing ρ. (This transitive relation is non-empty since $S \times S$ is a transitive relation.)

(**4.2.2**) A *poset* is a set with a reflexive, antisymmetric, transitive relation.

An acylic (no cyclic chains) relation ρ on S defines a partial order, by taking the transitive reflexive closure $TR(\rho)$.

Note that every relation in the interval $[\rho, TR(\rho)]_\subseteq$ (with respect to the inclusion partial order) is acyclic.

We may consider the set of all relations having the same transitive reflexive closure. If the closure is a poset then all the relations 'above' it are acyclic. A minimal such relation is a *transitive reduction* (of any element of this set).

If S is finite then there is a unique transitive reduction of an acylic relation. Otherwise there may be no (or one, or multiple) transitive reductions.

If there is a unique transitive reduction of an acyclic relation we call this the *covering relation.*

(**4.2.3**) Note that the opposite relation to a poset is a poset.

(**4.2.4**) If we use the notation $(S, \geq)$ for a poset then we may write $a > b$ for $a \geq b$ and $a \neq b$. In this case the relation $(S, >)$ induces the same poset.

(**4.2.5**) Further we may write $(S, \leq)$ for the opposite relation (which is another poset).
See what we did there!? There are actually a lot of subtleties to this notation. For example

the symbol we use needs to be sufficiently 'directed'. (But hopefully this is ok, because it is familiar.)

(**4.2.6**) Let $(S, \geq)$ be a poset, and $s, t \in S$. We say s *covers* t if $s > t$ and there does not exist $s > u > t$.

(**4.2.7**) The notion of cover/covering relation leads to the notion of *Hasse diagram*, as for example in [Bir48, BS81].

(**4.2.8**) A poset satisfies ACC (is *Noetherian*) if every ascending chain terminates.

For example, the poset of ideals, ordered by inclusion, of the ring $\mathbb{Z}$ satisfies ACC.

4.2.2 Lattice meet and join

(**4.2.9**) Notation: By convention if we declare a poset $(S, \leq)$ then $a \leq b$ can be read as *a is less than or equal to b* (although the opposite relation is a perfectly good poset, and we could in principle have associated the relation symbol $\leq$ to that).

(**4.2.10**) Let $(S, \geq)$ be a poset. A *lower bound* of subset $T \subset S$ is an element $b \in S$ such that $t \geq b$ for all $t \in T$.
A *greatest lower bound* of $T \subset S$ is a $b \in S$ such that b is a lower bound and for each lower bound c we have $b \geq c$.

(**4.2.11**) Upper bounds and *least upper bounds* reverse these definitions in the obvious way.

(**4.2.12**) With the above convention, a poset S is a *join semilattice* if every pair $s, t \in S$ has a least upper bound (*join*) in S.
A poset is a *lattice* if both it and its opposite are join semilattices.

We may write $(L, \leq, \vee, \wedge)$ to give names to the set L, relation $\leq$, join $\vee$ and meet (opposite join) $\wedge$.

(**4.2.13**) Example. The power set $\mathtt{P}(S)$ of a finite set with the inclusion order is a lattice. An upper bound of $s, t \in \mathtt{P}(S)$ is any set containing sets s, t; and the least upper bound is the union. That is

$$s \vee t \;=\; s \cup t.$$

A lattice is *distributive* if $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$ (and similarly with $\wedge \leftrightarrow \vee$).

So $(\mathtt{P}(S), \subseteq, \cup, \cap)$ is distributive.

(**4.2.14**) A lattice is *modular* if

$$S \wedge (T \vee U) \;=\; (S \wedge T) \vee U \qquad \Rightarrow \qquad S \geq U.$$

(**4.2.15**) For $(P, \leq)$ a lattice, the interval

$$[a, b]_{\leq} \;:=\; \{c \in P \mid a \leq c \leq b\}$$

is sometimes called a *quotient* (see e.g. Faith [**?**]).

4.2.3 Digraphs and graphs

(**4.2.16**) A *digraph* is a triple (V, E, f) where V, E are (finite) sets and f a function $f : E \to V \times V$.

(**4.2.17**) Given a digraph $G = (V, E, f)$, we say V is the set of 'vertices' and E the set of 'edges'. An edge $e \in E$ with $f(e) = (a, a)$ is a 'loop'. If $f(e) = (a, b)$ then e is an edge 'on' (a, b) or from a to b.

An *edge colouring* of a digraph is a map from E to a set of 'colours', and hence a partition of E into same-coloured subsets.

(**4.2.18**) A digraph can be represented by a picture with a labeled node for each vertex and a directed labeled arc for each edge. Examples:

(4.2)

(**4.2.19**) A *simple digraph* is a digraph (V, E, f) in which f is an inclusion.

This amounts to saying that we can use a subset of $V \times V$ as the edge set. Thus we do not need labels on edges in a picture. That is, a simple digraph is just a relation on V.

(**4.2.20**) Given a digraph, if there is a proper path (along directed edges) from a to b then the *'distance'* from a to b is the minimum number of edges in such a path. (Note that this is not a true distance function. The distance from b to a may be different, for example.)

A digraph is *acyclic* if there is no proper path (along directed edges) from a to a for any $a \in V$.

(**4.2.21**) A digraph is *rooted* with root $r \in V$ if there is a vertex $r \in V$ such that every vertex is reachable by a directed path from r.

Note that if a digraph is acyclic then it has at most one root.

(**4.2.22**) Two simple digraphs (V, E, f) and (V', E', f') are *isomorphic* if there is a bijection $\psi : V \to V'$ such that (a, b) is in $f(E)$ iff $(\psi(a), \psi(b))$ is in $f'(E')$.

(**4.2.23**) We will say that two (not necessarily simple) digraphs are isomorphic if there is a bijection $\psi : V \to V'$ such that $f^{-1}(a, b)$ has the same order as $f'^{-1}(\psi(a), \psi(b))$ for all a, b (thus each of these pairs of sets could be placed in explicit bijection, but such a set of bijections is not necessarily given).

That is, two digraphs are isomorphic if their pictures can be 'morphed' into each other, using ψ, but ignoring the edge labels.

(**4.2.24**) Remark. We do not require the *finite* set condition for digraphs here. In practice our digraphs are either finite or inverse limits of sequences of finite graphs. This means in particular that there are only finitely many edges associated to any given pair of vertices, i.e. $f^{-1}(v, w)$ is always finite.

(**4.2.25**) Let G be a digraph with a countable vertex set. The *adjacency matrix* $M^G = A(G)$ is a vertex indexed square array such that entry M^G_{ij} is the number of edges from i to j in G.

Example from (4.2) above:

$$A(G_1) = \left(\begin{array}{c|ccc} & a & b & c \\ \hline a & 0 & 2 & 0 \\ b & 0 & 0 & 0 \\ c & 1 & 1 & 0 \end{array} \right)$$

(**4.2.26**) The opposite graph of a digraph has the same V and E but $f^{op}(e) = f(e)^{op}$ (i.e. if $f(e) = (a, b)$ then $f^{op}(e) = (b, a)$).

(**4.2.27**) A *graph* is a digraph that is isomorphic to its opposite.

(**4.2.28**) We say a graph is *connected* if for any pair of vertices there is a finite chain of edges connecting them.

(**4.2.29**) Example. Let G be a group and S a set of elements. The *Cayley graph* $\Gamma(G, S)$ is the digraph with vertex set G and an edge s_a on (a, b) whenever $b = as$ for some $s \in S$.

(**4.2.30**) Notes:

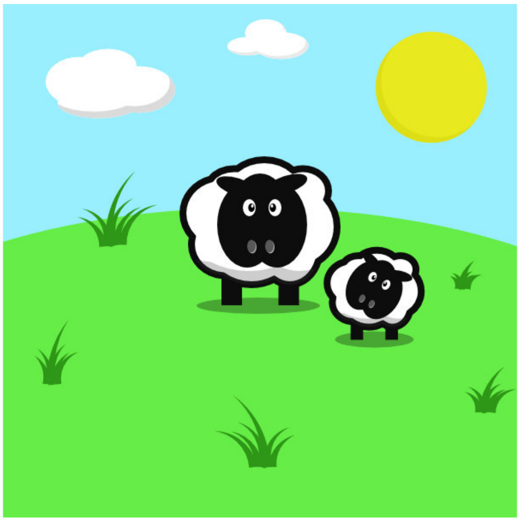
1. $s = a^{-1}b$ so there is at most one edge on (a, b), i.e. $\Gamma(G, S)$ is simple.
2. If $s \in S$ is an involution then edges involving s are effectively undirected. Some workers define S to include inverses (write this as $S = S^{-1}$), so again $\Gamma(G, S)$ is undirected.
3. We consider that S excludes the identity, so $\Gamma(G, S)$ is loop-free.
4. Some workers require that S generates G. Then $\Gamma(G, S)$ is connected. If $S = G$ then $\Gamma(G, S)$ is the complete graph.
5. If all generators are involutions (or $S = S^{-1}$) then the graph is effectively undirected by construction. However one can sometimes 'direct' such a $\Gamma(G, S)$, by using a *length function*... The root vertex is the identity 1, and there is a well-defined distance (really minimum distance, since there are undirected adjacent pairs) from 1 to any vertex g, denoted $l(g)$. If there is no edge between vertices of equal distance then we can 'direct' edges away from the root.

(**4.2.31**) Example. For $\Gamma(S_n, S)$ where S is the set of adjacent pair permutations, one can show that $l(gs) \neq l(g)$. See §**??**.

Chapter 5

The real field and geometry

For examples and motivation topology makes heavy use of properties of the real number field. These are familiar to us — the familiar properties of real arithmetic. But they are amazing and important. So let's review them a little. It is safe to skip this Chapter on first reading.



Our starting point here is the observation that the real line is a magical thing — meaning that it is amazing and familiar but hard to fully understand.

Here is a picture (of just a bit of it):

It is already quite amazing that I didn't need to tell you *which* bit that was. This is because, from one point of view, one chunk of the real line is much like another. If I mark a point anywhere on it, then that partitions the line into three parts: left of the point; the point; right of the point.

Notice that is not a property held by just any old set. If I have a random set and a pick an element, no further structual implications arise.

We often do mark a point on the real line. We call the first such marked point 0. We then have an additive structure on the line, with respect to which 0 is the identity. We can do 'slide-rule' addition, meaning we can compute $a + b$ by taking another copy of the line and sliding until the second copy of 0 is at a then looking at the position of the second copy of b. (Caveat: To do the slide we had to place the line in a bigger universe in which the second copy could also live! This was a physical-world operation rather than a maths one. To do it in maths-world we have to be a bit more patient.)

Notice that there was not actually anything special about the point 0 as far as the line was concerned. We chose it and made it special.

We often mark another special point on the line: we call it 1. Again this is a choice (except on the real line it should not be the same point as 0). We can use it as the identity of the $\times$ operation.

(**5.0.1**) Recall that the real numbers form a field, with $+$ and $\times$ as operations and 0 and 1 as their identities.

As usual we write $-a$ for the additive inverse of a.

(Aside: should we have a refresher on what a field is? Just in case, there are some reminders at the end of this Chapter.)

5.1 Ordered rings and fields

(**5.1.1**) An ordered field (or ring) F is one with a non-empty subset P_F (P for positive) that is closed under $+, \times$, does not contain 0, and for all $a \in F$ either a or $-a$ is in P_F.

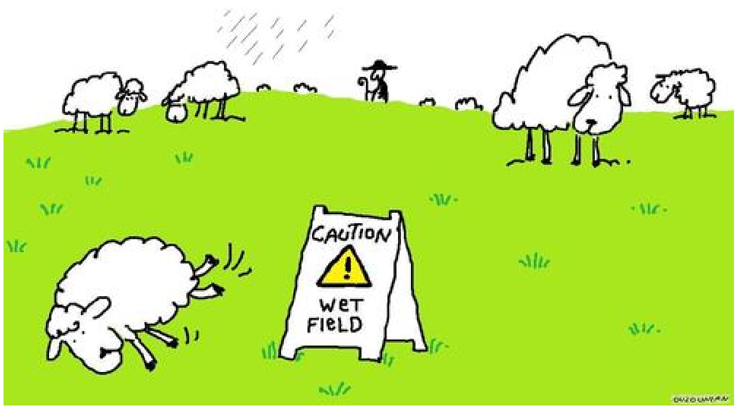
In an ordered field we write $a > b$ if $a + (-b) \in P_F$.

Is this a partial order on F? Is it true that $a > b$ and $b > c$ implies $a > c$?

(**5.1.2**) Anyway, an example is our real field $\mathbb{R}$, taking the positive numbers for $P_{\mathbb{R}}$. Indeed here $>$ has its usual meaning corresponding to ordering on the real line.

Another example is $\mathbb{Q}$.

Neither of these examples is really obvious. A simpler example is the ring of integers $\mathbb{Z}$, where $P_{\mathbb{Z}}$ is the natural numbers. The natural numbers (and zero) have the construction we already noted coming from set cardinalities — which also gives a natural order $\geq$; and addition (via union); and multiplication (via Cartesian product). We think of $\mathbb{Z}$ as the extension of $\mathbb{N}_0$ by another disjoint copy of $\mathbb{N}$ denoted $-\mathbb{N}$. Imposing the familiar arithmetic we have a ring. ...Which is an ordered ring.



(**5.1.3**) Given an ordered ring R and $r \in R$ then $|r|$ means the element from $\{r, -r\}$ that is in P_R (or else 0).

(**5.1.4**) When we pass from $\mathbb{Z}$ to its field of quotients $\mathbb{Q}$ we find that again we have an ordered ring (now field), with arithmetic and order playing nicely together.

5.1.1 Sequences and a construction of $\mathbb{R}$

See for example Mac Lane and Birkoff [**?**].

(**5.1.5**) In an ordered field F, suppose we have an infinite sequence $f_- = (f_1, f_2, f_3, ...)$. We say f_- converges to a limit $f \in F$ if for each $t \in P_F$ there exists an n such that $|f_k - f| < t$ for all $k \geq n$.

(**5.1.6**) Example. The sequence $(1, 1/2, 1/3, /1/4, ...)$ converges to 0 in $\mathbb{Q}$.

(**5.1.7**) Example. The rational sequence given by $f_1 = 1$ and then $f_{i+1} = \frac{(f_i^2+2)}{2f_i}$ has no rational limit.
(It is secretly converging to $\sqrt{2}$. Always rational, but approximating something in geometry that doesn't fit in the rationals.)
(Hint: make sure you know about the Newton–Raphson method...)

(**5.1.8**) A Cauchy sequence in ordered field F is a sequence $f_- = (f_1, f_2, f_3, ...)$ such that for any $t \in P_F$ there is an integer m such that $j, k > m$ implies $|f_j - f_k| < t$.

(**5.1.9**) Note that this is a bit like convergence, except that it doesn't require a 'number' to converge to. (Handy for us, since we want to use it to construct 'new' numbers, not old ones.)

(**5.1.10**) Our example 5.1.7 is a Cauchy sequence.

(**5.1.11**) Note that this Cauchy condition is nominally a weaker condition than converging to a limit (in the sense that converging to a limit implies Cauchy).

(**5.1.12**) A nice feature of the set $\mathbb{S}$ of Cauchy sequences of rational numbers is arithmetic:

If f_- and g_- are Cauchy then so are $f_- + g_-$ (defined in the obvious 'pointwise' way); and $f_- g_-$ and $-f_-$.

And if no $f_i = 0$ and the sequence does not tend to 0 then $\frac{1}{f_-}$ (to use a natural shorthand) is also Cauchy.

(**5.1.13**) Define a relation on the set of rational Cauchy sequences by $f_- \equiv g_-$ if $f_- - g_-$ converges to 0.

(**5.1.14**) **Proposition.** This $\equiv$ is an equivalence relation.

(**5.1.15**) **Proposition.** If $f_- \equiv g_-$ then $f_- + h_- \equiv g_- + h_-$ and $f_- h_- \equiv g_- h_-$.
(These results are far from obvious. Note that the Cauchy property for f_- in the case $t = 1$ implies that the sequence is bounded.)

(**5.1.16**) It follows that the quotient set $\mathbb{S}/\equiv$ is a ring. And indeed a field.

And indeed, defining $f_- > 0$ in $\mathbb{S}/\equiv$ to mean there is a $t > 0$ with $a_k > t$ for all sufficiently large k; then $\mathbb{S}/\equiv$ is an ordered field.

(**5.1.17**) So altogether we have an ordered field that contains an isomorphic copy of the rationals and also contains a square root of 2.

This is (one version of) the mathematicians real line.

(**5.1.18**) **Definition.** An ordered field F is 'complete' if every non-empty set of positive elements in F has a greatest lower bound in F.

(**5.1.19**) **Proposition.** The real numbers $\mathbb{R} = \mathbb{S}/\equiv$ form a complete ordered field.
(This is far from obvious. It is an exercise to try to prove.)

(**5.1.20**) There are other very nice but essentially equivalent constructions of $\mathbb{R}$.

(**5.1.21**) **Question.** Does this capture your intuition for what a mathematical model of the real real line should be?

(**5.1.22**) Notice that given that we have greatest lower bounds, we also have least upper bounds.

The least upper bound of a subset is sometimes called supremum or sup.
The GLB is sometimes called infimum or inf.

5.2 Arithmetic

(**5.2.1**) For $r \in \mathbb{R}$ we define $|r|$ to be the element of $\{r, -r\}$ that lies in $P_{\mathbb{R}}$ (as above) if $r \neq 0$, and define $|0| = 0$.

Note that $|r| = 0$ implies $r = 0$. Otherwise $|r| > 0$.

(**5.2.2**) For $a, b \in \mathbb{R}^n$ recall the notation

$$a.b = \sum_{i=1}^{n} a_i b_i$$

Note that $a.a > 0$. An interesting notation is to set $|a|^2 = a.a$.

An interesting property of $\mathbb{R}$ is that for $r \in \mathbb{R}$ if $r > 0$ then the equation $x^2 - r = 0$ has a positive solution in $\mathbb{R}$ (and also a negative solution). We write $\sqrt{r}$ for the positive solution. Then for $a \in \mathbb{R}^n$ the notation $|a|$ means the positive square root of $|a|^2$.

(**5.2.3**) Consider the inequality:

$$a \cdot b \leq \sqrt{|a|^2}\sqrt{|b|^2} \quad \forall\, a, b \in \mathbb{R}^n. \tag{5.1}$$

The inequality (5.1) is sometimes called the *Cauchy-Schwartz (CS) inequality.*
Proof. The sum of positive numbers is positive, so

$$\begin{aligned} 0 \;\leq\; & \sum_{i=1}^{n}\sum_{j=1}^{n}(a_i b_j - a_j b_i)^2 \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^2 b_j^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} a_j^2 b_i^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_i a_j b_j \\ &= 2\sum_{i=1}^{n} a_i^2 \sum_{j=1}^{n} b_j^2 - 2\left(\sum_{i=1}^{n} a_i b_i\right)^2 . \end{aligned}$$

So $|a|^2|b|^2 - (a \cdot b)^2 \geq 0$, and rearranging yields (5.1).

5.3 A little algebra

See for example Paul's "Rings, Polynomials and Fields" notes online for more on this Section.

(**5.3.1**) **Definition**. A **group** $(G, \cdot)$ consists of a set G and a map $\cdot : G \times G \to G$, $(g,h) \mapsto g \cdot h$ satisfying the following conditions:

(G1) (Associativity) $(g \cdot h) \cdot k = g \cdot (h \cdot k)$ for all $g, h, k \in G$.

(G2) (Identity) There exists an element $e \in G$ such that $g \cdot e = g = e \cdot g$ for all $g \in G$.

(G3) (Inverses) For every $g \in G$ there exists an element $g^{-1} \in G$ such that $g^{-1} \cdot g = e = g \cdot g^{-1}$.

(**5.3.2**) **Definition**. If in addition $g \cdot h = h \cdot g$ for every $g, h \in G$ the group is called **abelian** (or *commutative*).

(**5.3.3**) We call the dot map above a 'closed binary operation'.

(**5.3.4**) The integers with addition are a nice example of a group. ...But of course there is a lot more to integer arithmetic than just addition. ...Which thought takes us next to the notion of a 'ring'.

(**5.3.5**) A **ring** is a set K (say) together with two closed associative binary operations, $\cdot$ and $+$ (say), satisfying certain further axioms as follows.
Both operations have an identity — for the moment denoted 1 and 0 respectively (caveat: K may not be a set of numbers, so these may not be the usual 1 and 0).
The $+$ operation has inverses. (The 'additive' inverse of $a \in K$ is denoted $-a$.)
The group $(K, +)$ is abelian.
We have distributivity: $a \cdot (b + c) = a \cdot b + a \cdot c$ (and similarly the other way round).

(**5.3.6**) We may write K^* for $K \setminus \{0\}$.

(**5.3.7**) Example: the integers with addition and multiplication form a ring.

(**5.3.8**) A **field** is a ring in which $(K^*, \cdot)$ is an abelian group. (With inverse of $a \in K^*$ denoted a^{-1}.)

(**5.3.9**) Note that the integers do not form a field! But the rational numbers do.

5.4 (What about) Linearity and vector spaces

Another kind of 'space' is a vector space. How do topological spaces and vector spaces play together? ...

One of our favourite ideas (in mathematical modelling) is linearity — the vector space property of $\mathbb{R}^n$. It appears in various ways: "smoothness" and diffeomorphisms; Taylor series; approximation; convexity; ...

The space $\mathbb{R}^n$ is already a vector space! Others such as

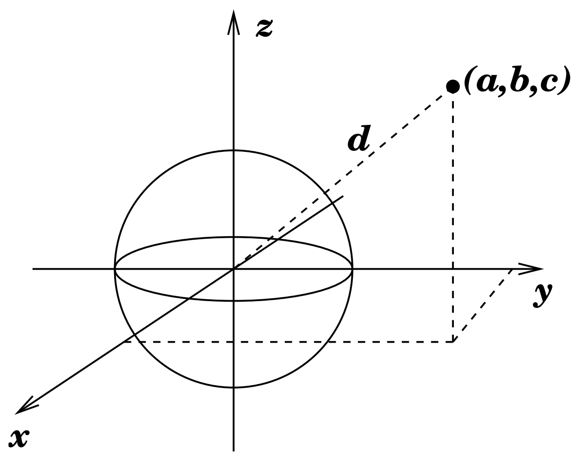
...

and

..

not obviously so (how add vectors or scalar-multiply?).

Ask me about:
The physical line and the real line and points on the physical line...
the circle
tangents on the sphere and linearisation and Lie theory...
...



Chapter 6

Metric Spaces

6.1 Metrics and metric spaces

In this section we assume you know quite a lot about the real numbers — the set $\mathbb{R}$ together with its properties as an ordered field. (Do not worry. These are properties with which you *are* familiar, like $3\pi > 0$; and that positive numbers have real square roots, as we'll see.)

Definition 6.1. Let X be a set. A function $d : X \times X \to \mathbb{R}$ is a **metric** on X if:

(M1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;

(M2) $d(x, y) = d(y, x)$ for all $x, y \in X$ (symmetry);

(M3) $d(x, y) + d(y, z) \geq d(x, z)$ for all $x, y, z \in X$ (triangle inequality).

The pair (X, d) is called a **metric space**.

Example 6.2. On any set X, the function

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

is a metric. Axioms (M1) and (M2) are satisfied directly. (M3) holds if $x = z$, for then $d(x, y) + d(y, z) \geq 0 = d(x, z)$. If $x \neq z$ then y cannot coincide with both x and z, so $d(x, y) + d(y, z) \geq 1 = d(x, z)$.

Example 6.3. Let $X = \mathbb{R}$ and let

$$d(x, y) = |x - y|.$$

(M1) and (M2) are satisfied here by (5.2.1). (M3) can be checked case-by-case. We can assume that $x \leq z$ without loss of generality. If $x \leq y \leq z$ then $d(x, z) = d(x, y) + d(y, z)$. If $y < x \leq z$ or $x \leq z < y$ then $d(x, z) < d(x, y) + d(y, z)$.

This d is thus a metric. It is sometimes called the **standard metric** on $\mathbb{R}$.

Example 6.4. Fix $n \in \mathbb{N}$. Let $X = \mathbb{R}^n$ and let

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

Here (M1) and (M2) follow from (5.2.2). To prove (M3), we can proceed as follows. Consider the inequality (5.1):

$$a \cdot b \leq \sqrt{|a|^2}\sqrt{|b|^2} \quad \forall\, a, b \in \mathbb{R}^n. \tag{6.1}$$

Using (5.1) we have that

$$\begin{aligned}(d(x, y) + d(y, z))^2 &= |x - y|^2 + |y - z|^2 + 2\sqrt{|x - y|^2}\sqrt{|y - z|^2} \\ &\geq |x - y|^2 + |y - z|^2 + 2(x - y) \cdot (y - z) \\ &= (x - y + y - z) \cdot (x - y + y - z) \\ &= d(x, z)^2.\end{aligned}$$

Example 6.5. Let $X = \mathbb{R}^n$ and let

$$d(x, y) = \max_{i=1,\ldots,n} |x_i - y_i|.$$

Then d is a metric. For (M3), suppose j is such that $d(x, z) = |x_j - z_j|$. We know from Example 6.3 that $|x_j - z_j| \leq |x_j - y_j| + |y_j - z_j|$. So

$$\begin{aligned}d(x, z) &\leq |x_j - y_j| + |y_j - z_j| \\ &\leq \max_{i=1,\ldots,n} |x_i - y_i| + \max_{i=1,\ldots,n} |y_i - z_i| \\ &= d(x, y) + d(y, z).\end{aligned}$$

6.2 Open balls in metric spaces

Definition 6.6. Let (X, d) be a metric space. An **open ball** is a subset $B_\varepsilon(x)$ of X of the form

$$B_\varepsilon(x) = \{y \in X \ : \ d(x, y) < \varepsilon\}$$

where x is any element of X and $\varepsilon > 0$ is any positive real number.

Lemma 6.7. *Let (X, d) be a metric space, let $x \in X$ and let $\delta > 0$. If $y \in B_\delta(x)$ then there exists an $\varepsilon > 0$ such that $B_\varepsilon(y) \subseteq B_\delta(x)$.*

Proof. Let $\delta > 0$. We need to choose $\varepsilon > 0$ such that $z \in B_\varepsilon(y) \Rightarrow z \in B_\delta(x)$. By the triangle inequality, any $z \in B_\varepsilon(y)$ satisfies

$$d(z, x) \leq d(z, y) + d(y, x) < \varepsilon + d(y, x)$$

So if $\varepsilon = \delta - d(y, x)$ then every $z \in B_\varepsilon(y)$ satisfies $d(x, z) < \delta$ as required. This choice of ε is positive, because $d(x, y) < \delta$. $\square$

Definition 6.8. Let (X, d) be a metric space. We say that $U \subseteq X$ is ***d*-open** if and only if: for all $x \in U$ there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$.

Lemma 6.9. *In a metric space (X, d), open balls are d-open.*

Proof. Follows from Lemma 6.7. $\square$

Lemma 6.10. *Let (X, d) be a metric space. Denote*

$$\tau_d \ := \ \{U \subseteq X : U \text{ is } d\text{-open}\} \ \subseteq \ \mathtt{P}(X)$$

Then we have

(T0) $\emptyset \in \tau_d$

(T1) $X \in \tau_d$.

(T2) *If $U \in \tau_d$ and $V \in \tau_d$ then $U \cap V \in \tau_d$.*

(T3) *If $\{U_\lambda\}_{\lambda \in \Lambda}$ is a family of subsets of X such that $U_\lambda \in \tau_d$ for all $\lambda \in \Lambda$, then $\bigcup_{\lambda \in \Lambda} U_\lambda \in \tau_d$.*

Proof. For (T0) there is nothing to show. For (T1) it is true by construction.
(T2): if $x \in U \cap V$ there exits $\varepsilon_1, \varepsilon_2 > 0$ such that $B_{\varepsilon_1}(x) \subset U$ and $B_{\varepsilon_2}(x) \subset V$. Letting $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ gives $B_\varepsilon(x) \subset U \cap V$.
(T3): exercise. $\square$

Definition 6.11. We call the set of subsets τ_d the **metric topology** on metric space (X, d).

Our idea in what follows will be to turn things around and consider (T0-3) not as consequences of a construction, but as conditions on a set of subsets.

The conditions (T2) and (T3) have a different flavour. What would happen if we consider arbitrary intersections in (T2) instead of only pairwise? Consider for example the Euclidean metric above. We claim:

$$\cap_{e>0} B_e(x) \ = \ \{x\}$$

By definition this intersection is the set of elements that are in every such ball around x. The element x itself is in every such ball. But for any $y \neq x$ we have $d(y, x) > 0$ so one of the smaller balls does not contain y.

The above argument says that if we modified (T2) to allow arbitrary intersections then each $\{x\}$ would be an open set. Since we allow arbitrary unions, then every subset would be open.

Definition 6.12. Let X be a metric space. A subset A of X is called ***d*-closed** if $X \setminus A$ is open.

6.3 Equivalent metrics

Definition 6.13. Two metrics d and d' on a set X are called **equivalent** if there exist real numbers $D \geq C > 0$ such that

$$Cd'(x,y) \leq d(x,y) \leq Dd'(x,y) \quad \forall x, y \in X.$$

Lemma 6.14. *Let d and d' be equivalent metrics on a set X. Then $U \subset X$ is d-open if and only if $U \subset X$ is d'-open. Thus $\tau_d = \tau_{d'}$.*

Proof. We need to show that any set which is open with respect to one metric is also open with respect to the other.

Suppose that $U \subset X$ is d'-open. This means that for each $x \in U$ there exists an $\varepsilon > 0$ such that $B'_\varepsilon(x) \subset U$, where $B'_\varepsilon(x)$ denotes the open ball in the metric d'. We must construct an open ball in the metric d that contains x and is contained in U.

By definition $B_{C\varepsilon}(x) \subset B'_\varepsilon(x)$; if $z \in B_{C\varepsilon}(x)$ then $d(x,z) \leq C\varepsilon \implies Cd'(x,z) \leq d(x,z) \leq C\varepsilon$ so $z \in B'(x)$. Thus we have shown that U is open with respect to d.

Now, if U is d-open a similar argument (exercise) shows that it is also d'-open.

□

Example 6.15. The metrics on $\mathbb{R}^n$ defined in examples 6.5 and 6.4 are equivalent and therefore define the same set of d-open sets. This is because

$$\left(\max_{i=1,\ldots,n} |x_i - y_i|\right)^2 = \max_{i=1,\ldots,n} |x_i - y_i|^2 \leq \sum_{i=1}^n (x_i - y_i)^2$$

and

$$\sum_{i=1}^n (x_i - y_i)^2 \leq n \max_{i=1,\ldots,n} |x_i - y_i|^2 = n\left(\max_{i=1,\ldots,n} |x_i - y_i|\right)^2.$$

6.4 Metric continuous functions

Recall that $f : \mathbb{R} \to \mathbb{R}$ is said to be continuous at x if and only if: for all $\varepsilon > 0$ there exists $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$.

We can write this in a different way using the standard metric on $\mathbb{R}$, $d(x,y) = |x - y|$. Function $f : \mathbb{R} \to \mathbb{R}$ is continuous at x if and only if: for all $\varepsilon > 0$ there exists $\delta > 0$ such that

(I) $d(x,y) < \delta$ implies $d(f(x), f(y)) < \varepsilon$.

(II) $f(B_\delta(x)) \subset B_\varepsilon(f(x))$.

(III) $B_\delta(x) \subset f^{-1}(B_\varepsilon(f(x)))$.

Definition 6.16. A function $f : (X, d) \to (Y, d')$ between metric spaces is **metric continuous** at $x \in X$ if and only if: for all $\varepsilon > 0$ there exists $\delta > 0$ such that

$$B_\delta(x) \subset f^{-1}(B'_\varepsilon(f(x))).$$

Furthermore we say that $f : X \to Y$ is metric continuous if and only if it is continuous for all $x \in X$.

Remark 6.17. NB: This definition is the usual definition of continuity when X, Y are the vector spaces $\mathbb{R}^n, \mathbb{R}^m$ equipped with the standard metrics, only now it is suitable for general metric spaces.

The notation B' denotes an open ball with respect to the metric on Y, d'. But from here we drop this notation since it's clear from the context in which space the balls lie!

Theorem 6.18. *$f : (X, d) \to (Y, d')$ is metric continuous if and only if for all d'-open $U \subset Y$, $f^{-1}(U) \subset X$ is d-open.*

Proof. Suppose that for all open $U \subset Y$, $f^{-1}(U) \subset X$ is open. Then for all $\varepsilon > 0$ we know that $f^{-1}(B_\varepsilon(f(x))) \subset X$ is open, since $B_\varepsilon(f(x))$ is open in Y. By definition (since $x \in f^{-1}(B_\varepsilon(f(x))))$ this implies there exists $\delta > 0$ so that $B_\delta(x) \subset f^{-1}(B_\varepsilon(f(x)))$, thus f is metric continuous.

For the converse suppose that f is metric continuous. Let $U \subset Y$ be open so we want to prove that $f^{-1}(U)$ is open. Pick $x \in f^{-1}(U)$ giving $f(x) \in U$. Since U is open, there exists $\varepsilon > 0$ so that $B_\varepsilon(f(x)) \subset U$ and since f is metric continuous, there exists $\delta > 0$ so that $B_\delta(x) \subset f^{-1}(B_\varepsilon(f(x))) \subset f^{-1}(U)$ so we are done (remember that, since X is a metric space, $f^{-1}(U)$ is open if and only if, for all $x \in f^{-1}(U)$ there exists $\delta > 0$ so that $B_\delta(x) \subset f^{-1}(U)$). □

Chapter 7

Topological spaces

7.1 Definition

To define continuity on metric spaces we only need the notion of open sets (Theorem 6.18).

A topology on a set X is a collection of subsets τ which we **declare** to be 'open', requiring only that this collection satisfies some axioms. (Notice (after reading the definition!) that the 'metric topology' is indeed a topology (cf. Lemma 6.10).)

Definition 7.1. Let X be a set. A **topology** on X is a set τ of subsets of X having the properties that

(T0) $\emptyset \in \tau$

(T1) $X \in \tau$.

(T2) If $U \in \tau$ and $V \in \tau$ then $U \cap V \in \tau$.

(T3) If $\{U_\lambda\}_{\lambda \in \Lambda}$ is a family of subsets of X such that $U_\lambda \in \tau$ for all $\lambda \in \Lambda$, then $\bigcup_{\lambda \in \Lambda} U_\lambda \in \tau$.

A **topological space** is a pair (X, τ) where X is a set and τ is a topology on X. The elements of X are called the **points** of the space. The elements of τ are called the **open sets** of the space, and are said to be **open** in (X, τ).

Remark 7.2. It follows by induction from (T2) that the intersection of any *finite* collection of open sets is open. Note however that the intersection of an *infinite* collection of open sets need not be open. By contrast, in axiom (T3) the indexing set Λ could be finite, countably infinite, or uncountably infinite.

Example 7.3. If as in examples 6.3 and 6.5 $X = \mathbb{R}^n$ and d is the standard (or Euclidean) metric, the associated metric topology τ_d is a topology, sometimes called the **standard topology**.

Remark. Normally when we define a class of algebraic structures (the concept of group, say; or vector space) then we *study and understand* such things by studying the relationships (maps, morphisms) between them. The morphisms we use are the ones that preserve the axiomatic behaviour of the structure. So for a group homomorphism f we require $f(a.b) = f(a) * f(b)$ and so on. (Such a behaviour-preserving identity is sometimes called a 'commuting square'.) For topological spaces it is a *little* different — because a certain kind of special function between the underlying sets (coming from the extrinsic analytical notion of continuous maps) was really the whole point of making the axiomatization in the first place. So there are more than one kind of morphism that we could think of for topological spaces. Rather than starting with commuting squares, we start with continuity.

Definition 7.4. Let X and Y be two topological spaces and let $f : X \to Y$ be a function. Function f is called **continuous** if for every open subset $A \subset Y$, $f^{-1}(A)$ is an open subset of X.

Remark 7.5. Theorem 6.18 tells us that if X and Y are metric spaces equipped with the metric topologies then $f : X \to Y$ is continuous if and only if it is metric continuous. So there is only one notion of continuity.

7.1.1 More Examples

Example 7.6. Let X be any set and let $\tau = \{\emptyset, X\}$. Then certainly τ satisfies (T1). (T2) holds because $\emptyset \cap X = \emptyset \in \tau$, and (T3) holds because $\emptyset \cup X = X \in \tau$. So τ is a topology on X; it is known as the *trivial topology* or **indiscrete topology**.

Example 7.7. Let X be any set and let τ be the power set of X (i.e. the set of all subsets of X). You can check for yourself that τ satisfies the axioms (T1), (T2) and (T3). So τ defines a topology on X; it is known as the **discrete topology** on X.

Example 7.8. Let (X, d) be as in Example 6.2. Then the associated metric topology is the discrete topology.

To prove this, we need to first consider what open balls look like in this metric. If $\varepsilon > 1$ then $B_\varepsilon(x)$ is equal to the whole of X, whereas if $\varepsilon \leq 1$ $B_\varepsilon(x) = \{x\}$ is a singleton (a set with one element). Thus the collection of all open balls consists of all singletons in X, together with X itself. In particular *any* subset of X is open (since it is either empty, or a union of singletons).

Example 7.9. Let $X = \{1, 2\}$ and let $\tau = \{\emptyset, \{1\}, \{1, 2\}\}$. Then τ is a topology on X. Checking this is a matter of going through the axioms once more. (T1) is immediate. (T2) holds because

$$\emptyset \cap \{1\} = \emptyset \in \tau, \quad \emptyset \cap \{1,2\} = \emptyset \in \tau, \quad \{1\} \cap \{1,2\} = \{1\} \in \tau.$$

(T3) holds, because

$$\emptyset \cup \{1\} = \{1\} \in \tau, \quad \emptyset \cup \{1,2\} = \{1,2\} \in \tau, \quad \{1\} \cup \{1,2\} = \{1,2\} \in \tau, \quad \emptyset \cup \{1\} \cup \{1,2\} = \{1,2\} \in \tau.$$

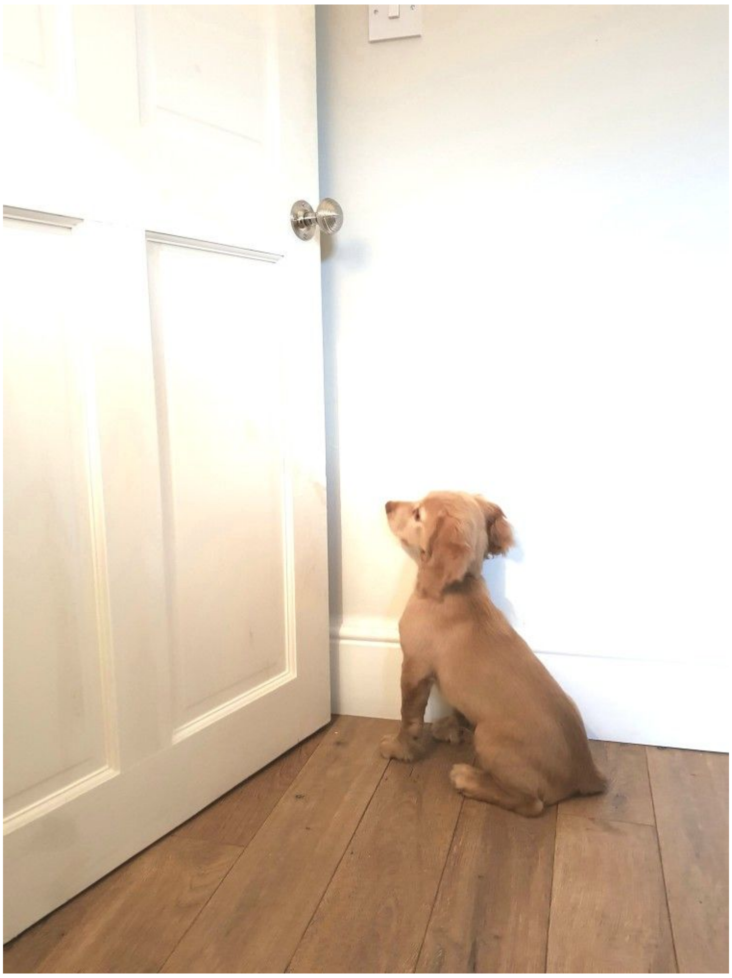
Note that this contruction made no use of any intrinsic properties of the elements of the set $\{1, 2\}$. We could have replaced it with $\{a, b\}$ or $\{elephant, dog\}$ and everything would have gone through analogously. Indeed we could have reversed the roles of 1 and 2.

Example 7.10. Let X be any set and let

$$\tau = \{A \subset X \ : \ X \setminus A \text{ is finite}\} \cup \{\emptyset\}.$$

Then τ defines a topology on X – checking this is an Exercise. This topology is known as the **finite complement topology** or *general Zariski topology*.

7.1.2 Closed sets



Definition 7.11. Let X be a topological space. A subset A of X is called **closed** if $X \setminus A$ is open.

Example 7.12. In the discrete topology on a set X, any subset is both open and closed.

Example 7.13. In the indiscrete topology on set X, the only closed subsets are X and $\emptyset$. All other sets are neither open nor closed.

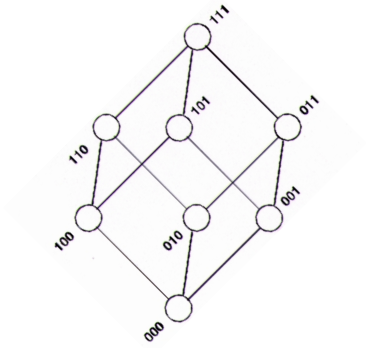
The above examples show that sets can be open, closed, both, or neither!

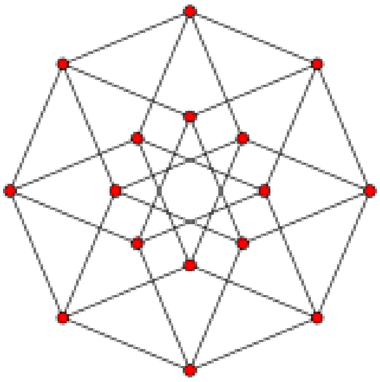
Proposition 7.14. *Let X be a topological space. Then*

1. *X and $\emptyset$ are closed;*
2. *The union $A \cup B$ of two closed sets A and B is closed;*
3. *If $\{A_\lambda\}_{\lambda\in\Lambda}$ is any indexed family of closed subsets of X, the intersection $\bigcap_{\lambda\in\Lambda} A_\lambda$ is closed.*

Proof. These all follow from the axioms for a topological space (using de Morgan's rules for the second two). □

7.1.3 Coarser/finer





Recall that if X is a set then $\mathtt{P}X$ is the set of subsets. (The first figure above codifies $\mathtt{P}\{1,2,3\}$, for example. The vertex 111 means $\{1,2,3\}$; then 011 means $\{2,3\}$ and so on.) Thus a topology on X is a (special) element of $\mathtt{PP}X$. (The second figure above is a schematic for $\mathtt{P}\{1,2,3,4\}$ and hence indirectly for $\mathtt{PP}\{1,2\} = \mathtt{P}\{\emptyset, \{1\}, \{2\}, \{1,2\}\}$.)

If we know two different topologies on the same set, then it is useful to be able to compare them:

Definition 7.15. Let τ and σ be two topologies on a set X. If τ is a subset of σ (i.e. if A is any subset of X then $A \in \tau \Rightarrow A \in \sigma$) then τ is said to be **coarser** than σ, and σ is said to be **finer** than τ.

It follows from this definition that the discrete topology on X is finer than any other topology on X. Similarly, axiom (T1) implies that the indiscrete topology is coarser than any other topology.

7.2 Neighbourhoods and separated spaces

Definition 7.16. Let X be a topological space and let $x \in X$. A **neighbourhood** U of x is an open set $U \subset X$ such that $x \in U$.

Definition 7.17. A topological space X is called a **separated (or Hausdorff) space** if

(H) for each pair x, y of distinct points in X there exist neighbourhoods U of x and V of y which are disjoint.

The Hausdorff condition (H) will allow us to prove strong theorems for separated spaces that do not apply to non-separated spaces.

Theorem 7.18. *Metric topologies are separated.*

Proof. Let (X, d) be a metric space equipped with the metric topology. Suppose that x_1, x_2 are two distinct points in X. Then $d(x_1, x_2) > 0$. Let $\varepsilon = d(x_1, x_2)/2$, $B_1 = B_\varepsilon(x_1)$ and $B_2 = B_\varepsilon(x_2)$. Then certainly $x_1 \in B_1$ and $x_2 \in B_2$. It remains to check that $B_1 \cap B_2 = \emptyset$. Suppose for contradiction that $B_1 \cap B_2$ is non-empty, and let $y \in B_1 \cap B_2$. Then, by the triangle inequality,

$$d(x_1, x_2) \leq d(x_1, y) + d(y, x_2) < \varepsilon + \varepsilon = d(x_1, x_2).$$

It cannot happen that $d(x_1, x_2) < d(x_1, x_2)$, so $B_1 \cap B_2$ must be empty. $\square$

It is not true that every separated topology is induced by a metric.

7.3 Constructing new topologies from old

Let X be a set. Let $\tau \in \mathsf{PP}X$. Define

$$\mathsf{C}_\cup : \mathsf{PP}X \to \mathsf{PP}X \tag{7.1}$$

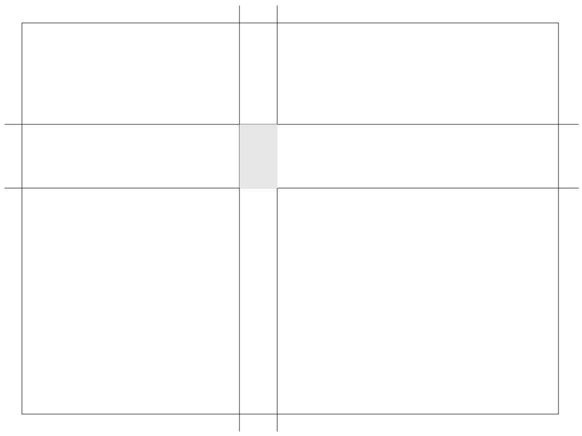
by $\tau \mapsto$ union-closure: τ together with all unions formed from subsets of τ (including 'union' over no elements).

Example: Let P_1X denote the set of single-element subsets. Then $\mathsf{C}_\cup(\mathsf{P}_1X) = \mathsf{P}(X)$.
Example: If τ is a topology then $\mathsf{C}_\cup(\tau) = \tau$.

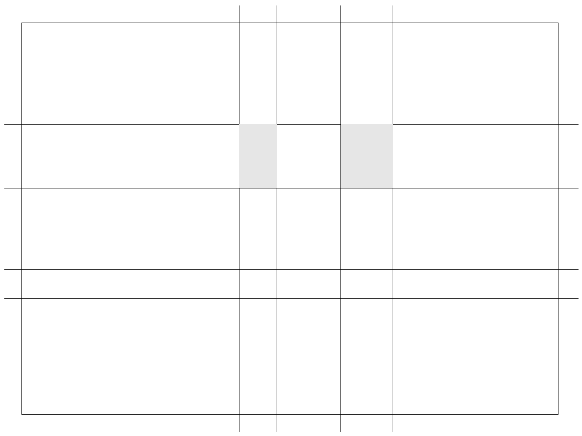
7.3.1 Asides on Cartesian products

(**7.3.1**) Note that if X, Y are sets and $S \subset X$ and $T \subset Y$ then $S \times T \subset X \times Y$.

Example: Here $X \times Y$ is represented by the big rectangle built from two intervals X (horizontal say) and Y; and S and T are the two smaller intervals marked on them; and $S \times T$ is shaded:



A collection of 'overlapping' subsets of X is not so easy to illustrate. But we can do a couple of non-intersecting ones...



Let's call them $S, S' \subset X$ and $T, T' \subset Y$. Then the union closures of the two sets of subsets separately are $\mathsf{C}_\cup(\{S, S'\}) = \{\emptyset, S, S', S \cup S'\}$ and $\{\emptyset, T, T', T \cup T'\}$.

Now consider $\{S, S'\} \times \{T, T'\} = \{(S, T), (S, T'), (S', T), (S', T')\}$. We can understand (S, T) as giving rise to a subset $S \times T$ of $X \times Y$.

Now let σ be a union-closed subset of the set of subsets $\mathsf{P}X$ and τ a union-closed subset of $\mathsf{P}Y$. Define $\beta_{\sigma\tau} = \{U \times V \mid U \in \sigma, V \in \tau\}$. Is $\beta_{\sigma\tau}$ union-closed?

7.3.2 Product and subspace topologies

Definition 7.19. /Proposition. Let (X, σ) and (Y, τ) be two topological spaces. We claim that the set

$$\beta_{\sigma\tau} = \{U \times V \mid U \in \sigma, V \in \tau\}$$

is WIC. The corresponding **product topology** on $X \times Y$ is that generated by $\beta_{\sigma\tau}$. (That is, the topology is $\mathsf{C}_{\cup}\beta_{\sigma\tau}$.)

Proof. Let us give two ways to do this. First using Proposition 3.1.16; then a direct proof.

Proof 1: By Prop.3.1.16 it is enough to show that $\beta_{\sigma\tau}$ is WIC. Consider $A, B \in \beta_{\sigma\tau}$, and $(x, y) \in A \cap B$. We have to show that there is an element in $\beta_{\sigma\tau}$ that contains (x, y) and is a subset of $A \cap B$.
Note that because $A \in \beta_{\sigma\tau}$ there exist U and V such that $A = U \times V$. Let us fix a choice of U and V for each A, denoted U_A and V_A respectively (using the axiom of choice; it will not matter exactly which ones we choose, just that some can be chosen). Then $(x, y) \in A$ implies $x \in U_A \in \sigma$ and $y \in V_A \in \tau$. So $(x, y) \in A \cap B$ implies $x \in U_A$ and $x \in U_B$, so $x \in U_A \cap U_B$ and similarly $y \in V_A \cap V_B$. So $(x, y) \in (U_A \cap U_B) \times (V_A \cap V_B)$.
We also have that

$$(U_A \cap U_B) \times (V_A \cap V_B) \subseteq U_A \times V_A = A$$

by (7.3.1), and similarly $(U_A \cap U_B) \times (V_A \cap V_B) \subseteq U_B \times V_B = B$, so

$$(U_A \cap U_B) \times (V_A \cap V_B) \subseteq A \cap B.$$

But $U_A \cap U_B \in \sigma$ since σ obeys (T2); and $V_A \cap V_B \in \tau$. So $(U_A \cap U_B) \times (V_A \cap V_B) \in \beta_{\sigma\tau}$. Since $(x, y) \in (U_A \cap U_B) \times (V_A \cap V_B)$ and $(U_A \cap U_B) \times (V_A \cap V_B) \subseteq A \cap B$, this element satisfies the requirement. Done.

Proof 2: For the definition to make sense we must explain why the collection $C_{\cup}\beta_{\sigma\tau}$ of subsets of $X \times Y$ is a topology. (T1) and (T3) are straightforward to check using the definition (exercise).

For (T2) we need to show that if W_1, W_2 are in $C_{\cup}\beta_{\sigma\tau}$ then so is $W_1 \cap W_2$. For this it would be enough to show that for every $(x, y) \in W_1 \cap W_2$ we have a $U \in \sigma$ and $V \in \tau$ such that $(x, y) \in U \times V$ and $U \times V \subseteq W_1 \cap W_2$ (enough because of the union closure). With this aim in mind then given $(x, y) \in W_1 \cap W_2$ we will try to come up with a suitable U and V. Each of W_1 and W_2 is a union of elements from $\beta_{\sigma\tau}$, so in particular every element (x, y) of W_1 lies in some $U_1 \times V_1$ say, where the factors are open in their respective spaces, and $U_1 \times V_1$ is contained in W_1. And similarly for some $U_2 \times V_2$, say, in W_2. So let $U_1, U_2 \subset X$ open and $V_1, V_2 \subset Y$ open be such that $(x, y) \in U_1 \times V_1 \subset W_1$ and $(x, y) \in U_2 \times V_2 \subset W_2$. Now set $U = U_1 \cap U_2$ (which is open in X), and note that $x \in U$. Similarly set $V = V_1 \cap V_2 \ni y$ (open in Y). Altogether then, we have

$$(x, y) \in U \times V \subset W_1 \cap W_2$$

as required. □

Example 7.20. Consider $\mathbb{R}^p$ and $\mathbb{R}^q$ for $p, q \in \mathbb{Z}_+$ and $p + q = n \geq 2$. Let each space be equipped with the standard topology. Then the product topology on $\mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^q$ coincides with the standard topology on $\mathbb{R}^n$.

Let τ_n denoted the product topology and τ the standard topology. We will show that $W \subset \mathbb{R}^n$ is τ_n-open (i.e. in τ_n) if and only if it is τ-open.

First suppose that W is τ_n-open and pick $x \in W$. We need to show that there exists $\varepsilon > 0$ so that $B^n_\varepsilon(x) \subset W$, where the open ball is defined with respect to the standard metric on $\mathbb{R}^n$. Since W is τ_n-open there exist open $U \subset \mathbb{R}^p$ and open $V \subset \mathbb{R}^q$ such that $x = (x_1, x_2) \in U \times V \subset W$.

Since they are both open there exits $\varepsilon_1, \varepsilon_2 > 0$ such that $B^p_{\varepsilon_1}(x_1) \subset U$ and $B^q_{\varepsilon_2}(x_2) \subset V$. Set $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ and notice that $B^n_\varepsilon(x) \subset B^p_{\varepsilon_1}(x_1) \times B^q_{\varepsilon_2}(x_2) \subset U \times V \subset W$ and we are done. The first inclusion follows since if $y = (y_1, y_2) \in B^n_\varepsilon(x) = B^n_\varepsilon((x_1, x_2))$ we have

$$\min\{\varepsilon_1, \varepsilon_2\}^2 = \varepsilon^2 > |y - x|^2 = |y_1 - x_1|^2 + |y_2 - x_2|^2,$$

giving $y_1 \in B^p_{\varepsilon_1}(x_1)$ and $y_2 \in B^q_{\varepsilon_2}(x_2)$.

The converse is similar so I won't spell it out. If W is τ-open. Pick $x = (x_1, x_2) \in W$ and notice that we can choose $\varepsilon > 0$ so that $B^n_\varepsilon(x) \subset W$. Now set $\varepsilon_1 = \frac{\varepsilon}{\sqrt{2}}$ and $\varepsilon_2 = \frac{\varepsilon}{\sqrt{2}}$ to give $B^p_{\varepsilon_1}(x_1) \times B^q_{\varepsilon_2}(x_2) \subset B^n_\varepsilon(x) \subset W$ since if $y = (y_1, y_2) \in B^p_{\varepsilon_1}(x_1) \times B^q_{\varepsilon_2}(x_2)$ then

$$\varepsilon^2 = \varepsilon_1^2 + \varepsilon_2^2 > |y_1 - x_1|^2 + |y_2 - x_2|^2 = |y - x|^2.$$

Thus setting $U = B^p_{\varepsilon_1}(x_1)$, $V = B^q_{\varepsilon_2}(x_2)$ gives $x \in U \times V \subset W$ and W is τ_n-open.

Definition 7.21. / Proposition. Let (X, τ) be a topological space and let $A \subset X$ be a subset of X. The **subspace topology** on A is

$$\tau_A = \{A \cap U \ : \ U \in \tau\}, \tag{7.2}$$

and the topological space (A, τ_A) is called a **subspace** of (X, τ).

Proof. To justify this definition we must explain why the collection τ_A of sets is a topology. Axiom (T0/1) is satisfied because $A = A \cap X$ and $\emptyset = A \cap \emptyset$ both belong to τ_A. The set τ_A satisfies axiom (T2) because τ does: if $A \cap U$ and $A \cap V$ are members of τ_A then their intersection $A \cap U \cap V$ is again in τ_A, because $U \cap V \in \tau$. Similarly, τ_A satisfies axiom (T3) because τ does: if $A \cap U_\lambda \in \tau_A$ is a collection of sets in τ_A indexed by $\lambda \in \Lambda$ then

$$\bigcup_{\lambda \in \Lambda} (A \cap U_\lambda) = A \cap \left(\bigcup_{\lambda \in \Lambda} U_\lambda \right) \in \tau_A.$$

Done.

Example 7.22. Consider $\mathbb{R}^2$ with the Euclidean metric topology, and consider the subset $\{(x, 0) | x \in \mathbb{R}\}$.
This subset can be considered as a copy of the real line in $\mathbb{R}^2$ — sitting on the x-axis, and it has a topology inherited from the usual metric topology on $\mathbb{R}$. One can check that this is the same as the subspace topology.

Proposition 7.23. *Let A be a subspace of a topological space X, and let $B \subset A$. Then B is closed in A if and only if there exists a closed subset C of X such that $B = C \cap A$.*

Proof. See exercise sheet 2. □

The following lemma gives a useful way to identify open sets in some special circumstances:

Lemma 7.24. *Let X be a topological space and let Y be an open subset of X. If $U \subset Y$ is open in the subspace topology then U is an open subset of X.*

Proof. Let $U \subset Y$ be any open subset of Y. Then by definition there exists an open subset $V \subset X$ such that $U = Y \cap V$. Then U must be open, because both Y and V are. □

Remark 7.25. The analogous result to Lemma 7.24 holds for closed subsets of closed subspaces as an easy corollary to Proposition 7.23: i.e. if $B \subset X$ is closed, and $C \subset B$ is closed in B (with the subspace topology), then $C \subset X$ is closed.

Recall that separated spaces are special kinds of topological spaces that satisfy the additional axiom (H). The operations of product and subspace preserve the Hausdorff property:

Proposition 7.26. *Any subspace of a separated space is separated. The product of any two separated spaces is separated.*

Proof. First suppose that X is a separated space and that A is a subspace. Let $x, y \in A$ be two points such that $x \neq y$. Since X is separated there exist open sets U, V such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$. Then the open subsets $A \cap U$ and $A \cap V$ satisfy the condition that $x \in A \cap U$, $y \in A \cap V$, and $(A \cap U) \cap (A \cap V) = \emptyset$. So A is separated.

The case of a product is left as an exercise! □

7.4 More on Continuous functions on topological spaces

Example 7.27. Constant functions are continuous.

Let X, Y be two topological spaces, pick a point $c \in Y$ and let $f(x) = c \; \forall x \in X$. Now suppose that $U \subset Y$ is an open subset of Y. Then either $c \in U$ and $f^{-1}(c) = X$, or $c \notin U$ and $f^{-1}(U) = \emptyset$. In either case, $f^{-1}(U)$ is open, so f must be continuous.

Example 7.28. The identity function is continuous.

Let X be a topological space and let $f(x) = x \; \forall x \in X$. Suppose that $U \subset X$ is an open subset of X. Then $f^{-1}(U) = U$ so $f^{-1}(U)$ is open. Thus f is continuous

Example 7.29. Let Y be any topological space, let X be a set equipped with the discrete topology, and let $f : X \to Y$. Then for any open set $U \subset Y$, $f^{-1}(U)$ is open, since all subsets of X are open by definition. Therefore f is continuous.

Example 7.30. Let X be any topological space, let Y be a set equipped with the indiscrete topology, and let $f : X \to Y$. The only open sets in Y are Y and $\emptyset$: $f^{-1}(Y) = X$ and $f^{-1}(\emptyset)$ are both open, so f is continuous.

Proposition 7.31. *Let X, Y be topological spaces. Then $f : X \to Y$ is continuous if and only if for every closed subset C of Y, $f^{-1}(C)$ is closed in X.*

Proof. See exercise sheet 2. □

Proposition 7.32. *Let X, Y and Z be topological spaces and let $f : X \to Y$ and $g : Y \to Z$ be continuous functions. Then*

$$g \circ f : X \to Z$$

is continuous.

Proof. Suppose that $U \subset Z$ is open. Since g is continuous, $g^{-1}(U) \subset Y$ is open, and since f is continuous $(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U))$ is open. Thus $g \circ f$ is continuous. □

Recall that we sometimes write $Hom(S, T)$ for the set of functions from set S to set T. We may now also write $Top((S, \sigma), (T, \tau))$ for the set of continuous functions from a topological space (S, σ) to topological space (T, τ).

In practice people often leave off the explicit declaration of the topology, in cases where there is a 'natural' one. So for example as a topological space $[0, 1]$ would be assumed to be given by the subspace topology derived from the Euclidean metric topology on $\mathbb{R}$. Then

$$Top([0, 1], (T, \tau))$$

is an interesting kind of set! ...A set of 'paths' in T.

7.4.1 Some toolkits for continuous functions

Proposition 7.33. *Suppose that A is a subspace of a topological space X. Then*

(a) the inclusion map $\iota_A : A \to X$, $x \mapsto x$ is continuous.

(b) if Y is any topological space and $f : X \to Y$ is continuous, the restriction $f_A : A \to Y$ of f to A is continuous.

(c) if Z is any topological space and $f : Z \to X$ satisfies $f(Z) \subset A$, then $f : Z \to X$ is continuous $\iff$ $f : Z \to A$ is continuous.

Proof. For part (a), suppose that U is an open subset of X. Then $\iota_A^{-1}(U) = A \cap U$. This set is by definition an open subset of A, so ι_A is continuous.

For part (b), note that $f_A = f \circ \iota_A$. So f_A is continuous by Proposition 7.32.

For part (c), ($\implies$), if $U \subset A$ is open then $U = A \cap \tilde{U}$ for $\tilde{U} \subset X$ open. Thus $f^{-1}(U) = f^{-1}(A \cap \tilde{U}) = f^{-1}(\tilde{U})$ since $f(Z) \subset A$. So $f^{-1}(U) \subset Z$ is open.

For ($\impliedby$), if $\tilde{U} \subset X$ is open, then again by the assumption on f we have $f^{-1}(\tilde{U}) = f^{-1}(\tilde{U} \cap A)$. But $\tilde{U} \cap A \subset A$ is open by definition, to $f^{-1}(\tilde{U}) \subset Z$ is open. □

Proposition 7.34. *Let X and Y be two topological spaces. Then*

(a) the projection maps $\pi_X : X \times Y \to X$, $(x, y) \mapsto x$ and $\pi_Y : X \times Y \to Y$, $(x, y) \mapsto y$ are continuous.

(b) if Z is any topological space then $f : Z \to X \times Y$ is continuous if and only if $\pi_X \circ f$ and $\pi_Y \circ f$ are both continuous.

Proof. For part (a), suppose that U is an open subset of X. Then $\pi_X^{-1}(U) = U \times Y$. This is clearly open in $X \times Y$.[1] Thus π_X is continuous; a similar argument shows that π_Y is continuous.

For the "only if" part of (b), suppose that $f : Z \to X \times Y$ is continuous. Then Proposition 7.32 implies that $\pi_X \circ f$ and $\pi_Y \circ f$ are both continuous, because π_X and π_Y are both continuous.

For the "if" part of (b), assume that $\pi_X \circ f$ and $\pi_Y \circ f$ are both continuous, and let W be an open subset of $X \times Y$. We must show that $f^{-1}(W)$ is an open subset of Z. By Definition 7.19 W can be written as:

$$W = \bigcup_{\lambda \in \Lambda} U_\lambda \times V_\lambda, \quad \text{with } U_\lambda \subset X,\ V_\lambda \subset Y \text{ open } \forall \lambda \in \Lambda.$$

By Proposition A.8, we have that

$$f^{-1}\left(\bigcup_{\lambda \in \Lambda} U_\lambda \times V_\lambda \right) = \bigcup_{\lambda \in \Lambda} f^{-1}(U_\lambda \times V_\lambda).$$

For each $\lambda \in \Lambda$,

$$\begin{aligned} f^{-1}(U_\lambda \times V_\lambda) &= \{z \in Z \ : \ f(z) \in U_\lambda \times V_\lambda\} \\ &= \{z \in Z \ : \ \pi_X \circ f(z) \in U_\lambda \text{ and } \pi_Y \circ f(z) \in V_\lambda\} \\ &= (\pi_X \circ f)^{-1}(U_\lambda) \cap (\pi_Y \circ f)^{-1}(V_\lambda) \end{aligned}$$

Since we have assumed that $\pi_X \circ f$ and $\pi_Y \circ f$ are continuous, this set is the intersection of two open sets and hence open. Therefore $f^{-1}(W)$ is a union of open sets, hence open. □

[1] Clearly $U \subset X$ and $Y \subset Y$ are open. Thus for all $(x, y) \in U \times Y$ we have $(x, y) \in U \times Y \subset U \times Y$

7.4.2 Gluing continuous functions together

Lemma 7.35 (glue lemma). *Let X and Y be two topological spaces and let A and B be subsets of X such that $X = A \cup B$. Let $g : A \to Y$ and $h : B \to Y$ be two continuous functions such that $g(x) = h(x)$ for all $x \in A \cap B$. Define $f : X \to Y$ by*

$$f(x) = \begin{cases} g(x) & \text{if } x \in A \\ h(x) & \text{if } x \in B. \end{cases}$$

Then

(a) if A and B are both open subsets of X then f is continuous; and

(b) if A and B are both closed subsets of X then f is continuous.

Proof.

(a) Suppose that A and B are open. Let U be an open subset of Y. Then

$$f^{-1}(U) = g^{-1}(U) \cup h^{-1}(U).$$

Since g and h are both continuous the subsets $g^{-1}(U) \subset A$ and $h^{-1}(U) \subset B$ are open in the subspace topology. The subsets $A \subset X$ and $B \subset X$ are open, so by Lemma 7.24, $g^{-1}(U)$ and $h^{-1}(U)$ are open subsets of X. Therefore $f^{-1}(U)$ is a union of open sets, hence open.

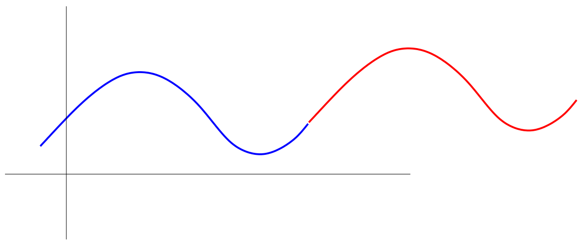
(b) Suppose now that A and B are closed. Let C be a closed subset of Y. By Proposition 7.31, it suffices to show that $f^{-1}(C)$ is closed. We have that

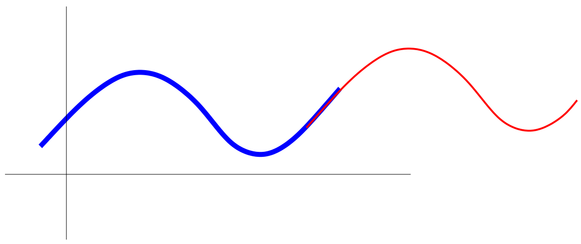
$$f^{-1}(C) = g^{-1}(C) \cup h^{-1}(C).$$

Since g and h are both continuous the subsets $g^{-1}(C) \subset A$ and $h^{-1}(C) \subset B$ are closed. Then by Remark 7.25, since $A \subset X$, $B \subset X$ are closed, we know that $g^{-1}(C) \subset X$, $h^{-1}(C) \subset X$ are closed. Since the union of two closed sets is closed we have $f^{-1}(C) \subset X$ is closed.

□

Examples:





7.5 Quotient topologies

This Section will allow us to make rigorous sense of gluing and/or deforming topological spaces to make new ones via equivalence relations.

Lemma 7.36. *Let (X, τ) be a topological space, and Y be any* set, *and let $f : X \to Y$ be any function. Then*

$$\tau_f := \{U \subset Y \ : \ f^{-1}(U) \in \tau\}$$

is a topology on Y.

Remark 7.37. τ_f is the largest/finest topology on Y which makes f continuous.

Proof. To prove this lemma we just need to check the axioms.
(T0/1) holds because $f^{-1}(Y) = X \in \tau_f$ and $f^{-1}(\emptyset) = \emptyset \in \tau_f$.
(T2) holds because if $U, V \in \tau_f$ then

$$f^{-1}(U \cap V) = f^{-1}(U) \cap f^{-1}(V) \in \tau.$$

(T3) holds because if $\{U_\lambda\}_{\lambda \in \Lambda}$ is a collection of open sets in τ_f then

$$f^{-1}(\bigcup_{\lambda \in \Lambda} U_\lambda) = \bigcup_{\lambda \in \Lambda} f^{-1}(U_\lambda) \in \tau.$$

□

This is a rather abstract-looking construction. But in fact we can use it to define 'gluings of points' in a topological space — to make a new space — that are relatively intuitive, as we shall see now.

7.5.1 Equivalence relations and topologies

Recall that an equivalence relation $\sim$ on a set X is a relation that partitions X into equivalence classes, denoted $[x]$. Such a relation satisfies

- $x \sim x$ (reflexivity)
- $x \sim y \implies y \sim x$ (symmetry)
- $x \sim y, y \sim z \implies x \sim z$ (transitivity)

The equivalence class

$$[x] = \{y \in X : x \sim y\}.$$

The **quotient set** of the relation is simply the collection of all equivalence classes

$$X/_\sim = \{[x] : x \in X\}.$$

There is a natural map

$$p : X \to X/_\sim,$$

given by $p : x \mapsto [x]$ called the **quotient map**. This allows us to equip $X/_\sim$ with a topology:

Definition 7.38. Let (X, τ) be a topological space, let $\sim$ be an equivalence relation on X, and let $p : X \to X/_\sim$ be the map $x \to [x]$. The **quotient topology** on $X/_\sim$ is

$$\tau_\sim = \tau_p = \{U \subset X/_\sim \ : \ p^{-1}(U) \in \tau\}.$$

The quotient set $X/_\sim$ equipped with the topology $\tau_\sim$ is called the **quotient space**.

Remark 7.39. The quotient map p is continuous by construction.

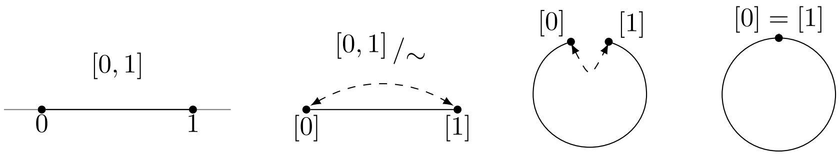
This forms a topology because of Lemma 7.36.

Example 7.40. Let $X = [0,1] \subset \mathbb{R}$ and $\sim$ be defined by $0 \sim 1$: it goes without saying that $x \sim x$ for all $x \in [0,1]$. The equivalence classes are thus

$$[x] = \{x\} \text{ when } x \in (0,1) \quad \text{and} \quad [0] = [1] = \{0,1\}.$$

Defining $S^1 = \{(x_1,x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ with the subspace topology, we will see later that there is a homeomorphism between the topological spaces $[0,1]/_\sim$ and S^1.

The following picture indicates this:



This **picture** is really quite magical, and deserves a book to itself. For now let us simply note again the following. 'Pictures' of this kind are subsets of the page. And the page is a certain copy of $\mathbb{R}^2$ (with or without coordinatisation). So (without the coordinatisation) a picture of $[0,1]$ is a realisation of the interval in $\mathbb{R}^2$.

Very many different realisations are of course possible. When we do put axes in to coordinatise $\mathbb{R}^2$ we often think of them as copies of $\mathbb{R}$ sitting in $\mathbb{R}^2$. They are copies embedded in $\mathbb{R}^2$ in a very 'straight' way: no curving, and so speeding up and slowing down as we pass along the line. Our $[0,1]$ is a subset of $\mathbb{R}$, so this gives us a nice straight picture... ...the first picture above.

However we could also draw it embedded with a bit of 'curvature' (so far in the very general sense of simply not being straight. That is what the third picture illustrates. (The second picture is an anticipation of the quotient we are constructing in the example — it notes that we are going to identify the ends of the interval... but of course this is not something we can easily to while the interval is embedded rigidly, straightly on the page.)

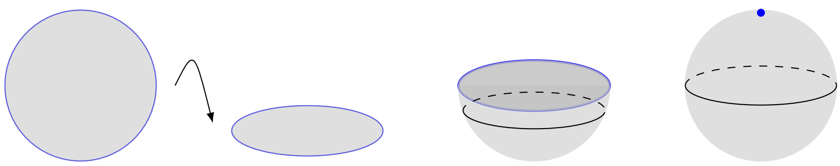
Now that we have got some curvature, we can ('smoothly') bring the ends closer and closer together. If they touch then this is no longer a very good picture of $[0,1]$, since the touching identifies 0 and 1. (Worse sins happen in the world of picture making — where the picture may not be a faithful likeness of the subject! ...That is a story for another day.) But for our quotient this is exactly what we want. So the final picture (far above now) is a picture of the quotient space.

7.5.2 Examples with surfaces

Example 7.41. Let $X = D^2 \subset \mathbb{R}^2$ where $D^2 := \{x = (x_1, x_2) : |x|^2 = x_1^2 + x_2^2 \leq 1\}$ and $\sim$ be defined by $x \sim y$ if and only if $x = y$ or $|x| = |y| = 1$. The equivalence classes are thus

$$[x] = \{x\} \text{ when } |x| < 1 \quad \text{and} \quad [x] = \{y \in S^1\} \text{ when } |x| = 1.$$

Defining $S^2 = \{z = (z_1, z_2, z_3) \in \mathbb{R}^3 : |z|^2 = 1\}$, we will see later that there is no difference between the topological spaces ${D^2}/{\sim}$ and S^2:



The examples above fall into an important category of quotient spaces In general, let X be a topological space and $A \subset X$. Define an equivalence relation on X via

$$x \sim y \quad \iff x = y, \quad \text{or} \quad x, y \in A.$$

You should check this is an equivalence relation. Then the equivalence classes are A and the sets $\{x\}$ such that $x \in X \setminus A$.

Notation 7.42. In the case of the above equivalence relation, the resulting quotient space is often denoted by ${X}/{A}$.

Example 7.43. Let $X = [0,1] \times [0,1] \subset \mathbb{R}^2$. Let $\sim$ be the equivalence relation on X such that

$$(0, y) \sim (1, y) \quad \forall y \in [0, 1].$$

The equivalence classes for this equivalence relation consist of 1-point sets,

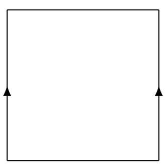
$$\{(x, y)\} = [(x, y)] \quad \text{when } x \in (0, 1) \text{ and } y \in [0, 1],$$

and 2-point sets

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

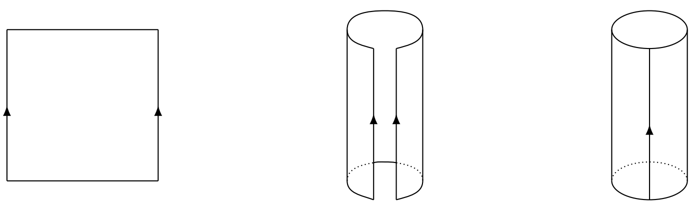
We say this equivalence relation 'identifies' the left- and right-hand edges of the square. The quotient ${X}/{\sim}$ is a topological space known as the **cylinder**.

We represent the cylinder construction with the following picture:



The arrows indicate the direction in which the edges have been 'identified', i.e. matched up, point by point.

You should be able to see why this space is called a cylinder:



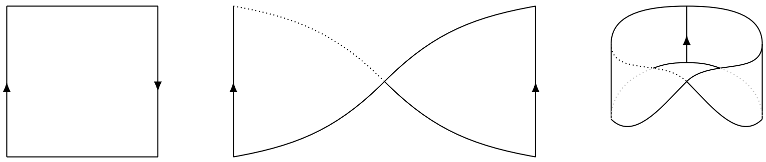
Here we've gone a step further and *visualised* the identification by embedding the square in $\mathbb{R}^3$ and then curving it up until the edges to be identified are literally matched together. This is just a visualisation though. The actual cylinder in the final drawing is a representation of a subspace of $\mathbb{R}^3$ which is homeomorphic to (but not identically equal to) our construction.

Example 7.44. Another example of an equivalence relation on $X = [0,1] \times [0,1]$ is

$$(0, y) \sim (1, 1-y) \quad \forall y \in [0, 1].$$

The quotient of $X/\sim$ in this case is known as the **Möbius strip**.

The picture in this case looks like this (you should see a band with a twist in it):



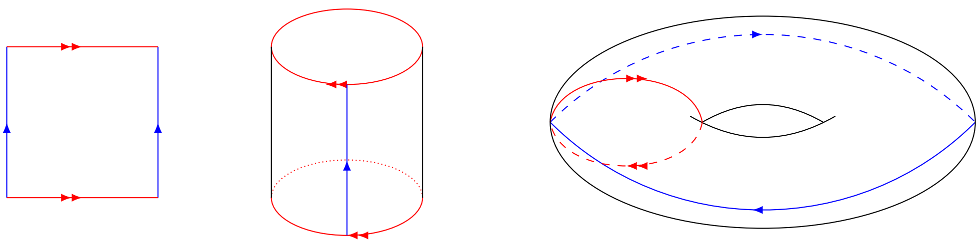
Once again, the arrows indicate the directions in which the edges have been identified.

Example 7.45. Consider the equivalence relation (again on $X = [0,1] \times [0,1]$)

$$(0, y) \sim (1, y) \quad \forall y \in [0, 1],$$
$$(x, 0) \sim (x, 1) \quad \forall x \in [0, 1],$$

The first identification is essentially the same as we did in making the cylinder above. So in effect the remaining edges of the old square actually become 'loops'. The second identification is also quite similar to the edge-to-edge join, but now we are identifying subsets that have become 'ends of a cylinder'. Because of this, in a sense the shape 'closes up' entirely. This gives a topological space called the **torus**.

In the following picture, we again imagine the square embedded in $\mathbb{R}^3$ and then gradually 'wrap it up' to make the identifications. The single arrows indicate the identification of the left and right edges (blue), and the double arrows indicate the identification of the upper and lower edges (red). Once one makes this identification one ends up with something that looks like the surface of a ring donut.



You may also like to go the other way. Start with the surface of a ring donut — just the surface (the wrapping of a lovely donut gift perhaps) and then cut around a line that passes through the hole in the donut; and also cut 'equatorially' — you end up with a (possible crumpled) square when you open this out! All the 'hole-y-ness' has been cut away.

7.5.3 Funkier examples: maths-worlds

So far the examples have some kind of physical underpinning — gluing together certain points in a shape to make a new shape. But of course one of the nice things about maths is that, having formalised operations like this, we can then 'game' them, sometimes far beyond the underpinning ideas...

Example 7.46. Consider the equivalence relation (again on $X = [0,1] \times [0,1]$)

$$(0,y) \sim (1,y) \quad \forall y \in [0,1],$$
$$(x,0) \sim (1-x,1) \quad \forall x \in [0,1],$$

Notice that this looks a bit like a mix of the identifications that we have done already, but we have not done this exact combination before.

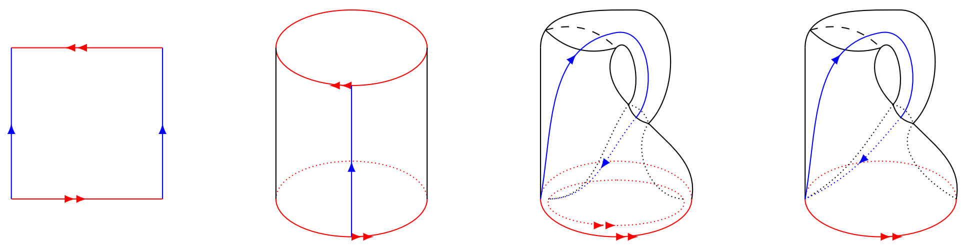
Our abstract prescription for building topological spaces works (as we have shown) for any such construction. So this latest one is certainly a space. But with our recent constructions we have also been visualising them — as 'surfaces'...

Usually the word surface applies when we have a solid 3d thing and we want to think about the 'boundary' between the thing and the rest of the world. That boundary is the surface. Hence the surface of the ocean, or a donut or a ball or a pretzel or a tiger. Our picture for some of the recent constructions has essentially related them to surfaces in this nice sense — making the 'square of wrapping paper' we started with wrap the awkward shape.

But that conceptual relationship was, to some extent, just lucky for us. Now things are getting much wilder...

This latest construction gives a topological space — often called the **Klein bottle** — that is not really the surface of any nice solid 3d thing. [2]

Unlike the previous examples, this one can't be realised/imagined (in three dimensions) without self-intersections:

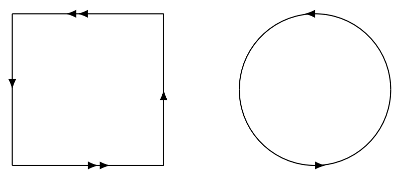


The 'picture' we have ended up with is not really a proper 3d thing at all any more. You can take it or leave it. The picture is not the important thing. The important thing is the construction. And the real questions are ones like: do you think this is homeomorphic to the torus or not?!

Example 7.47. The equivalence relation (again on $X = [0,1] \times [0,1]$)

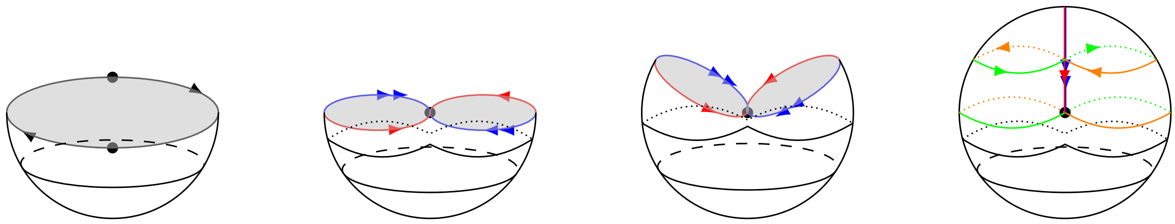
$$(0,y) \sim (1,1-y) \quad \forall y \in [0,1],$$
$$(x,0) \sim (1-x,1) \quad \forall x \in [0,1],$$

gives a topological space called the **projective plane**, denoted $\mathbb{RP}^2$. By smoothing out the edges of the square, you can imagine that the pictures below are both fair representations of this topological space:



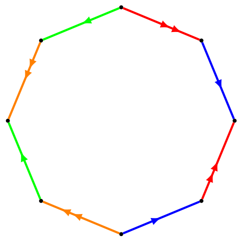
[2]Would this surface be any use as a bottle? Notice that it doesn't have a well-defined inside or outside: see a youtube video of the construction here: https://www.youtube.com/watch?v=yaeyNjUPVqs

Here, the right hand picture is $D^2/\sim$ where $x \sim y$ if $|x| = 1$ and $x = -y$. In other words we identify opposite points on the boundary of D^2. Once again the projective plane cannot be realised/imagined in three-dimensions without self intersections. Here's one way to do this, called the cross-cap:[3]

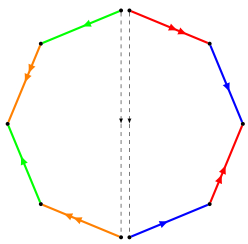


You should be able to convince yourself now that in fact $\mathbb{RP}^2$ can be equivalently obtained/defined via $S^2/\sim$ where $\sim$ is defined via $x \sim -x$.

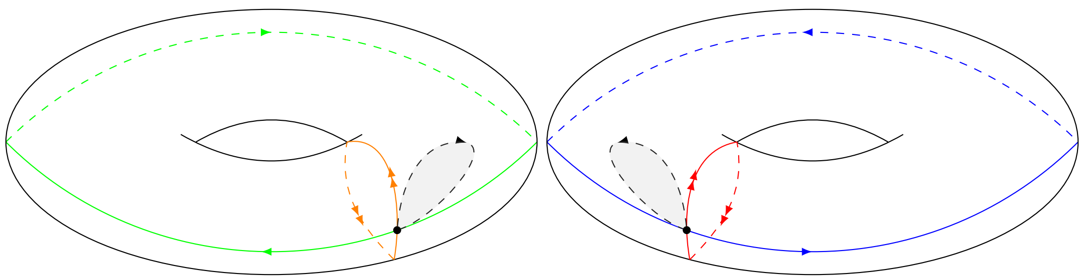
Example 7.48. We won't write the next one down precisely, but consider an octagon with its edges identified as suggested in the picture below



To see what we end up with it is best to first split the picture down the middle and cut it into two:

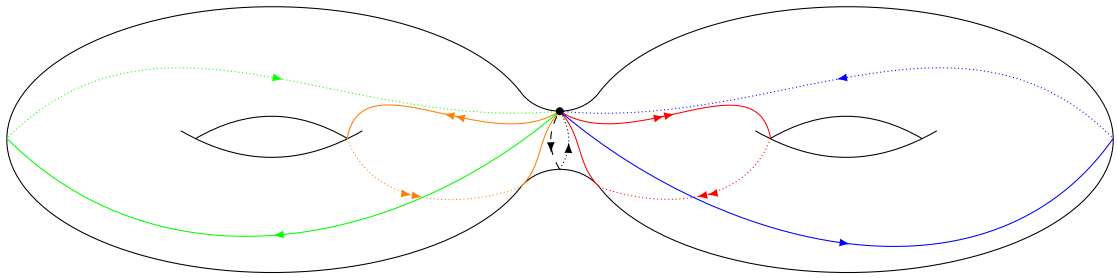


We end up with two pieces that each look like a torus, but with an extra curve and a piece missing. Try to imagine why we end up with the following (the grey areas indicate parts that are missing):



Now we must stitch the surface together again along the black dashed line - notice that the vertices of the hexagon have now all been identified to the same point. We end up with:

[3]find a youtube video of this construction here: https://www.youtube.com/watch?v=W-sKLN0VBkk



Recall that the product, subspace and metric constructions of topologies respect the separated property. The quotient topology does not, as the following counterexample shows:

Counterexample 7.49. Let $X = \mathbb{R}$ with its Euclidean topology and let $\sim$ be the equivalence relation

$$x \sim y \;\Leftrightarrow\; \exists\, \lambda \in \mathbb{R} \setminus \{0\} \text{ such that } x = \lambda y,$$

You may check for yourself that this is symmetric, reflexive, and transitive. Then the quotient space has just two elements, namely

$$X/_{\sim} = \{\{0\}, \mathbb{R} \setminus \{0\}\}.$$

This is true since any non-zero $x, y \in \mathbb{R}$ are related via $\lambda = x/y$. In order to calculate its topology, let $p : X \to X/_{\sim}$ denote the projection. The subset $\{\{0\}\} \subset X/_{\sim}$ is not open since $p^{-1}\{\{0\}\} = \{0\}$ and $\{0\}$ is not an open subset of $\mathbb{R}$ in the standard topology. All other subsets are open, so the topology is given by

$$\tau = \left\{\emptyset, \{\mathbb{R} \setminus \{0\}\}, X/_{\sim}\right\}.$$

Then $X/_{\sim}$ is not separated: for if $x = \{\{0\}\}$ and $y = \{\mathbb{R} \setminus \{0\}\}$ then the only open subset of $X/_{\sim}$ which contains x is $U = X/_{\sim}$. But then any open set V containing $\mathbb{R} \setminus \{0\}$ has to intersect U, because $\mathbb{R} \setminus \{0\} \in U$.

Despite this, the five quotient spaces defined above (cylinder, Möbius strip, torus, Klein bottle, projective plane) *are* separated spaces.

7.5.4 Continuous maps

Thinking about continuous maps between quotient spaces can sometimes be tricky. The next result helps us to do this.

Proposition 7.50. *Let X and Y be two topological spaces, let $\sim_X$ and $\sim_Y$ be equivalence relations on X and Y, and let $f : X \to Y$ be a continuous function. Suppose that f has the property that $\forall x, x' \in X$,*

$$x \sim_X x' \quad \Rightarrow \quad f(x) \sim_Y f(x').$$

Then

$$\widetilde{f} : X/_{\sim_X} \to Y/_{\sim_Y}, \quad \widetilde{f} : [x] \to [f(x)]$$

is a well-defined continuous function with respect to the quotient topologies.

Proof. First we check that $\widetilde{f}$ is well-defined, i.e. that if $[x] = [x']$ then $\widetilde{f}([x]) = \widetilde{f}([x'])$. Note that $[x] = [x']$ if and only if $x \sim_X x'$. Thus by assumption $[x] = [x']$ implies $f(x) \sim_Y f(x')$, which implies that $[f(x)] = [f(x')]$, or in other words, $\widetilde{f}([x]) = \widetilde{f}([x'])$.

Now we check that $\widetilde{f}$ is continuous. Let $p_X : X \to X/_{\sim_X}$ and $p_Y : Y \to Y/_{\sim_Y}$ be the two projections.

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle p_X}\downarrow & & \downarrow{\scriptstyle p_Y} \\ X/{\sim_X} & \xrightarrow{\widetilde{f}} & Y/{\sim_Y} \end{array}$$

Observe that

$$\widetilde{f} \circ p_X = p_Y \circ f$$

(because $p_Y(f(x)) = [f(x)] = \widetilde{f}([x]) = \widetilde{f}(p_X(x)))$.

Let $U \subset Y/{\sim}$ be open. We must show that $\widetilde{f}^{-1}(U)$ is open, which by definition means that $p_X^{-1}(\widetilde{f}^{-1}(U)) \subset X$ is open. By the identity above, $p_X^{-1}(\widetilde{f}^{-1}(U)) = (\widetilde{f} \circ p_X)^{-1}(U) = (p_Y \circ f)^{-1}(U) = f^{-1}(p_Y^{-1}(U))$. Now $p_Y^{-1}(U)$ is open because U is open, and $f^{-1}(p_Y^{-1}(U))$ is open because f is continuous. □

The proof of the above proposition also allows us to understand continuous functions from $X/{\sim}$ to Y via continuous functions $f : X \to Y$.

Corollary 7.51. *Let X and Y be two topological spaces, let $\sim$ be an equivalence relation on X, and let $f : X \to Y$ be a continuous function. Suppose that f has the property that $\forall x, x' \in X$,*

$$x \sim x' \quad \Rightarrow \quad f(x) = f(x').$$

Then

$$\widetilde{f} : X/{\sim} \to Y, \quad \widetilde{f} : [x] \to f(x)$$

is a well-defined continuous function with respect to the quotient topology on X.

Proof. This time we have $\widetilde{f} \circ p = f$ and so exactly the same reasoning as above gives the result. □

Remark 7.52. If $f : X \to Y$ is continuous and $\sim$ is an equivalence relation on Y then $p : Y \to Y/{\sim}$ is continuous so $\widetilde{f} = p \circ f : X \to Y/{\sim}$ is continuous by Proposition 7.32 - i.e. $\widetilde{f}(x) = [f(x)]$, is always a continuous function when f is.

Chapter 8

Sequences and limit points

A point of terminology: a **neighbourhood of a point** x is an open set U containing x.

Definition 8.1. Let X be a topological space and let A be a subset of X. The **closure** $\overline{A}$ of A is the intersection of all closed sets containing A:

$$\overline{A} = \bigcap_{\lambda \in \Lambda} C_\lambda,$$

where $\{C_\lambda\}_{\lambda \in \Lambda}$ is the collection of all closed sets containing A.

Subset A is called **dense** if $\overline{A} = X$.

Note that $\overline{A}$ is closed, by Theorem 7.14. Note also that $A \subset \overline{A}$. In fact, $\overline{A}$ is the smallest closed set containing A.

Exercise 8.1. Prove that A is closed if and only if $A = \overline{A}$.

8.1 Limit points

Definition 8.2. Let X be a topological space and let A be a subset of X. A point $x \in X$ is called a **limit point** of A if for every neighbourhood U of x, $U \cap A \neq \emptyset$.

Proposition 8.3. *Let A be any subset of a topological space X. Then*

$$\overline{A} = \{\textit{limit points of } A\}.$$

Proof. We prove each set is contained in the other.

Let $x \in \overline{A}$. Then for any closed set $C \supset A$, $x \in C$. Suppose, for contradiction, that x is not a limit point of A. Then there exists a neighbourhood U of x, such that $U \cap A = \emptyset$. Then $K = X \setminus U$ is closed, and $A \subset K$. But $x \notin K$, contradicting $x \in \overline{A}$. Hence x is a limit point of A.

Conversely, let x be a limit point of A and assume for contradiction that $x \notin \overline{A}$. Then there is a closed set C such that $C \supset A$ and $x \notin C$. Therefore $U = X \setminus C$ is an open set containing x, and $U \cap A = \emptyset$, contradicting the fact that x is a limit point of A. Hence $x \in \overline{A}$. □

Proposition 8.4. *Let X be a separated space and let $x \in X$. Then $\overline{\{x\}} = \{x\}$. In particular, $\{x\}$ is closed.*

Proof. Note that $x \in \overline{\{x\}}$ because $\{x\} \subset \overline{\{x\}}$. Let $y \in X$ be any point with $y \neq x$. Then there exist neighbourhoods U of x and V of y such that $U \cap V = \emptyset$. Then $V \cap \{x\} = \emptyset$, so y is not a limit point of $\{x\}$. So by Proposition 8.3, $y \notin \overline{\{x\}}$. □

Now we turn our attention to sequences.

Definition 8.5. Let X be a topological space, let $x \in X$, and let $\{x_n\}_{n\in\mathbb{Z}^+}$ be a sequence of points in X. The sequence x_n is said to **converge to** x, and x is called a **limit** of the sequence x_n, if for every neighbourhood U of x there exists $N \in \mathbb{Z}^+$ such that $x_n \in U$ for all $n \geq N$. We write $x_n \to x$.

Example 8.6. (I) In general, limits may not be unique!

In the indiscrete topology, all sequences converge to everything. Let X be set with the indiscrete topology, let x_n be any sequence in X and let x be any point in X. The only neighbourhood of x is X, and $x_n \in X$ whenever $n \geq 1$. So $x_n \to x$.

(II) Consider the space $\mathbb{R}$ with the usual topology, and the sequence $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, ...$ This sequence converges to 0. (Exercise: Verify.)

(III) The sequence $1, -1, 1, -1, 1, -1, ...$ in $\mathbb{R}$ does not converge. (Exercise: Verify.)

There is an equivalent way to check convergence in metric spaces:

Theorem 8.7. *Let (X, d) be a metric space and let x_n be a sequence in X. Then x_n converges to a point $x \in X$ if and only if for every $\varepsilon > 0$ there exists an $N \in \mathbb{Z}^+$ such that $x_n \in B_\varepsilon(x)$ whenever $n \geq N$.*

Exercise 8.2. Prove that, in a metric space (X, d), $x_n \to x$ if and only if $d(x_n, x) \to 0$. This follows by writing this out in the usual ε-N language.

This theorem implies that in $\mathbb{R}$ with the standard (metric-induced) topology, our definition of convergence agrees with one that you have seen before.

Proof. The "only if" part is straightforward, because any open ball $B_\varepsilon(x)$ is a neighbourhood of x.

For the "if" part, suppose that x_n is a sequence with the property that $\forall\varepsilon > 0\ \exists N \in \mathbb{Z}^+$ such that $x_n \in B_\varepsilon(x)$ whenever $n \geq N$. Let U be a neighbourhood of x. Since U is open and $x \in U$ there exists an $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U$. Let N be such that $x_n \in B_\varepsilon(x)$ whenever $n \geq N$. Then $x_n \in U$ whenever $n \geq N$. So $x_n \to x$. □

Now we return to the question of how many limits a sequence can have. Everything works nicely in separated spaces:

Theorem 8.8. *In separated spaces, limits of convergent sequences are unique.*

Proof. Let X be a separated space and suppose for contradiction that x_n is a sequence in X with two distinct limits x and y. Since X is separated there exist disjoint neighbourhoods U and V of x and y. Since $x_n \to x$ and $x_n \to y$ there exist $M, N \in \mathbb{Z}^+$ such that $x_n \in U$ whenever $n \geq M$ and $x_n \in V$ whenever $n \geq N$. If $n \geq \max\{M, N\}$ then $x_n \in U \cap V$, contradicting the fact that U and V are disjoint. So a sequence x_n can have only one limit. □

Is every limit point of a set the limit of some sequence within that set? In general no, but the answer is yes in metric spaces!

Lemma 8.9. *In a metric space (X, d), let $A \subset X$, and x be a limit point of A. Then there exists a sequence $\{x_n\}_{n \in \mathbb{Z}^+} \subset A$ such that $x_n \to x$.*

Proof. Consider the ball $B_{\frac{1}{n}}(x)$, for $n \in \mathbb{Z}^+$. This is certainly a neighbourhood of x for all $n \in \mathbb{Z}^+$. Hence, as x is a limit point of A, we must have

$$B_{\frac{1}{n}}(x) \cap A \neq \emptyset.$$

Thus, there are points $x_n \in B_{\frac{1}{n}}(x) \cap A$ for each $n \in \mathbb{Z}^+$ which define a sequence in A. Let $\varepsilon > 0$, and choose $N \in \mathbb{Z}^+$ such that $N > \frac{1}{\varepsilon}$. Then for all $n \geq N$,

$$d(x_n, x) < \frac{1}{n} \leq \frac{1}{N} < \varepsilon,$$

so $x_n \in B_\varepsilon(x)$ for all $n \geq N$, which means $x_n \to x$ by Theorem 8.7. □

Lemma 8.10. *Let X be a topological space and $C \subset X$. If C is closed, then for all convergent sequences $\{x_n\} \subset C$ such that $x_n \to x \in X$, we have $x \in C$.*

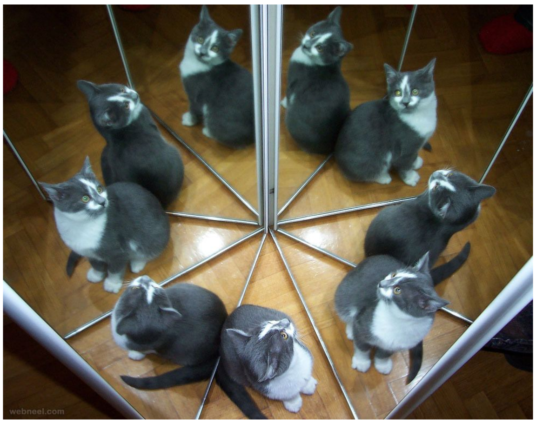
If (X, d) is a metric space and $C \subset X$ then the converse is true. If $C \subset X$ satisfies: for all convergent sequences $\{x_n\} \subset C$ such that $x_n \to x \in X$, we have $x \in C$. Then C is closed.

Proof. For the first part, let $\{x_n\} \subset C$ such that $x_n \to x \in X$, then x is a limit point of C since for all neighbourhoods U of x there exists some N s.t. $n \geq N$ implies $x_n \in U$. i.e. $U \cap C \neq \emptyset$ for all such U.

Assuming (X, d) is a metric space let $x \in \overline{C}$. Then by Lemma 8.9 there exists a sequence $\{x_n\} \subset C$ such that $x_n \to x$. By assumption we have $x \in C$ and thus $\overline{C} \subset C$, in which case $C = \overline{C}$ is closed. □

Chapter 9

Homeomorphisms



In this section we compare topological spaces. A key concept is that of “homeomorphism”: loosely speaking, two spaces which are homeomorphic are topologically the same, up to a relabelling of the points.

Definition 9.1. Let X and Y be two topological spaces and let $f : X \to Y$. Then f is called a **homeomorphism** if it is bijective and both f and f^{-1} are continuous. In this case X is **homeomorphic** to Y, denoted by $X \cong Y$.

Remark 9.2. Notice that a homeomorphism induces a bijection between the topologies on X and Y: for all open $V \subset Y$ there exists a unique open $U \subset X$ such that $f(U) = V$. In this sense they can essentially be considered to be the ‘same’ topological space.

Example 9.3. $\mathbb{R}$ is homeomorphic to $(0, \infty)$. The function $f : \mathbb{R} \to (0, \infty)$,

$$f : x \mapsto e^x$$

is a bijection; both f and its inverse $f^{-1} : (0, \infty) \to \mathbb{R}$, $f^{-1} : y \mapsto \ln y$ are continuous.

Example 9.4. Let $X = \{1, 2\}$, $\sigma = \{\emptyset, \{1\}, \{1, 2\}\}$, $Y = \{3, 4\}$, and $\tau = \{\emptyset, \{3\}, \{3, 4\}\}$. Then $f : X \to Y$, $f(1) = 3$, $f(2) = 4$ is a homeomorphism but $g : X \to Y$, $g(1) = 4$, $g(2) = 3$ is not ($g^{-1}(\{3\}) = \{2\}$ is not open).

9.1 A little algebraic structure

(**9.1.1**) Recall the category hom consisting of sets and functions between them. Recall that hom^i is the subcategory of hom of bijections.

Recall that Top is like a subcategory of hom, in that arrow are functions between sets (but not all functions). We can 'restrict' further to Top^i, formally, but the above (and below) example shows that it is better to restrict to Top^h.

(**9.1.2**) Lemma. If $f, g \in \mathsf{Top}^h$ are composable, i.e. $f \circ g \in$ hom, then $f \circ g \in \mathsf{Top}^h$.
Proof. Exercise.

(**9.1.3**) Thus Top^h is a groupoid. And each $\mathsf{Top}^h(X, X)$ is a group.

Proposition 9.5. *Homeomorphism is an equivalence relation.*

Proof. First check symmetry: suppose that X is homeomorphic to Y. Then there exists a homeomorphism $f : X \to Y$. The inverse $f^{-1} : Y \to X$ is also a homeomorphism, so Y is homeomorphic to X.

Next check reflexivity: the identity map $\mathrm{id}_X : X \to X$, $\mathrm{id}_X : x \mapsto x$ is a homeomorphism, so X is homemorphic to X.

Finally, check transitivity: if $f : X \to Y$ is a homemorphism and $g : Y \to Z$ is a homemorphism then $g \circ f : X \to Z$ is also a homeomorphism, hence X is homeomorphic to Z. □

9.2 And more examples

First a non-example.

Example 9.6. Let $X = [0, 1)$ and $Y = S^1 = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$, and equip X and Y with their subspace topologies inherited from $\mathbb{R}$ and $\mathbb{R}^2$. Let $f : X \to Y$ be the map

$$f : t \to (\cos(2\pi t), \sin(2\pi t)).$$

Then f is a continuous bijection, but f is *not* a homeomorphism. This is because $f^{-1} : Y \to X$ is not continuous.

To prove that f^{-1} is not continuous, it is enough to find one open subset $U \subset X$ such that $(f^{-1})^{-1}(U) = f(U)$ is not open in Y. Consider $U = [0, 1/2)$. Then U is an open subset of X (because $U = (-1/2, 1/2) \cap X$ — noting (7.2)), but $f(U)$ is not open. The reason for this is that

$$f(0) = \underline{(1, 0)} \in f(U)$$

— underline just to distinguish a vector from an interval!! — but any open set in Y containing $f(0)$ is not a subset of $f(U)$ (exercise: check the details; hint: draw some pictures!).

Note that just because some $f : X \to Y$ is not a homeomorphism does not mean X and Y are not homeomorphic (consider the $\{1, 2\}$ example above). In the present example it is perhaps intuitive that the spaces are not homeomorphic (i.e. that *no* homeomorphism exists). Can you see a way to prove it? We will come back to this shortly.

Example 9.7. You will prove in the exercises that for $a, b \in \mathbb{R}$, $a < b$ then the open interval $(a, b) \cong \mathbb{R}$.

Reminder: Continuity of multivariable functions Consider the spaces $\mathbb{R}^n$, $\mathbb{R}^m$ equipped with their standard topologies. By Proposition 7.34

$$f(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), f_2(x_1, \ldots x_n), \ldots f_m(x_1, \ldots, x_n))$$

is continuous if and only if $f_i(x_1, \ldots, x_n)$ is continuous for all $i = 1, \ldots, m$.

So we really only need to be concerned with $f_i(x_1, \ldots, x_n)$ in order to check continuity of f.

Example 9.8. Let $B_r^n(x_0) := \{x \in \mathbb{R}^n : |x - x_0| < r\}$, i.e. the open ball or radius r centred at x_0 in $\mathbb{R}^n$. Then for all $x_0, y_0 \in \mathbb{R}^n$ and $r, R > 0$ we have $B_r^n(x_0) \cong B_R^n(y_0)$ since $f : B_r^n(x_0) \to B_R^n(y_0)$ defined by

$$f(x) = \frac{R}{r}(x - x_0) + y_0$$

is a continuous bijection with inverse $f^{-1}(y) = \frac{r}{R}(y - y_0) + x_0$, which is also continuous.

Example 9.9. Define $f : B_1^n(0) \to \mathbb{R}^n$ by

$$f(x) = \frac{x}{1 - |x|}.$$

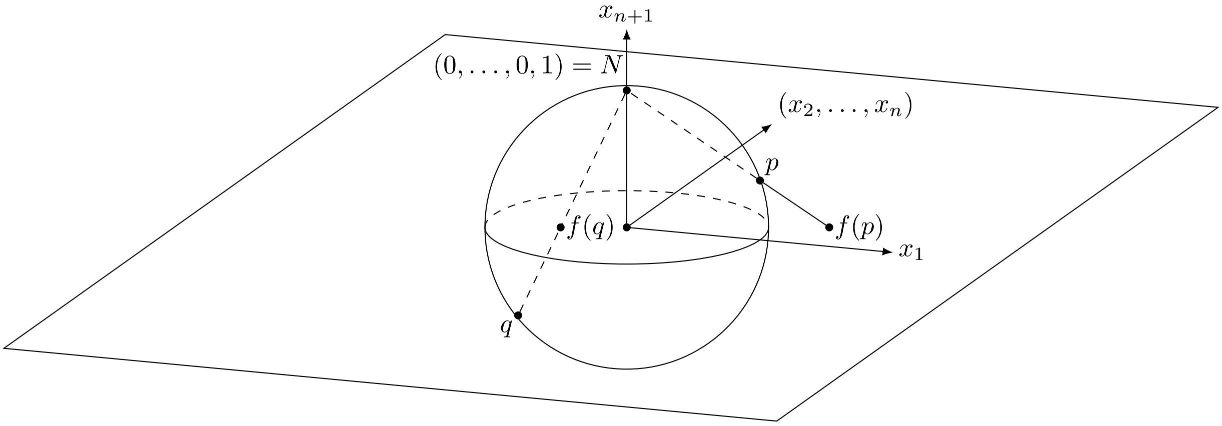
You can check that f is continuous, and directly compute its inverse to be $f^{-1} : \mathbb{R}^n \to B_1^n(0)$

$$f^{-1}(y) = \frac{y}{1 + |y|},$$

which is also continuous. Thus $B_1^n(0) \cong \mathbb{R}^n$.

Example 9.10. Let $S^n := \{x \in \mathbb{R}^{n+1} : |x|^2 = 1\}$.. Defining the North Pole $N = (0, \dots, 0, 1)$, we now define the *stereographic projection* $f : S^n \backslash \{N\} \to \mathbb{R}^n$ in the following way. First off, we make the identification $\mathbb{R}^n \cong \{x \in \mathbb{R}^{n+1} : x^{n+1} = 0\}$.

Pick $p \in S^n \backslash \{N\}$ and draw the unique straight line in $\mathbb{R}^{n+1}$ joining p and N. Extending this line *ad infinitum*, it will strike $\mathbb{R}^n$ once, and only once at a point, which we define to be $f(p)$.



As an exercise, you should be able to check that

$$f(x_1, \dots, x_{n+1}) = \frac{1}{1 - x_{n+1}}(x_1, \dots, x_n),$$

which is continuous on $S^n \backslash \{N\}$; since $f : \mathbb{R}^{n+1} \backslash \{N\} \to \mathbb{R}^n$ is continuous and $S^n \backslash \{N\} \subset \mathbb{R}^{n+1} \backslash \{N\}$ has inherited the subspace topology (see Proposition 7.33). Furthermore, you can check that

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

is also continuous (also by Proposition 7.33), proving that f is a homeomorphism, i.e. $S^n \backslash \{N\} \cong \mathbb{R}^n$.

A similar construction can be carried out if we single out any point of S^n (i.e. not just for the North Pole N). So in fact we have $S^n \backslash \{p\} \cong \mathbb{R}^n$ for any $p \in S^n$. The easiest way to do this is to let $R : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be a rotation or reflection (i.e. an orthogonal matrix[1])

[1] R satisfies $R^{-1} = R^T$

satisfying $R(p) = N$. Then $f \circ R : S^n \backslash \{p\} \to \mathbb{R}^n$ is our desired homeomorphism. In particular, if $S = (0, \ldots, 0, -1)$ is the South Pole, then we set $R_S(x_1, x_2, \ldots, x_{n+1}) = (x_1, \ldots, x_n, -x_{n+1})$ and $f \circ R_S : S^n \backslash \{S\} \to \mathbb{R}^n$ is a homeomorphism.

Example 9.11. Let $D^n = \overline{B_1^n(0)} \subset \mathbb{R}^n$, so that $D^n = \{y \in \mathbb{R}^n : |y| \leq 1\}$. Let

$$S_+^2 = \{x \in S^n : x_{n+1} \geq 0\}$$

be the upper hemisphere. Now we'll show that $D^n \cong S_+^n$.

Let $f : D^n \to \mathbb{R}^{n+1}$ be given by

$$f(y_1, y_2, \ldots, y_n) = (y_1, \ldots, y_n, \sqrt{1 - |y|^2}).$$

By now we have seen that an f of this form is continuous: we have the reminder above; and we recall that $\sqrt{1 - |y|^2}$ is continuous for the relevant y values. And furthermore $|f|^2 = 1$ for all $y \in D^n$ and $f_{n+1}(y_1, \ldots, y_n) \geq 0$. So that $f : D^n \to S_+^n$ is in fact continuous.

The inverse f^{-1} is given by

$$f^{-1}(x_1, x_2, \ldots, x_{n+1}) = (x_1, x_2, \ldots, x_n)$$

and again, is clearly continuous. Thus f is a homeomorphism.

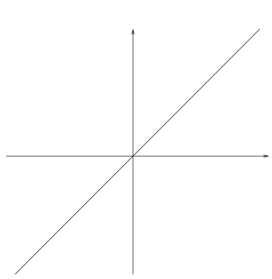
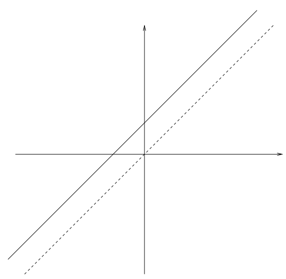
9.3 Self-homeomorphisms

Here are some examples from $\mathsf{Top}^h(\mathbb{R}, \mathbb{R})$.

For $a \in \mathbb{R}$,

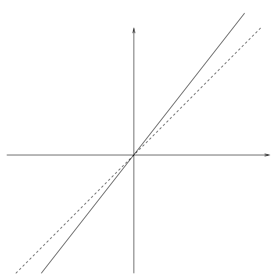
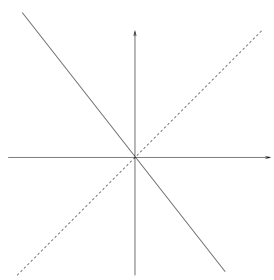
$$\begin{aligned} \sigma_a : \mathbb{R} &\to \mathbb{R} && (9.1) \\ x &\mapsto x + a && (9.2) \end{aligned}$$

Such functions can be graphed. Here are the cases $a = 0$ and $a = 1$:

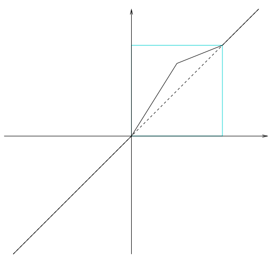
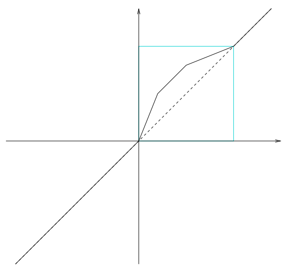



For $a \in \mathbb{R}^\times = \mathbb{R} \setminus \{0\}$

$$\begin{aligned} \mu_a : \mathbb{R} &\to \mathbb{R} && (9.3) \\ x &\mapsto ax && (9.4) \end{aligned}$$

NB these do not restrict to homeomorphisms (or even functions) on $[0,1]$ or $(0,1)$.
Here are some nice ones 'of finite support':

These ones do restrict to homeomorphisms of $[0,1]$.

(**9.3.1**) Exercise. Expressed in the graphing setting above, describe how to construct the inverse of $f \in \mathsf{Top}^h(\mathbb{R},\mathbb{R})$. (Indeed we can do this, in a sense, for $f \in \hom(\mathbb{R},\mathbb{R})$.)

(**9.3.2**) Exercise. Recall the $\tau_\circ$ topology from (3.2.36). Is there a path in $\mathsf{Top}(\mathbb{I},(\mathsf{Top}^h(\mathbb{R},\mathbb{R}),\tau_\circ))$ from the identity $id_\mathbb{R} = \sigma_0$ to μ_{-1}?

(**9.3.3**) Exercise. Why is it natural to organise elastic bands topologically? (This is a very open question! Just have a think.)

(**9.3.4**) Note that we can make a composition on $\mathsf{Top}(\mathbb{I},(\mathsf{Top}^h(\mathbb{R},\mathbb{R}),\tau_\circ))$ by $(f.g)_t = f_t \circ g_t$.

Chapter 10

Organisation and classification: Tigers at last?

10.1 Topological manifolds and the classification of curves and surfaces

In this section we will not provide proofs. We include it in order to see some beautiful classification results for one- and two-dimensional topological manifolds (as defined below).

Definition 10.1. Let X be any set. A **basis for a topology** on X is a collection β of subsets of X such that

(B1) For each $x \in X$, there is at least one $B \in \beta$ such that $x \in B$.

(B2) If $B_1, B_2 \in \beta$ and $x \in B_1 \cap B_2$ then there exists $B_3 \in \beta$ such that $x \in B_3$ and $B_3 \subset B_1 \cap B_2$. (The WIC condition.)

The sets $B \in \beta$ are called **basis elements**.

Definition 10.2. Let β be a basis for a topology on a set X. The **topology τ generated by β** is

$$\tau = \{U \subset X \ : \ \text{for each } x \in U \text{ there exists a } B \in \beta \text{ such that } x \in B \text{ and } B \subset U\}\,.$$

Equivalently

$$\tau = \{U \subset X : \exists \{B_\lambda\}_{\lambda \in \Lambda} \subset \beta \ s.t. \ U = \cup_\lambda B_\lambda\}.$$

One should check that this is a topology (see earlier). We have seen many examples of topologies generated by bases. For instance the metric topology on a metric space (X, d) is generated by the basis of open balls $\beta = \{B_r(x) : x \in X,\ r > 0\}$. Furthermore, the product topology on $X \times Y$ is generated by the basis $\{U \times V : U \subset X \text{ and } V \subset Y \text{ are open.}\}$.

Definition 10.3. Topological space (X, τ) is said to be **second countable** if it is generated by a basis with countably many elements.

For example, $\mathbb{R}^n$ (and any subspace of it) is second countable since all balls of rational radii centred at points with rational coordinates, is a basis for the standard topology (and thus a basis for any subspace). Similarly, *any* metric space (X, d) which has a countable subset $A \subset X$ such that $\overline{A} = X$, is second countable. Furthermore if X and Y are second countable, then $X \times Y$ is second countable.

If we let R denote the real numbers equipped with the discrete topology, then it is *not* second countable.

Definition 10.4. A topological space M is called a **topological manifold of dimension** n if

1. M is separated
2. M is second countable
3. For all $x \in M$ there exists a neighbourhood U_x and a homeomorphism $f : U_x \to B_1^n(0) \subset \mathbb{R}^n$ with $f(x) = 0$: this last condition can be stated that M is 'locally homeomorphic' to $\mathbb{R}^n$.

Note that $\mathbb{R}^n$ (and any open subset thereof) is a topological n-manifold. We also have that S^n, $\mathbb{RP}^n$, the torus, the Klein bottle and the genus two surface are topological manifolds.

If we let R be as above and $\mathbb{R}$ denote the real numbers with the standard topology, then $N = \mathbb{R} \times R$ is separated and locally homeomorphic to $\mathbb{R}$, but it is *not* second countable with the product topology.

Definition 10.5. A topological space M is called a **topological manifold with boundary** (again of dimension n), if M satisfies

1. M is separated
2. M is second countable
3. For all $x \in M$ one of the following holds, and we enforce that both scenarios occur[1]:
   (a) there exists a neighbourhood U_x and a homeomorphism $f : U_x \to B_1^n(0)$ with $f(x) = 0$
   (b) there exists a neighbourhood U_x and a homeomorphism $f : U_x \to B_1^n(0)^+ = \{x \in B_1^n(0) : x_n \geq 0\}$ with $f(x) = 0$.

If x satisfies $3(a)$ then it is an interior point of M, and if it satisfies $3(b)$ then it is a boundary point. One can check directly from the definition that the interior and boundary of M are disjoint from one-another (no point x can satisfy both (a) and (b)). Moreover the interior of M, denoted $\mathring{M}$, is open and not closed in M, and the boundary is closed. We then have that the boundary points of M are the set $\partial\mathring{M}$ in M, often just written simply as ∂M.

[1]If (b) does not occur then this is no different to the definition of a topological manifold. Furthermore, if (b) holds somewhere then there must be points where (a) holds, by definition.

Beware that this notation is far from being consistent - the boundary of a manifold is not the same definition of topological boundary we have seen in the exercise sheets!

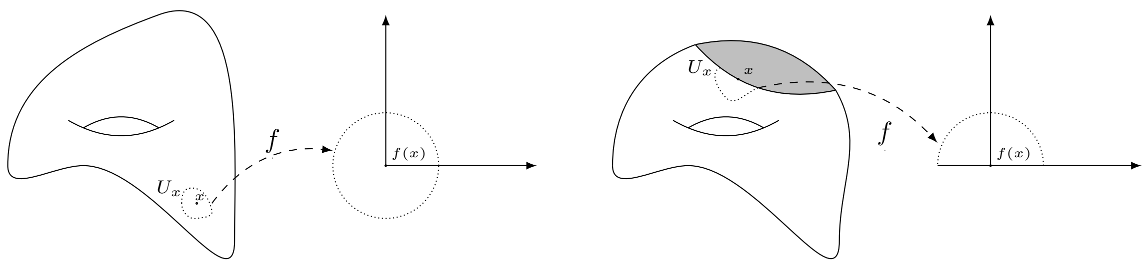
One can again prove using the definition, that the manifold boundary ∂M is a topological $(n-1)$-manifold (without boundary).

The examples of manifolds with boundary that we have seen are the discs $D^n := \{x \in \mathbb{R}^n : |x| \leq 1\}$, the cylinders $[0,1] \times S^n$ and the Möbius band. Be aware that the open ball

$$B^n = \{x \in \mathbb{R}^n : |x| < 1\}$$

and the open cylinder $(0,1) \times S^n$ are topological manifolds *without* boundary! This is in direct conflict with the definition of boundary we defined earlier in the course - so be careful when considering manifolds with boundary verses the boundary of a set $A \subset X$.

Below is an attempt to draw a topological surface without boundary (on the left) and with boundary (on the right - the grey area is *not* part of the surface).



We will deal exclusively with *connected* topological 1 and 2 manifolds.

10.2 (Abstract) 1d tigers

Theorem 10.6. *Let $M = M^1$ be a connected topological 1-manifold either with or without boundary. Then M is homeomorphic to one of four options depending on whether it is compact, or has boundary. To be precise, M is homeomorphic to one of the following*

1. *If $\partial M = \emptyset$ and M is not compact then $M \cong \mathbb{R} \cong (0,1)$.*
2. *If $\partial M = \emptyset$ and M is compact then $M \cong S^1$.*
3. *If $\partial M \neq \emptyset$ and M is not compact then $M \cong [0,\infty) \cong [0,1)$. In this case $|\partial M| = 1$.*
4. *If $\partial M \neq \emptyset$ and M is compact then $M \cong [0,1]$. In this case $|\partial M| = 2$.*

Once again be aware that the notation ∂M does not mean the same thing as the boundary you met in exercise sheet two - here we are thinking of boundary purely in the topological manifold sense defined above.

10.3 (Abstract) 2d tigers

We now consider topological surfaces, or topological 2-manifolds.

(We will from now on largely deal with connected and compact topological surfaces without boundary.)

10.3.1 Preliminaries: how to make a nice warm coat for a puppy or a centipuppy

Recall that a manifold $M = (M, \tau)$ can be equipped with a function

$$U : M \to \tau$$

taking x to a neighbourhood U_x of x that is homeomorphic to an n-ball by some given homeomorphism $f_x : U_x \to B_1^n(0)$. We omit the subscript on f_x where no ambiguity arises, and just write f. A specific quadruple (M, τ, U, f) is called an atlas for M.

Thus f is a collection of functions: for each $x \in M$ we have the local homeomorphism $f : U_x \to B_1^2(0)$.

Definition 10.7. If S, T are sets then $S \coprod T$ denotes (some fixed) disjoint union. Note that this lifts in a natural way to topological spaces.

Definition 10.8. Let M_1, M_2 be two connected and compact topological surfaces without boundary. Fix atlases for both: $A_i = (M_i, \tau_i, U, f)$. Note we do not notationally distinguish the two lots of U, f functions — they are generally distinguishable by their arguments (and we need the space!).

Pick $x_1 \in M_1$ and $x_2 \in M_2$. Let $f_1 : U_{x_1} \to B_1^2(0)$ be the local homeomorphism at x_1, and similarly f_2 for x_2. Define

$$\mathfrak{B}(x_i) := f_i^{-1}(B_{1/2}^2(0))$$

— this is the pre-image of the half-radius 2-ball (i.e. disk) inside our 2-ball. Let $\mathcal{S}_i^1(x_i) := f_i^{-1}(S_{1/2}^1(0))$.

Let $\widetilde{M_1} := M_1 \setminus \mathfrak{B}(x_1)$. This is the manifold with the small 2-ball around x_1 removed (think of puncturing the surface).

Define (omitting x_1, x_2 from notation for now)

$$M_1 \# M_2 \;:=\; \left(M_1 \setminus \mathfrak{B}(x_1)\right) \coprod \left(M_2 \setminus \mathfrak{B}(x_2)\right) /_{\sim}$$

where $x \sim y \iff x = y$ or $x \in \mathcal{S}_1^1(x_1)$, $y \in \mathcal{S}_2^1(x_2)$ and $f_1(x) = f_2(y)$.

This composition is called 'connected sum'.

Notes:
(I) The connected sum $M_1 \# M_2$ is another connected and compact topological surface without boundary.
(II) Up to homeomorphism, $M_1 \# M_2$ is independent of the choice of $x_i \in M_i$, the neighbourhoods U_{x_i} and the homeomorphisms f_i (hence dropping these from the notation).
(III) If $M_1 \cong N_1$ and $M_2 \cong N_2$ then

$$M_1 \# M_2 \cong N_1 \# N_2.$$

We do not prove these statements here. But we can use the definition to write down the classification of compact topological surfaces without boundary:

10.3.2 Preliminaries: group presentations

Recall from (3.2.25) the fundamental groupoid of a space X. The fundamental group of X at $x \in X$ is the group built on paths from x to x. If X is a connected manifold then every such group is isomorphic and we write it $\pi_1(X)$ for it up to isomorphism. In Section 12 we will study these groups in more detail.

Now suppose that $G = \{g_0, g_1, g_2, ...\}$ is any group. Let us assume that g_0 is the identity element — often written e. Any 'word' made up from the elements is another element. For example $w = g_1 g_2 g_1$ (perhaps think of the factors as 'letters' in the word).

Sometimes two such words may give the same element. For example $w = g_0 g_1 = e g_1$ and $w' = g_1$ obey $w = w'$ in the group.

We say a subset G' *generates* G if every element of G is equal in G to a word made up from 'letters' in G' (i.e. elements of G) and their inverses.

The notation

$$G = \langle a, b, c; aba, bbc \rangle$$

(a list of letters then a list of words in the letters and their inverses) means the group generated by a, b, c where the words aba and bbc are both equal to the identity element. (Note that this implies many other 'relations' among words as well.)

For example

$$G = \langle a; aa \rangle$$

gives the group of order 2 ($G = \{e, a\}$ since $aa = e$). Again we give more details in Section 12.

The notation $[a, b]$ means the word $aba^{-1}b^{-1}$. So the relation $[a, b] = e$ means $aba^{-1}b^{-1} = e$ which gives $ab = ba$ — 'commutativity'.

10.3.3 Classification

Theorem 10.9. *Suppose that M is a connected and compact topological surface* ***without boundary****. Then M is homeomorphic to either S^2; a finite connected sum of tori; or a finite connected sum of projective planes $\mathbb{RP}^2$. Moreover any two distinct members of this list are topologically distinct.*

If M is homeomorphic to a connected sum of g tori, it is said to have orientable genus g and

$$\pi_1(M) \cong \langle a_1, b_1, \ldots a_g, b_g; [a_1, b_1][a_2, b_2] \ldots [a_g, b_g] \rangle.$$

Similarly if M is homeomorphic to a connected sum of k projective planes, it is said to have non-orientable genus k and in this case

$$\pi_1(M) \cong \langle a_1, \ldots a_k; a_1^2 a_2^2 \ldots a_k^2 \rangle.$$

Remark 10.10. Some background and complementary facts that are worth bearing in mind when considering the above. Let M be a compact, connected topological surface without boundary. Then

1. S^2 is said to have orientable genus zero.
2. $M \# S^2 \cong M$ for all topological surfaces M.
3. The genus two surface we have seen earlier in the course should really be called the "orientable genus two surface".
4. $\mathbb{RP}^2 \# \mathbb{RP}^2$ is homeomorphic to the Klein bottle.
5. If M has orientable genus g then $M \# \mathbb{RP}^2$ has non-orientable genus $k = 2g + 1$, so is of the form $\mathbb{RP}^2 \# \mathbb{RP}^2 \# \dots \# \mathbb{RP}^2$ for $2g+1$ copies of $\mathbb{RP}^2$ - this is why "cross" terms do not occur in our list.

Chapter 11

Topological Invariants

These are properties of topological spaces which allow us to determine whether two spaces are 'different', in the sense of not-homeomorphic.

11.1 Topological invariants

Definition 11.1. A **topological invariant** is a property of a topological space which is preserved by homeomorphism. Specifically, it is an assignment $\varphi : \{\text{Top. spaces}\} \to \mathcal{S}$, where $\mathcal{S}$ is a set, such that if $X \cong Y$, then $\varphi(X) = \varphi(Y)$.

Remark 11.2. There are several important topological invariants that we will cover in this course:

- the Hausdorff property
- connectedness
- compactness
- the fundamental group.

Anyone interested in taking the study of topology further would do well to learn about homology and cohomology (and for surfaces, the Euler characteristic).

Topological invariants are useful because they let us prove that two spaces are not homeomorphic. Indeed, if $\varphi(X) \neq \varphi(Y)$, then $X \ncong Y$. Note however, we cannot use topological invariants to prove that two spaces are homeomorphic (there are many examples of invariants and spaces such that $\varphi(X) = \varphi(Y)$, but $X \ncong Y$).

Proposition 11.3. *The separated property is a topological invariant.*

Proof. Suppose that $f : X \to Y$ is a homeomorphism between two topological spaces X and Y and suppose in addition that Y is a separated space. We must show that X is a separated space.

Let x_1, x_2 be two distinct points in X. Then $f(x_1), f(x_2)$ are two distinct points in Y, and since Y is separated there exist disjoint open subsets $V_1, V_2 \subset Y$ such that $f(x_i) \in V_i$ for each $i = 1, 2$. Then $U_i = f^{-1}(V_i)$ satisfy $x_i \in U_i$ for each $i = 1, 2$ and $U_1 \cap U_2 = \emptyset$. So X is separated. □

Example 11.4. If X is a set with the indiscrete topology, Y is a set with the discrete topology and X has at least two elements, then X is not homeomorphic to Y. This is because Y is separated and X is not, and the previous proposition says that if X is homeomorphic to a separated space it must be separated.

It is **not** true that two separated spaces must be homeomorphic (for example, if X and Y are both separated but X has two elements and Y three then X and Y cannot be homeomorphic).

Proposition 11.5. *If $f : X \to Y$ is a homeomorphism and Z is a subspace of X then $f_Z : Z \to f(Z)$ is a homeomorphism.*

Proof. f_Z is surjective by definition, and it is injective because f is. Therefore $f_Z : Z \to f(Z)$ is a bijection. f_Z is continuous by Proposition 7.33. The only tricky part is proving that f_Z^{-1} is continuous. To do this we use the following lemma:

Lemma 11.6. *Let $g : X \to Y$ be an injective map between two sets. Then for any subsets $A, B \subset X$, $g(A \cap B) = g(A) \cap g(B)$.*

Proof. You should have already proved in problem set 1 that $g(A \cap B) \subset g(A) \cap g(B)$, so we only need to show that $g(A \cap B) \supset g(A) \cap g(B)$. Let $y \in g(A) \cap g(B)$. Then there exist $x \in A$ and $x' \in B$ such that $g(x) = y$ and $g(x') = y$. Since g is injective, $x = x'$, so $x \in A \cap B$. Therefore $y \in g(A \cap B)$. □

Returning to the proof of Proposition 11.5, suppose that $U \subset Z$ is open. By definition there exists a $V \subset X$ such that $U = V \cap Z$. By the lemma,

$$(f_Z^{-1})^{-1}(U) = f_Z(U) = f(U) = f(V \cap Z) = f(V) \cap f(Z).$$

The set $f(V) = (f^{-1})^{-1}(V)$ is open in Y because f^{-1} is continuous, so $f(V) \cap f(Z)$ is open in $f(Z) \subset Y$. Therefore f_Z^{-1} is continuous. □

Corollary 11.7. *If $f : X \to Y$ is a homeomorphism and $\{x_1, x_2, \ldots, x_n\} \subset X$ then $f : X \backslash \{x_1, x_2, \ldots, x_n\} \to Y \backslash \{f(x_1), f(x_2), \ldots, f(x_n)\}$ is a homeomorphism.*

Proof. Exercise. □

11.2 Connectedness

Definition 11.8. Let X be a topological space. A pair of **nonempty open** subsets U and V of X such that $U \cap V = \emptyset$ and $X = U \cup V$ is called a **partition** of X. X is called **disconnected** if it admits a partition, and **connected** if it does not.

Remark 11.9. Notice that this definition forces the empty topological space to be connected (it certainly does not admit a non-empty subset, let alone a pair of disjoint ones!).

X is connected if and only if the only subsets of X which are both open and closed are X and $\emptyset$ (you will prove this in the exercise sheet).

Example 11.10. Any set X with the indiscrete topology is connected. This is because there is only one non-empty open set (the set X itself).

Example 11.11. If X has at least 2 elements then X with the discrete topology is disconnected. This is because we can write $X = \{x\} \cup (X \setminus \{x\})$ for some $x \in X$; the sets $\{x\}$ and $X \setminus \{x\}$ are both open and do not intersect one another.

Example 11.12. The subspace $Y = [0, 1] \cup [2, 3]$ of $\mathbb{R}$ is disconnected. This is because $U = [0, 1] = (-1, 2) \cap Y$ and $V = [2, 3] = (1, 4) \cap Y$ are open subsets of Y that form a partition.

Example 11.13. The subspace $\mathbb{Q}$ of $\mathbb{R}$ is disconnected. Let a be any irrational number (e.g. $a = \sqrt{2}$); then $U = (-\infty, a) \cap \mathbb{Q}$ and $V = (a, \infty) \cap \mathbb{Q}$ are open subsets of $\mathbb{Q}$ that form a partition.

Proving that a topological space is connected turns out to be quite a bit harder than proving that a space is disconnected.

To help we introduce some terminology:

Definition 11.14. A nonempty subset J of $\mathbb{R}$ is called an **interval** if for every $x, y \in J$, $[x, y] \subset J$.

According to this definition, all of the following subsets of $\mathbb{R}$ are intervals:

$$(a, b),\ (a, b],\ [a, b),\ [a, b],\ (-\infty, a),\ (-\infty, a],\ (a, \infty),\ [a, \infty),\ (-\infty, \infty) = \mathbb{R}.$$

In addition, a singleton set of the form $\{a\}$ for some $a \in \mathbb{R}$ is an interval according to the definition.

Theorem 11.15. *A subset of* $\mathbb{R}$ *is connected if and only if it is an interval.*

Recall that $\mathbb{Q}$ is disconnected, so $\mathbb{R}$ is very different from $\mathbb{Q}$ as a topological space.

The proof of this theorem relies on the property that distinguishes $\mathbb{R}$ from $\mathbb{Q}$, namely that every bounded subset of $\mathbb{R}$ has a least upper bound (called the supremum, or sup — see Chapter 5).

Proof. First we prove the "only if" part:
Suppose that $J \subset \mathbb{R}$ is connected and suppose for contradiction that J is not an interval. Then there exist $x, y \in J$ and $c \notin J$ with $x < c < y$. Let $U = J \cap (-\infty, c)$ and $V = J \cap (c, \infty)$. Then U and V are open $U \cap V = \emptyset$ and $U \cup V = J \setminus \{c\} = J$. So U and V form a partition of J, contradicting the supposition that J is connected.

Now we prove the "if" part:
Suppose that J is an interval, and let U and V be two open non-empty subsets of J such that $U \cap V = \emptyset$. By definition, these take the form $U = U_0 \cap J$ and $V = V_0 \cap J$ for some open subsets U_0, V_0 of $\mathbb{R}$. We will prove that there exists a number $c \in J$ with $c \notin U \cup V$; this means that $J \neq U \cup V$, and hence that J is connected.

Since U and V are non-empty and non-intersecting we may pick points $a \in U, b \in V$ with $a \neq b$. By changing notation if necessary, we may arrange things so that $a < b$. Since J is an interval, $[a, b]$ is a subset of J.

Now let $U_1 = U \cap [a, b] = U_0 \cap [a, b]$ and $V_1 = V \cap [a, b] = V_0 \cap [a, b]$, and let

$$c = \sup U_1$$

be the least upper bound of U_1. We know that $c \in [a, b]$, because $a \in U_1$ and b is an upper bound for U_1. Therefore $c \in J$, and it remains to show that c does not belong to U_1 or V_1.

If $c \in U_1 \subset U_0$ then $c \neq b$, so $c \in [a, b)$. Then there exists an interval (d_1, d_2) with $c \in (d_1, d_2) \subset U_0$, and we may assume $d_2 < b$ by shrinking this interval if necessary. Then $(c + d_2)/2$ is an element of U_1 and is greater than c, contradicting the statement that c is an upper bound for U_1.

If $c \in V_1 \subset V_0$ there exists an interval (e_1, e_2) with $c \in (e_1, e_2) \subset V_0$. So no real numbers $x \in (e_1, e_2)$ belong to U_1. In fact, since c is an upper bound for U_1 and $c > e_1$, no real numbers $x > e_1$ belong to U_1. So $(c + e_1)/2$ is an upper bound for U_1 which is less than c, contradicting the statement that c is a least upper bound. □

11.2.1 Properties of connected spaces

Now we will begin to explore the properties of connected topological spaces

Proposition 11.16. *The continuous image of a connected topological space is connected.*

Proof. Let X and Y be two topological spaces and let $f : X \to Y$ be a continuous function. Suppose that the subspace $f(X) \subset Y$ is disconnected; we must show that X is disconnected.

Since $f(X)$ is disconnected there exist open subset U and V of Y such that $f(X) \cap U$ and $f(X) \cap V$ form a partition of $f(X)$. This means that $f(X) \cap U \neq \emptyset$, $f(X) \cap V \neq \emptyset$, $(f(X) \cap U) \cap (f(X) \cap V) = \emptyset$ and $(f(X) \cap U) \cup (f(X) \cap V) = f(X)$. We claim that $f^{-1}(U)$ and $f^{-1}(V)$ form a partition for X. These sets are open because f is continuous. They are non-empty because $f(X) \cap U$ and $f(X) \cap V$ are. Their intersections and unions are

$$f^{-1}(U) \cap f^{-1}(V) = f^{-1}(U \cap V) = f^{-1}(f(X) \cap U \cap V) = f^{-1}(\emptyset) = \emptyset$$

and

$$f^{-1}(U) \cup f^{-1}(V) = f^{-1}(U \cup V) = f^{-1}(f(X) \cap (U \cup V)) = f^{-1}(f(X)) = X.$$

□

This proposition is useful for proving that spaces are connected.

Example 11.17. The circle $S^1 = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ (with its subspace topology) is connected. This is because it is the image of the continuous function $f : [0, 1] \to \mathbb{R}^2$, $f : t \mapsto (\cos 2\pi t, \sin 2\pi t)$, and $[0, 1]$ is connected.

Proposition 11.16 has two important theoretical consequences. The first is that it generalises the intermediate value theorem that you learnt in your real analysis course.

Corollary 11.18. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then for any number y between $f(a)$ and $f(b)$ there exists an $x \in [a, b]$ such that $f(x) = y$.*

Proof. According to Theorem 11.15 $[a, b]$ is a connected topological space. By Proposition 11.16 its image $f([a, b])$ is connected. By Theorem 11.15 its image is an interval. Since $f(a), f(b) \in f([a, b])$ any y such that $f(a) < y < f(b)$ or $f(b) < y < f(a)$ is an element of $f([a, b])$. Then by definition there exists an $x \in [a, b]$ such that $f(x) = y$. □

The second important consequence is that

Corollary 11.19. *Connectedness is a topological invariant.*

Proof. Let X, Y be two topological spaces, let $f : X \to Y$ be a homeomorphism, and suppose that X is connected. Since f is surjective Y is equal to the image of $f(X)$ of X. By proposition 11.16 $f(X)$ is connected. Therefore Y is connected. □

We can use the property of connectedness to prove that topological spaces are not homeomorphic

Example 11.20. $[a, b]$ is not homeomorphic to (c, d) for any a, b, c, d such that $a \leq b$, $c < d$.

Suppose that $f : [a, b] \to (c, d)$ is a homeomorphism. Then $f_{(a,b]} : (a, b] \to (c, d) \setminus \{f(a)\}$ is also a homeomorphism, by Proposition 11.5. The topological space $(a, b]$ is connected by Theorem 11.15. The topological space $(c, d) \setminus \{f(a)\}$ is not, because it can be written as union of two disjoint non-empty open subsets: $(c, d) \setminus \{f(a)\} = (c, f(a)) \cup (f(a), d)$. These statements contradict corollary 11.19.

Using the same ideas you can prove that (for a, b, c, d as above) $[a, b]$ is not homeomorphic to $(c, d]$. Similarly $(a, b]$ is not homeomorphic to (c, d).

Another useful consequence of 11.16 is the following:

Corollary 11.21. *Let X and Y be two topological spaces. $X \times Y$ is connected if and only if X and Y are connected.*

Proof. ($\Longrightarrow$): Recall from Proposition 7.34 that the projections $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are continuous maps, and that the image of $\pi_X(X \times Y) = X$ and $\pi_Y(X \times Y) = Y$. Therefore X and Y are the images of connected maps, hence connected by Proposition 11.16.

($\Longleftarrow$): Assuming that X and Y are connected, we first note that for fixed $y \in Y$

$$f_y : X \to X \times Y \quad x \mapsto (x, y)$$

is continuous (exercise) and therefore $X \times \{y\} \subset X \times Y$ is connected for each fixed y.

Similarly, $\{x\} \times Y \subset X \times Y$ is connected for each fixed x.

For a contradiction, suppose that $X \times Y$ is not connected, so there exits a non-empty open partition U, V of $X \times Y$.

Claim: For each fixed $x \in X$ we have

$$U \cap (\{x\} \times Y) = \begin{cases} \{x\} \times Y & \text{or} \\ \emptyset. \end{cases}$$

To prove the claim, suppose it is false. In particular $U_x = U \cap \{x\} \times Y$ and $V_x = V \cap \{x\} \times Y$ is a non-empty open partition of $\{x\} \times Y$. This contradicts the connectedness of $\{x\} \times Y$ so the claim is true.

The claim also proves that

$$V \cap (X \times \{y\}) = \begin{cases} X \times \{y\} & \text{or} \\ \emptyset \end{cases}$$

for any fixed $y \in Y$. Now, letting $(x, y) \in U$ and $(x', y') \in V$ we see that $\{x\} \times Y \subset U$ and $X \times \{y'\} \subset V$. Therefore $(x, y') \in U \cap V \neq \emptyset$ and we have our contradiction.

□

11.3 Path-connectedness

In this section we study another topological invariant, called path connectedness. Path connectedness is related to connectedness, and is usually easier to work with. There is a difference, however, and it turns out that path-connectedness is a stronger condition.

Definition 11.22. Let X be a topological space. Given two points $x, y \in X$, a **path from** x **to** y is a continuous function $\alpha : [0,1] \to X$ such that $\alpha(0) = x$ and $\alpha(1) = y$. The point x is called the **initial point** of the path and the point y is called the **final point** of the path.

Definition 11.23. X is called **path-connected** if for any two points $x, y \in X$ there exists a path from x to y. Notice that the empty set automatically satisfies this, so it is itself path connected.

Proposition 11.24. *Any path-connected topological space is connected.*

Proof. Let X be a path-connected topological space. Suppose for contradiction that X is not connected. Let U, V be a partition of X, and let $a \in U$ and $b \in V$. Since X is path-connected there exists a continuous function $\alpha : [0,1] \to X$ such that $\alpha(0) = a$ and $\alpha(1) = b$.

Consider the subspace $\alpha([0,1]) \subset X$. It is straightforward to show that the open subsets $\alpha([0,1]) \cap U$ and $\alpha([0,1]) \cap V$ form a partition for $\alpha([0,1])$: they are nonempty because $a \in \alpha([0,1]) \cap U$ and $b \in \alpha([0,1]) \cap V$; their union is $\alpha([0,1])$ because $U \cup V = X$, and they are disjoint because U and V are. So $\alpha([0,1])$ is disconnected. However, Theorem 11.16 and Proposition 11.15 imply that $\alpha([0,1])$ is connected, because it is the continuous image of a connected set $[0,1]$. So we have a contradiction, and it must be the case that X is connected. $\square$

Using this proposition makes it much easier to check that a space is connected. Be warned however that the converse to Proposition 11.24 is false: connected spaces are not necessarily path-connected. We will give a counter-example below.

Example 11.25. $\mathbb{R}^n$ is path connected, hence connected.

Let x and y be two points in $\mathbb{R}^n$. Then

$$\alpha : [0,1] \to \mathbb{R}^n, \quad \alpha : t \mapsto (1-t)x + ty$$

is a continuous function such that $\alpha(0) = x$ and $\alpha(1) = y$.

Example 11.26. When $n \geq 2$, $\mathbb{R}^n \setminus \{(0,\ldots,0)\}$ is path connected, hence connected.

Let x and y be two points in $\mathbb{R}^n \setminus \{(0,0,\ldots,0)\}$. If x and y do not lie on opposite sides of the origin we may use the function α written down in the previous example to define a continuous $\alpha : ([0,1]) \to \mathbb{R}^n \setminus \{(0,\ldots,0)\}$ such that $\alpha(0) = x$ and $\alpha(1) = y$. If x and y do lie on opposite sides of the origin (i.e. $x = -\lambda y$ for some $\lambda > 0$) we must use a different path. Thinking geometrically, there is always some non-zero vector $m \in \mathbb{R}^n$ so that $x \cdot m = 0$ [1]. In particular $\tilde{x} = x + m$ is non-zero and does not lie on opposite sides of the origin to y or x. Thus we first run a straight line (as above) from x to $\tilde{x}$ and then from $\tilde{x}$ to y. But we do each one at double speed! The following one is suitable:

$$\alpha : t \mapsto \begin{cases} (1-2t)x + 2t\tilde{x} & \text{if } t \in [0,1/2] \\ (2-2t)\tilde{x} + (2t-1)y & \text{if } t \in [1/2,1]. \end{cases}$$

You should check for yourself that this function is continuous (you could use, for example, Lemma 7.35).

Example 11.27. $\mathbb{R}^2$ is not homeomorphic to $\mathbb{R}$.

To prove this, suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ is a homeomorphism. Then the restriction $f_{\mathbb{R}^2 \setminus \{(0,0)\}} : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R} \setminus \{f(0,0)\}$ is also a homeomorphism, by Proposition 11.5. However, $\mathbb{R}^2 \setminus \{(0,0)\}$ is connected and $\mathbb{R} \setminus \{f(0,0)\}$ is not, so they cannot be homeomorphic. Therefore there is no homeomorphism $f : \mathbb{R}^2 \to \mathbb{R}$.

Now we will prove an analog of Proposition 11.16

Proposition 11.28. *The continuous image of a path-connected topological space is path-connected.*

Proof. Let X and Y be topological spaces such that X is path-connected, and let $f : X \to Y$ be a continuous function. We must show that $f(X)$ is path-connected. Let $y_1, y_2 \in f(X)$. Then there exist $x_1, x_2 \in X$ such that $f(x_1) = y_1$ and $f(x_2) = y_2$. Since X is path-connected there exists a path α from x_1 to x_2. Then $f \circ \alpha$ is a path from y_1 to y_2. It is continuous because α is continuous by definition, and the composition of continuous functions is continuous, and it satisfies $f \circ \alpha(0) = y_1$, $f \circ \alpha(1) = y_2$ because $\alpha(0) = x_1$ and $\alpha(1) = x_2$. So $f(X)$ is path-connected. □

Corollary 11.29. *Path-connectedness is a topological invariant.*

Proof. Exercise. □

Proposition 11.30. *Let X and Y be two topological spaces. Then $X \times Y$ is path-connected if and only if X and Y are path-connected.*

Proof. Suppose first that X, Y are path connected. Let (x_1, y_1) and (x_2, y_2) be two points in $X \times Y$. Then there exist paths α fom x_1 to x_2 and β from y_1 to y_2. The function $F : [0,1] \to X \times Y$, $F(t) = (\alpha(t), \beta(t))$ is continuous by Proposition 7.34 and satisfies $F(0) = (x_1, y_1)$ and $F(1) = (x_2, y_2)$, so is a path from (x_1, y_1) to (x_2, y_2). Therefore $X \times Y$ is path-connected.

The proof of the converse (i.e. that $X \times Y$ path connected implies X and Y path connected) is similar to the proof of corollary 11.21 and left as an exercise. □

[1] actually there is an $n-1$-dimensional space of choices for such an m (exercise: prove why this is).

11.3.1 A counterexample, with a little look at embedding

The definitions of connected and path-connected have a somewhat different flavour as formal definitions. But at least intuitively they are related, and indeed path-connected implies connected as we have seen. To show that they are not equivalent, we could build a space that is connected but not path connected. How might we do that? (I could just give you a construction at this point. But actually some of the easiest such construction, while 'easy', do look a bit contrived. So it is illuminating to think 'around' the problem a bit.)

A useful tool to this end is embedding: embedding of one topological space in another. Thus for example there are many ways of embedding a line in $\mathbb{R}^2$. The x-axis is already an embedding of the line in $\mathbb{R}^2$, for example. And with a bit of stretching one could embed $[0, 1]$ as the part of the y-axis from $y = -1$ to $y = 1$ say. This latter is the space

$$Y = \{(0, 2t - 1) \in \mathbb{R}^2 \ : \ t \in [0, 1]\}$$

If we think about the space we get by taking the above two embeddings *together*, perhaps you can see that this combination is both connected and path connected. Meanwhile if we take the x-axis together with the interval of the y-axis from $y = 2$ to $y = 4$ say, then this space is neither.

Another embedding of $\mathbb{R}$ is

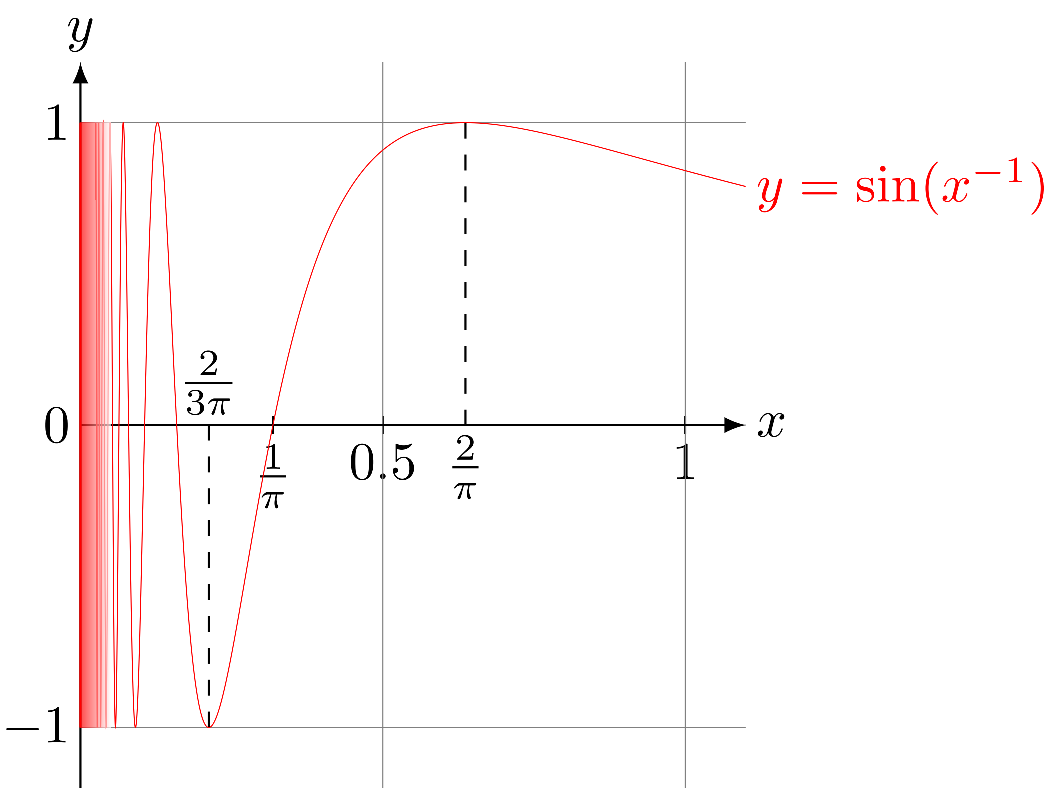
$$R = \{(x, \sin(x)) \ : \ x \in \mathbb{R}\}$$

— this one is curvy but smooth. The combined space $Y \cup R$ is again connected and path-connected.

A 'wilder' embedding of $\mathbb{R}_+$ ($\mathbb{R}_+ = (0, \infty)$) in $\mathbb{R}^2$ is:

$$S = \{(x, \sin(x^{-1})) \ : \ x \in (0, \infty)\}$$

This one looks something like this:



Maybe you like the look of that, and maybe you don't. But an interesting feature of it is that it is now a lot less clear what the relationship of this curve to our old friend Y is. It is clearly 'Y adjacent'. But how do they 'touch'? This combination: $Y \cup S$, is the kind of marginal thing we might be looking for, as we try to build something connected but not path-connected.

Before we look at Y and S together, let's look at just S for a bit longer:
As $x \to 0$, the value of $\sin(1/x)$ will rise and fall, hitting all values in the interval $[-1, 1]$ infinitely many times (see the image above). Thus the limit points of S regarded as a subset of $\mathbb{R}^2$ (recall that a limit point is here any point in $\mathbb{R}^2$ such that every neighbourhood of that point intersects S), that are not actually in S, are given exactly by Y. That is,

$$Y \cup S \ = \overline{S},$$

the closure of S in $\mathbb{R}^2$.

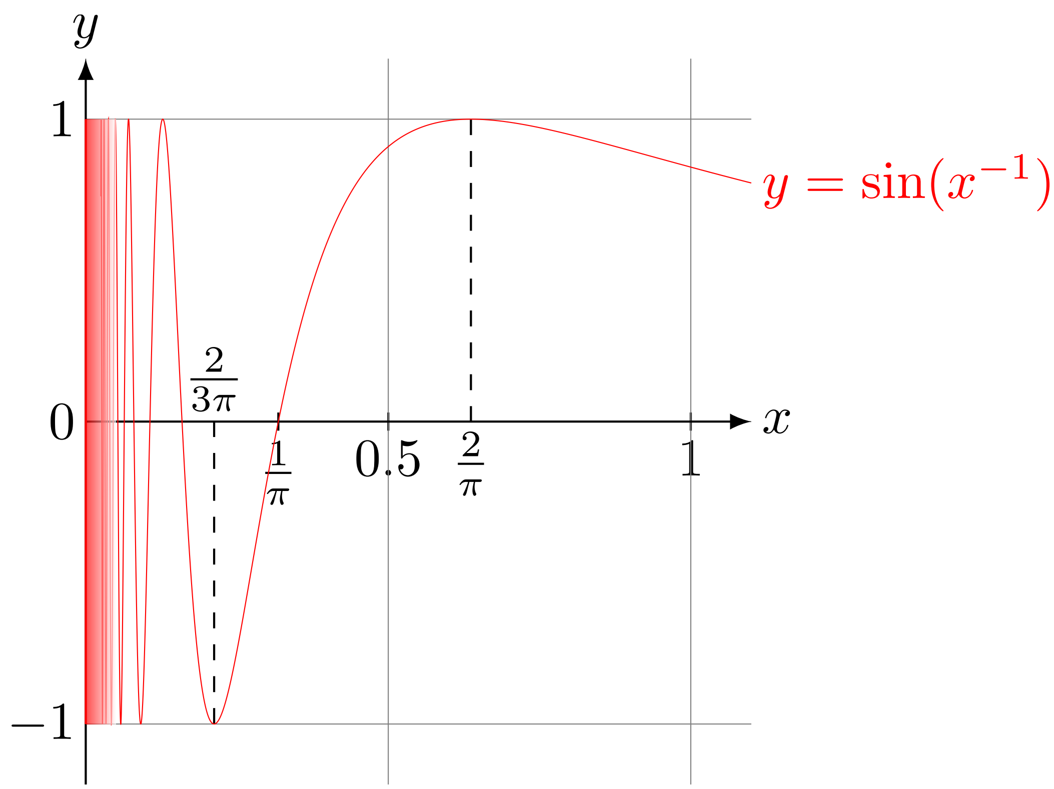
This is interesting on a number of levels. Different embeddings of the same (i.e. homeomorphic) space can have different kinds of closure. (Note that the x-axis embedding of $\mathbb{R}$ was already its own closure.)

Intuition would perhaps suggest that connected implies path-connected. Now we are ready for a counter-example. (The contrived nature of the following example perhaps shows that the intuition still has worth.)

Counterexample 11.31. The "topologist's sine curve". Let $X \subset \mathbb{R}^2$ be the space

$$X = S \cup Y, \quad Y = \{(x, y) \ : \ x = 0, \ y \in [-1, 1]\}, \quad S = \{(x, \sin(1/x)) \ : \ x \in (0, \infty)\}$$

i.e. X is the union of the graph of the function $\sin(1/x)$ with the interval $[-1, 1]$ in the y-axis. See the image below (actually the picture is the same picture as before - the fact that we can't tell the difference is an interesting point about the limitations of 'pictures'!). It turns out that X is connected but not path-connected.



To see that X is connected, suppose for a contradiction that U and V form a partition for X into open subsets. Neither of these sets can equal Y because it is not an open subset of X (consider a small neighbourhood around any point of Y, which includes the intersection of a ball in $\mathbb{R}^2$ around that point with our subspace X). Therefore $U \cap S$, $V \cap S$ form a partition for S. But S is the image of the connected set $(0, \infty)$ under the continuous function $x \mapsto (x, \sin(1/x))$, hence connected, so we have a contradiction.

To see that X is not path-connected we suppose for a contradiction that it is. If that is true then we can pick a continuous path $\alpha : [0, 1] \to X$ from $\alpha(0) = (0, 0)$ to $\alpha(1) = (2/\pi, 1)$. Recall that the projections $\pi_y : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto y$ and $\pi_x : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x$ would then be continuous.

Let

$$t_0 = \sup_{t \in [0,1]} \{\alpha(t) \in Y\} = \sup_{t \in [0,1]} \{(\pi_x \circ \alpha)(t) = 0\}$$

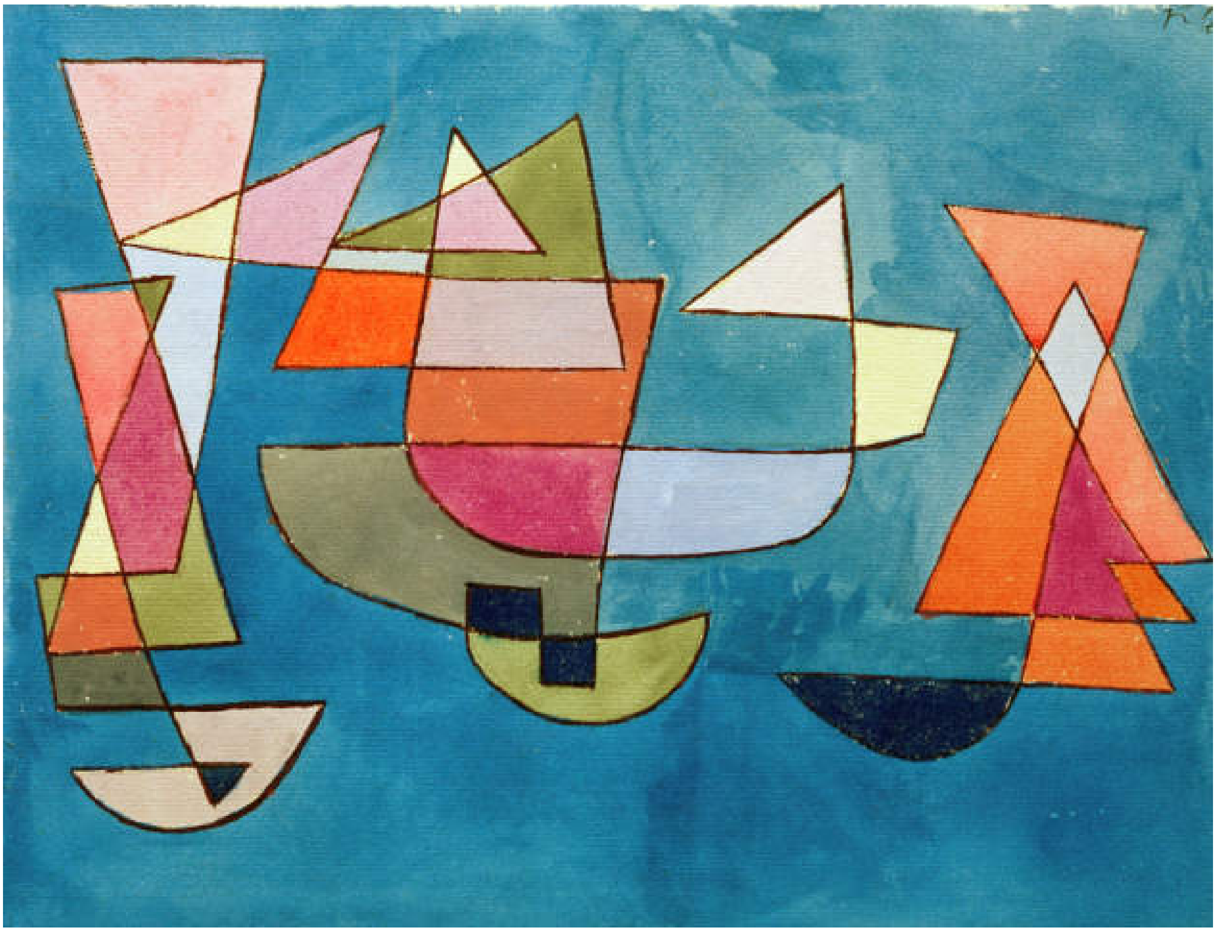
— the l.u.b. of instants spent on Y before heading east for good. By continuity of α we must have $\alpha(t_0) \in Y$. And note $t_0 < 1$. Let $\tilde{\alpha} = \pi_x \circ \alpha : (t_0, 1] \to \mathbb{R}$. By assumption $\tilde{\alpha}(t) > 0$ (equivalently $\alpha(t) \in S$) for all $t \in (t_0, 1]$ and thus $\alpha(t) = (\tilde{\alpha}(t), \sin(1/\tilde{\alpha}(t)))$ for all $t \in (t_0, 1]$. Now notice that $\hat{\alpha} = \pi_y \circ \alpha : [0, 1] \to [-1, 1]$ is continuous. But, for any $w \in [-1, 1]$ we can choose a sequence $\{x_n\} \subset (t_0, 1]$ with $x_n \to t_0$ so that

$$\hat{\alpha}(x_n) = \sin(1/\tilde{\alpha}(x_n)) \to w$$

(to see this pick a w and then imagine drawing a line in $\mathbb{R}^2$ heading eastwards from $P = (0, w)$ — this will intersect S many times, and the t values at which it does so can provide our sequence). Note that this contradicts the continuity of $\hat{\alpha}$ (in particular our conditions imply that $\hat{\alpha}(t_0) = w$ for any $w \in [-1, 1]$!). This contradiction shows that no such continuous α can exist and X is not path connected.

(**11.3.1**) More generally, one can show that (exercise) if $S \subset X$ is connected, then $\overline{S}$ is connected. Again, the above example shows that if $S \subset X$ is path connected, then we do not necessarily have $\overline{S}$ is path connected.

11.4 Compactness



Definition 11.32. Let X be a topological space and let A be a subset of X. An **open cover** for A is an indexed family $\{U_\lambda\}_{\lambda\in\Lambda}$ of **open** subsets of X such that

$$A \subset \bigcup_{\lambda\in\Lambda} U_\lambda.$$

An open cover is called **finite** if the set Λ contains finitely-many elements. If $\{U_\lambda\}_{\lambda\in\Lambda}$ is an open cover for A and $\Lambda' \subset \Lambda$ is such that

$$A \subset \bigcup_{\lambda\in\Lambda'} U_\lambda$$

then $\{U_\lambda\}_{\lambda\in\Lambda'}$ is called a **subcover** of $\{U_\lambda\}_{\lambda\in\Lambda}$.

Definition 11.33. Let X be a topological space and let A be a subset of X. Subset A is called a **compact subset** of X if every open cover for A has a finite subcover. Space X is called a **compact topological space** if the set X is a compact subset of X.

Example 11.34. If X is any set with the indiscrete topology, then every subset of X is compact. This is because the only possible open covers for any subset of X are $\{X\}$ and $\{X, \emptyset\}$; both of these are already finite, so they both admit finite subcovers.

Example 11.35. Let X be any topological space and let A be any subset of X with finitely many elements. Then A is compact.

To prove this, suppose that $\{U_\lambda\}_{\lambda\in\Lambda}$ is any open cover for A. Then for each $x \in A$ we may choose a λ_x such that $x \in U_{\lambda_x}$. Set

$$\Lambda' = \{\lambda_x \,:\, x \in A\}.$$

We claim that $\{U_\lambda\}_{\lambda\in\Lambda'}$ is a finite subcover. We need to check that $A \subset \bigcup_{\lambda\in\Lambda'} U_\lambda$: let $x \in A$; then $x \in U_{\lambda_x}$ and $U_{\lambda_x} \subset \bigcup_{\lambda\in\Lambda'} U_\lambda$, so $x \in \bigcup_{\lambda\in\Lambda'} U_\lambda$ as required. We also need to check that Λ' is finite: this is true because A is finite and $|\Lambda'| \leq |A|$.

Example 11.36. Let X be any set with the discrete topology and let $A \subset X$. Then A is compact if and only if A is finite.

We have already proved the "if" part of this statement. For the "only if" part, suppose that A is compact. Let $U_x = \{x\}$ for any x in A. Then $\{U_x\}_{x\in A}$ is an open cover for A. If we remove one set U_y from the collection $\{U_x\}_{x\in A}$ it will no longer be an open cover for A. So the only subcover of $\{U_x\}_{x\in A}$ is $\{U_x\}_{x\in A}$ itself.

Since A is compact we know that $\{U_x\}_{x\in A}$ admits a finite subcover. Therefore $\{U_x\}_{x\in A}$ must be a finite cover for A. Therefore A has finitely many elements.

Example 11.37. $\mathbb{R}$ with its standard topology is not a compact topological space.

To prove this we need to give an example of an open cover that does not admit a finite subcover. $\{(-n,n)\}_{n\in\mathbb{Z}^+}$ is such an example. It should be clear that this is an open cover. If $M \subset \mathbb{Z}^+$ is a finite subset then $\{(-n,n)\}_{n\in M}$ is not an open cover for $\mathbb{R}$: for example, the real number $x = \max M$ does not belong to $\bigcup_{n\in M}(-n,n)$.

Example 11.38. $\mathbb{R}^n$ with its standard topology is not a compact topological space, for any $n \geq 1$. Set $B^n_M(0) = \{|x| < M\}$. Then $\{B_M\}_{M\in\mathbb{Z}^+}$ is an open cover for $\mathbb{R}^m$ that does not admit a finite subcover.

Example 11.39. The subset $(0,1) \subset \mathbb{R}$ is not compact. $\{(1/3n, 1-1/3n)\}_{n\in\mathbb{Z}^+}$ is an example of an open cover that does not admit a finite subcover.

Proposition 11.40. *Let (X,τ) be a topological space, let $A \subset X$ and let τ_A be the subspace topology on A. Then A is a compact subset of X if and only if (A,τ_A) is a compact topological space.*

Proof. "if" part: suppose that (A,τ_A) is a compact topological space. Let $\{V_\lambda\}_{\lambda\in\Lambda}$ be an open cover for A, where $V_\lambda \in \tau$. Set $U_\lambda = V_\lambda \cap A$; then $\{U_\lambda\}_{\lambda\in\Lambda}$ is a cover for A by open subsets of A. Since A is a compact topological space this admits a finite subcover $\{U_\lambda\}_{\lambda\in\Lambda'}$. Then $\{V_\lambda\}_{\lambda\in\Lambda'}$ is an open subcover of $\{V_\lambda\}_{\lambda\in\Lambda}$.

"only if" part: suppose that A is a compact subset of X. Let $\{U_\lambda\}_{\lambda\in\Lambda}$ be an open cover for A, where $U_\lambda \in \tau_A$. By definition there exist open sets $V_\lambda \in \tau$ such that $U_\lambda = A \cap V_\lambda$. Then $\{V_\lambda\}_{\lambda\in\Lambda}$ is a cover for A by open subsets of X. Since A is a compact subset of X this admits a finite subcover $\{V_\lambda\}_{\lambda\in\Lambda'}$. Then $\{U_\lambda\}_{\lambda\in\Lambda'}$ is an open subcover of $\{U_\lambda\}_{\lambda\in\Lambda}$. □

Are there any compact subsets of $\mathbb{R}$? The following theorem, (which is a special case of the Heine-Borel theorem — see later) shows that there are:

Theorem 11.41. *Every interval $[a,b]$ such that $a < b$ is a compact subset of $\mathbb{R}$.*

Proof. Let $\{U_\lambda\}_{\lambda\in\Lambda}$ be an open cover for $[a,b]$. Clearly, for any $x \in [a,b]$, $\{U_\lambda\}_{\lambda\in\Lambda}$ is an open cover for $[a,x]$. The main idea of this proof is to consider the question: for which $x \in [a,b]$ does this open cover have a finite subcover? We will define

$$C = \big\{x \in [a,b] \,:\, \text{the open cover } \{U_\lambda\}_{\lambda\in\Lambda} \text{ of } [a,x] \text{ admits a finite subcover}\big\}.$$

We claim that there exists an $\varepsilon > 0$ such that

$$[a, a+\varepsilon) \subset C.$$

To prove this, let $\lambda \in \Lambda$ be such that $a \in U_\lambda$. Since U_λ is open there exists an ε such that $(a-\varepsilon, a+\varepsilon) \subset U_\lambda$. Suppose that $x \in [a, a+\varepsilon)$; then $\{U_\lambda\}$ is a finite subcover for $[a, x]$ and $x \in C$. Therefore $[a, a+\varepsilon) \subset C$.

Since the set C is non-empty, we may define

$$s = \sup C.$$

Since $[a, a+\varepsilon) \subset C$, $s \in (a, b]$. We will show below that (i) $s \in C$ and (ii) $s = b$. It follows that the open cover $\{U_\lambda\}_{\lambda\in\Lambda}$ for $[a, b]$ has a finite subcover, and hence that $[a, b]$ is compact.

(i) Let us show that $s \in C$. Let $\mu \in \Lambda$ be such that $s \in U_\mu$. Since U_μ is open we may choose a $\delta > 0$ such that $(s-\delta, s+\delta) \subset U_\mu$. There must exist an $x \in (s-\delta, s]$ such that $x \in C$, for otherwise $s-\delta$ would be an upper bound for C that is strictly less than s. Since this x belongs to C there exists a finite subcover $\{U_\lambda\}_{\lambda\in\Lambda'}$ for $[a, x]$. Then $\{U_\lambda\}_{\lambda\in\Lambda'\cup\{\mu\}}$ is a finite subcover for $[a, s]$, so $s \in C$.

(ii) Now we show that $s = b$. Suppose for contradiction that $s < b$. Again, let $\mu \in \Lambda$ be such that $s \in U_\mu$ and let $\delta > 0$ be such that $(s-\delta, s+\delta) \subset U_\mu$. Now let y be such that $y \in (s, s+\delta)$ and $y < b$. We claim that $y \in C$, in contradiction with the fact that s is an upper bound for C. To prove this claim, let $\{U_\lambda\}_{\lambda\in\Lambda''}$ be a finite subcover for $[a, s]$; then $\{U_\lambda\}_{\lambda\in\Lambda''\cup\{\mu\}}$ is a finite subcover for $[a, y]$ so $y \in C$. Since the assumption $s < b$ lead to a contradiction it must be that $s = b$. □

Now we explore some properties of compact topological spaces. The following is similar to Propositions 11.16 and 11.28 above

Proposition 11.42. *The continuous image of a compact set is compact.*

Proof. Let X, Y be topological spaces, let $f : X \to Y$ be a continuous function and let $A \subset X$ be a compact subset of X. We must show that $f(A)$ is a compact subset of Y.

Let $\{U_\lambda\}_{\lambda\in\Lambda}$ be an open cover for $f(A)$. Then $\{f^{-1}(U_\lambda)\}_{\lambda\in\Lambda}$ is an open cover for A, because

$$\bigcup_{\lambda\in\Lambda} f^{-1}(U_\lambda) = f^{-1}\left(\bigcup_{\lambda\in\Lambda} U_\lambda\right) \supset f^{-1}(f(A)) \supset A.$$

Since A is compact this open cover admits a finite subcover $\{f^{-1}(U_\lambda)\}_{\lambda\in\Lambda'}$. Then $\{U_\lambda\}_{\lambda\in\Lambda'}$ is a finite subcover for $f(A)$, because

$$f(A) \subset f\left(\bigcup_{\lambda\in\Lambda'} f^{-1}(U_\lambda)\right) = f\left(f^{-1}\left(\bigcup_{\lambda\in\Lambda'} U_\lambda\right)\right) \subset \bigcup_{\lambda\in\Lambda'} U_\lambda.$$

□

Proposition 11.42 may be used to prove that spaces are compact:

Example 11.43. The circle $S^1 = \{x \in \mathbb{R}^2 : |x|^2 = 1\}$ is compact. To prove this, note that S^1 is equal to the image of the following continuous function f from the compact topological space $[0, 1]$ to $\mathbb{R}^2$:

$$f : t \mapsto (\cos 2\pi t, \sin 2\pi t).$$

Corollary 11.44. *Compactness is a topological invariant.*

Proof. Exercise. □

The compact open category.

Next we investigate compactness of the product topology.

Theorem 11.45 (Tychonoff)**.** *Let X and Y be two topological spaces. Then $X \times Y$ is compact if and only if X and Y are compact.*

Proof. For the "only if" part, suppose that $X \times Y$ is compact. Recall from Proposition 7.34 that the projections $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are continuous. Then X and Y are compact by Proposition 11.42 because they are equal to the images of the projection maps.

For the "if" part, suppose that X and Y are compact and let $\{W_\lambda\}_{\lambda\in\Lambda}$ be an open cover for $X \times Y$. For each $(x, y) \in X \times Y$ there exists a $\lambda_{(x,y)} \in \Lambda$ such that $(x, y) \in W_{\lambda_{(x,y)}}$. By definition there exist open sets $U_{(x,y)} \subset X$ and $V_{(x,y)} \subset Y$ such that $(x, y) \in U_{(x,y)} \times V_{(x,y)} \subset W_{\lambda_{(x,y)}}$.

Consider the open cover

$$\{U_{(x,y)} \times V_{(x,y)}\}_{(x,y)\in X\times Y}.$$

We will show that this open cover has a finite subcover $\{U_{(x,y)} \times V_{(x,y)}\}_{(x,y)\in\Omega}$, where $\Omega \subset X \times Y$ is a finite subset. It will follow that the original open cover $\{W_\lambda\}_{\lambda\in\Lambda}$ has a finite subcover $\{W_{\lambda_{(x,y)}}\}_{(x,y)\in\Omega}$.

For each $b \in Y$ the family $\{U_{(x,b)}\}_{x\in X}$ is an open cover for X. Therefore it admits a finite subcover $\{U_{(x,b)}\}_{x\in\Xi_b}$, where Ξ_b is a finite subset of X. We will define

$$V_b = \bigcap_{x\in\Xi_b} V_{(x,b)}.$$

Now $\{V_y\}_{y\in Y}$ is an open cover for Y. Therefore it admits a finite subcover $\{V_y\}_{y\in\Upsilon}$, where Υ is a finite subset of Y.

We claim that

$$\{U_{(x,y)} \times V_{(x,y)} \ : \ y \in \Upsilon \text{ and } x \in \Xi_y\}.$$

is a finite open cover of $X \times Y$. To prove this, note first that

$$X \times V_y \subset \bigcup_{x\in\Xi_y} U_{(x,y)} \times V_y \subset \bigcup_{x\in\Xi_y} U_{(x,y)} \times V_{(x,y)}.$$

It follows that

$$X \times Y \subset \bigcup_{y\in\Upsilon} X \times V_y \subset \bigcup_{y\in\Upsilon} \bigcup_{x\in\Xi_y} U_{(x,y)} \times V_{(x,y)}.$$

□

Remark 11.46. It follows by induction that any finite product $X_1 \times X_2 \times \ldots \times X_n$ of compact topological spaces is compact.

Example 11.47. The rectangle $[a, b] \times [c, d]$ is compact for any $a < b$ and $c < d$. Equally the n-dimensional cubes

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$$

are compact.

11.4.1 Closed sets and compact sets; and the separated/Hausdorff property

There are close relations between closed subsets of topological spaces and compact subsets of topological spaces. We will investigate some of these below.

Proposition 11.48. *A closed subset of a compact space is compact.*

Proof. Let X be a compact topological space and let C be a closed subset of X. Let $\{U_\lambda : \lambda \in \Lambda\}$ be an open cover for C. We must find a finite subcover.

The collection $\{U_\lambda : \lambda \in \Lambda\} \cup \{X \setminus C\}$ is an open cover for X. Therefore it admits a finite subcover of the form $\{U_\lambda : \lambda \in \Lambda'\} \cup \{X \setminus C\}$, where Λ' is a finite subset of Λ. Then $\{U_\lambda : \lambda \in \Lambda'\}$ is a finite open cover for C, because $C \cap (X \setminus C) = \emptyset$. □

Example 11.49. $S^n = \{x \in \mathbb{R}^{n+1} : |x|^2 = 1\}$ is a closed subset of the compact space $[-1,1] \times \cdots \times [-1,1]$, since $S^n = f^{-1}(\{1\})$, where

$$f(x) = |x|^2$$

is continuous. Therefore S^n is compact.

Remark 11.50. Both hypotheses in this theorem are necessary:

- $(0,1)$ is a subset of the compact space $[0,1]$ but is not compact.
- $[0,\infty)$ is a closed subset of $\mathbb{R}$ but is not compact.

Note that in a compact space, compact does not imply closed:

Counterexample 11.51. Let $X = \{1,2\}$ and let $\tau = \{\emptyset, \{1\}, \{1,2\}\}$. Then X is compact because it contains finitely many elements. The set $\{1\}$ is compact (because it is finite), but is not closed (because $\{2\} \notin \tau$).

However, compact *does* imply closed in a separated space:

Proposition 11.52. *A compact subset of a separated space is closed.*

Proof. Let X be a separated space and let A be a compact subset of X. We must show that $X \setminus A$ is open.

Let $x \in X \setminus A$. For any $a \in A$ there exist disjoint open sets U_a, V_a such that $a \in U_a$ and $x \in V_a$. The family $\{U_a\}_{a \in A}$ forms an open cover for A. Since A is compact, it admits a finite subcover $\{U_{a_1}, U_{a_2}, \ldots, U_{a_n}\}$ (where $a_1, \ldots, a_n$ are finitely many points in A). Define

$$V_x = \bigcap_{i=1}^{n} V_{a_i}.$$

This set is open, because it is a finite intersection of open sets. Moreover, $x \in V_x$ and $V_x \cap A = \emptyset$ (the latter because V_x is disjoint from all the sets U_{a_i} in the finite open cover of A).

We claim that

$$X \setminus A = \bigcup_{x \in X \setminus A} V_x.$$

$X \setminus A \subset \bigcup_{x \in X\setminus A} V_x$ because every $x \in X \setminus A$ belongs to at least one set V_x in the union, and $X \setminus A \supset \bigcup_{x \in X \setminus A} V_x$ because $V_x \cap A = \emptyset$ for all x. So $X \setminus A$ is a union of open sets, hence open. □

Corollary 11.53. *In a separated space, arbitrary intersections of compact subsets are compact.*

Proof. Let X be a separated space and let $\{A_\lambda\}_{\lambda\in\Lambda}$ be a family of compact subsets of X. The sets A_λ are closed, by Proposition 11.52. The intersection $\bigcap_{\lambda\in\Lambda} A_\lambda$ is closed, by Theorem 7.14. Pick some $\mu \in \Lambda$; then $\bigcap_{\lambda\in\Lambda} A_\lambda$ is a closed subset of the compact space A_μ, hence compact by Proposition 11.48. □

Without the Hausdorff condition, it is not even true that finite intersections of compact subsets are compact. Here's a counterexample that shows why:

Counterexample 11.54. Let X be any infinite set and let a and b be any two distinct points in X. Let $\tau = \{U \subset X \setminus \{a, b\}\} \cup \{X\}$. Then τ defines a topology on X (check this!).

Let $A = X \setminus \{a\}$ and $B = X \setminus \{b\}$. Then A is compact: any open cover for A must contain the set X because X is the only open set containing b; therefore any open cover for A has the finite subcover $\{X\}$. Similar reasoning shows that B is compact.

However, $A \cap B$ is not a compact topological space (and hence not a compact subset of X). This is because the subspace topology on $A \cap B$ is equal to the discrete topology on $A \cap B$. We showed earlier that the only compact subsets in the discrete topology are the finite sets, so $A \cap B$, being an infinite set, is not compact.

Corollary 11.55. *A continuous bijection from a compact topological space to a separated space is a homeomorphism.*

Proof. Let X be a compact space, let Y be a separated space and let $f : X \to Y$ be a continuous bijection. We must show that f^{-1} is continuous.

f^{-1} is continuous if and only if for every closed subset $C \subset X$, $(f^{-1})^{-1}(C) = f(C)$ is closed. Let $C \subset X$ be any closed subset. Then C is compact by Theorem 11.48, $f(C)$ is compact by Proposition 11.42, and $f(C)$ is closed by Proposition 11.52. □

The above result is **very** useful.

Example 11.56. Going back to Example 7.40, setting $F : [0, 1] \to S^1$ to be

$$F(x) = (\cos(2\pi x), \sin(2\pi x))$$

then by Proposition 7.50 F induces a continuous bijection $\tilde{F} : {}^{[0,1]}/_{\sim} \to S^1$ which must be a homeomorphism since the domain is compact, and the target is Hausdorff. In particular ${}^{[0,1]}/_{\sim} \cong S^1$.

Corollary 11.57. *The only separated topology on a finite set is the discrete topology.*

Proof. Let X be a finite set and let τ be a separated topology on X. Let τ_0 denote the discrete topology on X. Let $f : (X, \tau_0) \to (X, \tau)$ be the function $f : x \mapsto x$. Then f is a bijection and is continuous (because any function from a set with the discrete topology is continuous). The topological space (X, τ_0) is compact because X is finite and the topological space (X, τ) is separated by assumption, so by corollary 11.55 f is a homeomorphism. It follows that $\tau = \tau_0$. □

One may wonder how these results come in handy in practice. Corollary 11.55 is very powerful as it enables us to show homeomorphisms without having to explicitly construct them or their inverse.

11.4.2 Compactness in metric spaces

The first result coming up below shows that continuous functions defined on a compact metric space satisfy a strengthened notion of 'uniform' continuity:

(**11.4.1**) Let (X, d_1), (Y, d_2) be metric spaces. Let $f : X \to Y$ be continuous. Then f is *uniformly continuous* if for all $\varepsilon > 0$ there exists $\delta > 0$ s.t. for any $x \in X$, $B_\delta(x) \subset f^{-1}(B_\varepsilon(f(x)))$.

The point here is that $\delta > 0$ is uniform in x (not choosable afresh for each x). For instance the function $f : (0,1) \to \mathbb{R}$, $x \mapsto 1/x$ is of course continuous, but *not* uniformly continuous.

Lemma 11.58. *Let (X, d_1), (Y, d_2) be metric spaces with X compact. Let $f : X \to Y$ be continuous, then f is uniformly continuous.*

Remark 11.59. In the simple case that $f : [a, b] \to \mathbb{R}$ is continuous, the Lemma before this one shows that f is uniformly continuous. In the usual analysis language: for all $\varepsilon > 0$ there exists $\delta > 0$ so that for any $x, y \in [a, b]$, $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$.

On the other hand, we noted that the function $f : (0, 1) \to \mathbb{R}$, $x \mapsto 1/x$ is *not* uniformly continuous ...and of course $(0, 1)$ is not compact.

Proof. Let $\varepsilon > 0$. f is continuous so for all $x \in X$ there exist $\delta_x > 0$ (that depends on x) s.t. $B_{2\delta_x}(x) \subset f^{-1}(B_{\varepsilon/2}(f(x)))$. The family $\{B_{\delta_x}(x)\}_{x\in X}$ covers X and thus admits a finite subcover, given by $x_1, \ldots, x_n$ say (with their associated $\delta_1 \ldots \delta_n$). Set $\delta = \min_i \delta_i > 0$.

Now pick $x \in X$ and notice that $x \in B_{\delta_i}(x_i)$ for some i.

For arbitrary $y \in B_\delta(x)$ then

$$d_1(y, x_i) \le d_1(y, x) + d_1(x, x_i) < \delta + \delta_i \le 2\delta_i.$$

This implies that $y \in B_{2\delta_i}(x_i)$ so that $f(y) \in B_{\varepsilon/2}(f(x_i))$. But since $x \in B_{\delta_i}(x_i)$ we also have $f(x) \in B_{\varepsilon/2}(f(x_i))$ giving

$$d_2(f(x), f(y)) \le d_2(f(x), f(x_i)) + d_2(f(x_i), f(y)) < \varepsilon.$$

In particular $f(y) \in B_\varepsilon(f(x))$ and we are done: for all $\varepsilon > 0$ there exists $\delta > 0$ s.t. for any $x \in X$, $B_\delta(x) \subset f^{-1}(B_\varepsilon(f(x)))$. ☐

Definition 11.60. Let (X, d) be a metric space and let A be a subset of X. Then A is called **bounded** if there exists an $x \in X$ and radius $M > 0$ such that

$$A \subset B_M(x).$$

(**11.4.2**) For example, a finite interval of the real line is bounded, open or closed.
But the x-axis is not a bounded subset of $\mathbb{R}^2$.

Proposition 11.61. *Every nonempty compact subset of a metric space is closed and bounded.*

Proof. Let (X, d) be a metric space and let A be a compact subset of X. The topology on X is separated by Theorem 7.18, so A is closed by Proposition 11.52. It remains to show that A is bounded.

Let x be any point in A. We note that $A \subset \bigcup_{n\in\mathbb{Z}^+} B_n(x)$, because if $y \in A$ then $y \in B_n(x)$ for any n large enough that $n > d(x, y)$. So $\{B_n(x)\}_{n\in\mathbb{Z}^+}$ form an open cover for A, and since A is compact there exists a finite subset $\Gamma \subset \mathbb{Z}^+$ such that $\{B_n(x)\}_{n\in\Gamma}$ is a finite subcover. It follows that $A \subset B_M(x)$, where M is equal to the largest element in Γ. □

In general it is not true that closed and bounded subsets of metric spaces are compact, however, there is one important example of a metric space whose closed and bounded subsets are compact.

Theorem 11.62 (Heine-Borel)**.** *A nonempty subset of* $\mathbb{R}^n$ *(with the standard topology) is compact if and only if it is closed and bounded.*

Proof. Proposition 11.52 shows that if $A \subset \mathbb{R}^n$ is compact it must be closed and bounded. Suppose then that A is a closed and bounded subset of $\mathbb{R}^n$; we must show that A is compact. Since A is bounded there exists an $a \in A$ and $M > 0$ such that $A \subset B_M(a)$. But

$$B_M(a) \subset [a_1 - M, a_1 + M] \times [a_2 - M, a_2 + M] \times \ldots \times [a_n - M, a_n + M].$$

Each set $[a_i - M, a_i + M]$ is compact by Theorem 11.41, so the product is compact by Tychonoff's Theorem 11.45. So A is a closed subset of a compact set, and therefore compact by Proposition 11.48. □

Warning 11.63. A closed and bounded set is **not** compact in a general metric space. For example a set X with the discrete topology is the topological space induced by the discrete metric

$$d(x, y) = \begin{cases} 1, & x \neq y, \\ 0, & x = y. \end{cases}$$

X is bounded in this metric since $X = B_2(x)$ for any $x \in X$. However, X is compact if and only if X is finite, hence for infinite discrete spaces, we have a counter example.

A better example is the space of bounded infinite sequences

$$l^\infty = \{\{x_n\}_{n\in\mathbb{Z}^+} \subset \mathbb{R} : \sup_n |x_n| < \infty\}$$

with the metric

$$d(\{x_n\}, \{y_n\}) = \sup_n |x_n - y_n|.$$

You can check that this is a metric if you like (this space is an example of an interesting class of spaces sometimes called *Banach spaces*, and is infinite-dimensional as a vector space).

Consider the set

$$D_1(0) = \{\{x_n\} : d(\{x_n\}, \{0\}) \leq 1\}$$

which turns out to be closed and bounded. However, this is by no means compact: let $\{x_n\}_{\lambda_\pm} = \{0, \ldots, 0, \pm 1, 0 \ldots 0, \ldots\}$ be the sequence with ± 1 in the $\lambda_\pm^{th}$ position. Then

$$\{B_1(\{x_n\}_{\lambda_\pm}), B_1(\{0\})\}_{\lambda_\pm\in\mathbb{Z}^+}$$

is an open cover of D_1 with no finite subcover.

You may recall that connectedness provided a means to generalise a fundamental theorem from real analysis, the *intermediate value theorem*. The notion of compactness allows us to generalise another important theorem of real analysis: the *maximum value theorem*.

Corollary 11.64 (Maximum value theorem)**.** *Let X be a compact topological space and let $f : X \to \mathbb{R}$ be a continuous function. Then there exist points $c, d \in X$ such that $f(c) \leq f(x) \leq f(d)$ for all $x \in X$.*

In the particular case $X = [a, b]$ for some $a < b$ this is just the usual maximum value theorem.

Proof. Let $A = f(X)$. By Proposition 11.42, A is a compact subset of $\mathbb{R}$. Thus A is closed and bounded by Heine-Borel. Let $m = \inf A$ and $M = \sup A$ so that $A \subset [m, M]$. By Lemma 8.10 $m, M \in A$, thus there exist $c, d \in X$ s.t. $f(c) = m$, $f(d) = M$. $\square$

11.4.3 Lebesgue numbers and sequential compactness

Again we will focus on metric spaces (X, d). Our goal is to prove that a certain notion of **sequential compactness** is equivalent to compactness in this setting:

Definition 11.65. A subspace $C \subset X$ of a metric space (X, d) is **sequentially compact** if for any sequence $\{x_n\} \subset C$, there exists a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to x$ for some $x \in C$.

One can see using Theorem 8.8 and Lemma 8.10 that this implies C is closed.

More directly intuitively, to get an idea about sequential compactness, consider $\mathbb{R}$ and $[0, 1] \subset \mathbb{R}$.
For $\mathbb{R}$ itself the sequence $1, 2, 3, 4, ...$ has no subsequence with a limit. But note that we need an infinite amount of 'space' for a sequence like this.
Note that if we want a sequence $\{x_n\}$ that disproves sequential compactness then it cannot visit the same point infinitely many times (else this point can be the limit point of the subsequence).
Hence the (infinite) sequence must visit infinitely many distinct points (possibly multiple times each, but finitely many times each).
But then if we try to do this - visiting infinitely many distinct points - in a compact space (e.g. a finite closed interval like $[0, 1]$) the average separation of points (across all points in the sequence) tends to zero. In that case there will be ordered subsets with the same property. So either there is a limit of a subsequence or there is a limit but it lies just outside the subset (i.e. the subset is open).
This is roughly the idea. Now let's look at it formally.

Theorem 11.66. *A subspace $C \subset X$ of a metric space (X, d) is compact if and only if it is sequentially compact.*

We will start with the first half of the above theorem:

Proposition 11.67. *If a subspace $C \subset X$ of a metric space (X, d) is compact then it is sequentially compact.*

Proof. First of all let's re-phrase what it means to find a convergent subsequence. There is a convergent subsequence $\{x_{n_k}\}$ if and only if: there exists $x_0 \in C$ so that for all $\varepsilon > 0$, there is a $K \in \mathbb{Z}^+$ so that $k \geq K$ implies $x_{n_k} \in B_\varepsilon(x_0)$. This is true if and only if: there exists $x_0 \in C$ so that for all $\varepsilon > 0$, $\sup\{n : x_n \in B_\varepsilon(x_0)\} = \infty$.

By this reasoning, if there is no convergent subsequence: for any $y \in C$ there exists $\varepsilon(y) > 0$ such that $\sup\{n : x_n \in B_\varepsilon(y)\} < \infty$.

The family $\{B_{\varepsilon(y)}(y)\}_{y \in C}$ is surely an open cover of C to which there exists a finite subcover $\{B_{\varepsilon(y_i)}(y_i)\}_{i=1}^L$. But notice that for each i, there are only finitely many elements of the sequence in each $B_{\varepsilon(y_i)}(y_i)$, and thus the sequence only has finitely many elements in C. This contradiction finishes the proof. □

Definition 11.68. Let $\{U_\lambda\}_{\lambda\in\Lambda}$ be an open cover of a subset $A \subset X$ of a metric space (X, d). We call $\varepsilon > 0$ a **Lebesgue number** of the open cover if: for all $x \in A$ there exits $\lambda' \in \Lambda$ so that $B_\varepsilon(x) \subset U_{\lambda'}$.

Proposition 11.69. *Suppose that $C \subset X$ is a sequentially compact subset of a metric space (X, d). Then for any open cover $\{U_\lambda\}_{\lambda\in\Lambda}$ of C, there exits a Lebesgue number $\varepsilon > 0$ for $\{U_\lambda\}_{\lambda\in\Lambda}$.*

Proof. Suppose the statement is false. In other words, there exits an open cover $\{U_\lambda\}_{\lambda\in\Lambda}$ of C, such that that for all $\varepsilon > 0$, there exits $x \in C$ with $B_\varepsilon(x) \not\subset U_\lambda$ for all λ.

For this open cover, let $\varepsilon(n) = 1/n$ for $n \in \mathbb{Z}^+$ and pick x_n so that

$$B_{\varepsilon(n)}(x_n) = B_{1/n}(x_n) \not\subset U_\lambda \tag{11.1}$$

for all λ. Since C is sequentially compact, this sequence $\{x_n\}$ admits a convergent subsequence $\{x_{n_k}\}$ converging to some $x \in C$. Pick λ' so that $x \in U_{\lambda'}$ and notice that for all $\eta > 0$ there exists K so that $k \geq K$ implies $x_{n_k} \in B_\eta(x)$.

Since η is arbitrary we pick it so that $B_{2\eta}(x) \subset U_{\lambda'}$ (which is possible since $U_{\lambda'}$ is open). Now we can pick $K' \geq K$ so that $1/n_k < \eta$ whenever $k \geq K'$.

We have, for $k \geq K'$, $B_{1/n_k}(x_{n_k}) \subset B_{2\eta}(x) \subset U_{\lambda'}$ which contradicts (11.1). □

We now finish the proof of Theorem 11.66 by proving the converse to Proposition 11.67.

Proof. Suppose that C is sequentially compact and let $\{U_\lambda\}_{\lambda\in\Lambda}$ be an arbitrary open cover. By Proposition 11.69 there exits a Lebesgue number $\varepsilon > 0$ for this open cover.

We now **claim** that there exits a finite number of points $x_1, \ldots, x_n$ so that $C \subset \cup_{i=1}^n B_\varepsilon(x_i)$ and thus, for all i there exits U_{λ_i} such that $B_\varepsilon(x_i) \subset U_{\lambda_i}$. In particular $C \subset \cup_{i=1}^n U_{\lambda_i}$ and we have found our finite subcover.

To prove the claim, suppose that it is false. Now pick x_1 arbitrary: we must have $C \not\subset B_\varepsilon(x_1)$. Thus there is $x_2 \in C\backslash B_\varepsilon(x_1)$ and again we must have $C \not\subset B_\varepsilon(x_1) \cup B_\varepsilon(x_2)$. Since the claim is false, for all $n \in \mathbb{Z}^+$, we can inductively find $x_n \in C\backslash \cup_{i=1}^{n-1} B_\varepsilon(x_i)$ such that $C \not\subset \cup_{i=1}^n B_\varepsilon(x_i)$.

This sequence $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\}$ with $x_{n_k} \to x \in C$, by assumption. Since we have convergence of this subsequence, for any $\eta > 0$ there exits $K \in \mathbb{Z}^+$ so that $k \geq K$ implies $x_{n_k} \in B_\eta(x)$. So pick $\eta = \varepsilon/2$ and notice that $x_{n_k} \in B_{\varepsilon/2}(x) \subset B_\varepsilon(x_{n_k})$ for all $k \geq K$. But this contradicts that $x_{n_{k+1}} \notin B_\varepsilon(x_{n_k})$. □

Chapter 12

Homotopy theory

12.1 Paths and Homotopies

Here $\mathbb{I}^n$ means $[0, 1]^n$ with the corresponding Euclidean metric topology, as before.

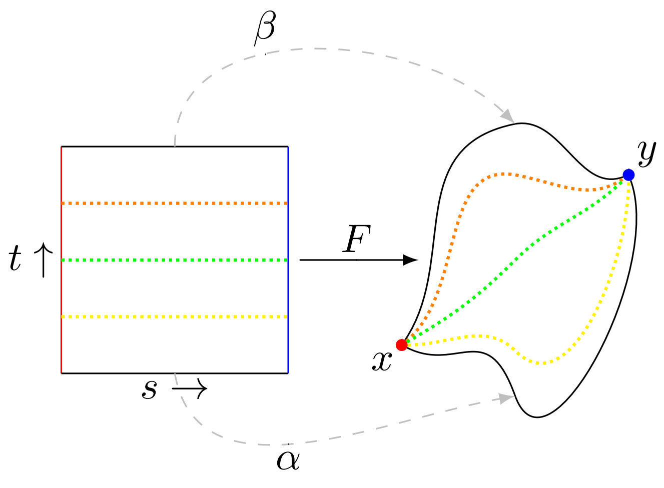
Remember from §3.2.3 and §11.3 that a path from x to y in a space (X, τ) is an element $f \in \mathsf{Top}(\mathbb{I}, (X, \tau))$ (thus $f \in hom([0, 1], X)$) such that $f(0) = x$ and $f(1) = y$.

Remarks: The use of the particular interval $[0, 1]$ for the domain of a path is just an arbitrary choice here. The term 'path' is a bit of a misnomer since we pay attention to the 'time' as we travel along — so it is more like a journey. However we have not said anything about units, and shortly we will make an equivalence which will make a version in which (amongst other things) the details of the rate of travel are washed out.

Definition 12.1. Let (X, τ) be a topological space. Let $x, y \in X$ and let α, β be two paths from x to y. Then α is **path homotopic** to β (written $\alpha \simeq \beta$) if there exists a continuous function $F : \mathbb{I}^2 \to (X, \tau)$ such that

$$\begin{aligned} F(s, 0) &= \alpha(s) \quad \forall s \in [0, 1] \\ F(s, 1) &= \beta(s) \quad \forall s \in [0, 1] \\ F(0, t) &= x \quad \forall t \in [0, 1] \\ F(1, t) &= y \quad \forall t \in [0, 1]. \end{aligned}$$

Such a function F is a **path homotopy** from α to β.



Think of F as being a path of paths from α to β: for **fixed** t, the path $f_t(s) = F(s, t)$ is another path from x to y. If we let t vary from 0 to 1 the resulting paths f_t will go from $\alpha = f_0$ to $\beta = f_1$. The above diagram is showing $f_{1/4}$ (yellow), $f_{1/2}$ (green) and $f_{3/4}$ (orange).

Remarks.

In general the intermediate paths f_t may intersect one another.

It is possible that F doesn't change the "image" of α as t moves, but just the parametrisation. For example, consider

$$F(s,t) = \begin{cases} \alpha\left(\frac{2s}{2-t}\right) & \text{if } s \leq 1 - t/2 \\ \alpha(1) & \text{if } s \geq 1 - t/2 \end{cases}$$

One can show that this is a continuous function, and hence gives a homotopy from $\alpha(s) = F(s,0)$ to

$$\beta(s) = \begin{cases} \alpha(2s) & \text{if } s \leq 1/2 \\ \alpha(1) & \text{if } s \geq 1/2 \end{cases}$$

which is just a re-parametrisation of α (β runs along α fast and then waits at the end point!).

This kind of construction raises the question of how we can tell in practice when a real function is continuous. As we showed at the begining, for the usual topology on the real line, our notion of continuity agrees with the calculus idea of continuity. Next we quickly recall some aspects from calculus.

12.1.1 Aside: Paths and continuity properties from analysis

Note that the real line is fundamental to all path constructions. We will need many of its properties. Functions from $\mathbb{R}$ to $\mathbb{R}$, or from subsets of $\mathbb{R}$, are often built using the field property. For example $x \mapsto \pi x$; $x \mapsto x^2$; $x \mapsto e + x$; $x \mapsto |x|$; $x \mapsto x^{-1}$. What can we say about continuity in this specific setting? We showed earlier that here continuity is the same as 'calculus continuity'. So we can check continuity with ϵ, δ constructions:

$f : \mathbb{R} \to \mathbb{R}$ is continuous iff

for $\epsilon > 0$

(a real parameter that is 'positive' in the ordered field construction)

there exists $\delta > 0$

(another one!)

such that $|x - y| < \delta$

(heavy repeated use of the ordered field properties required to state this part of the axiom)

implies $|f(x) - f(y)| < \epsilon$

(more!).

(**12.1.1**) Example. Consider $f(x) = 3x + 4$. Our target tolerance is $\epsilon > 0$. To achieve this tolerance try $\delta = \epsilon/3$. That is, we impose $|x-y| < \epsilon/3$. (Note that this is at once familiar if you have done analysis; and not a manipulation that has any meaning in an arbitrary topological space!) We have

$$|f(x) - f(y)| = |3x + 4 - 3y - 4| = 3|x - y| < \epsilon$$

as required.

(**12.1.2**) Example. If $f, g : \mathbb{R} \to \mathbb{R}$ are continuous, then the 'pointwise composition' fg given by $fg(x) = f(x)g(x)$ is continuous. Also the pointwise sum given by $(f + g)(x) = f(x) + g(x)$.

Since we already know that the constant and identity functions are always continuous, we deduce that polynomial functions are continuous.

12.1.2 Homotopy constructions and properties

Example 12.2. Let $x, y \in \mathbb{R}^n$ and let $\alpha, \beta : [0,1] \to \mathbb{R}^n$ be any two paths from x to y. Then we claim that the following gives a function, and this function is a homotopy from α to β:

$$F(s,t) = (1-t)\alpha(s) + t\beta(s).$$

To see that the expression gives a function we just note closure of the appropriate vector arithmetic on the codomain (NB we will have important examples later where this kind of construction fails). For continuity we may proceed as follows. Clearly $F(s,0) = \alpha(s)$ and $F(s,1) = \beta(s)$. The functions $(s,t) \mapsto t$, $(s,t) \mapsto \alpha(s)$ and $(s,t) \mapsto \beta(s)$ are continuous by Proposition 7.34, and so F is continuous by the 'calculus of limits' as recalled in §12.1.1.

This homotopy is called the **straight line homotopy** from α to β.

Example 12.3. Consider the formulae

$$\alpha(s) = (\cos(k\pi s), \sin(k\pi s))$$
$$\beta(s) = (\cos(k\pi s), -\sin(k\pi s))$$

with $k = 2$. They describe paths from $(1,0)$ to $(1,0)$ both in $\mathbb{R}^2$ and in $\mathbb{R}^2 \setminus \{(0,0)\}$.
In fact they are path homotopic in the former space, but not in the latter.
It is an interesting question how to prove these statements. (But intuitively they may be 'clear'.)
In $\mathbb{R}^2$ both paths are homotopic to the constant path given by $\gamma(s) = (1,0)$. This can be seen by considering appropriate straight line homotopies.
Note that these constructions do not work in $\mathbb{R}^2 \setminus \{(0,0)\}$, since $(0,0)$ is an image point at some value of t. (Exercise: work out which one.)
But... How do we know that there is not an alternative construction of a homotopy in this case? For this, see later.

Lemma 12.4. *Let (X,τ) be any space, and $x, y \in X$. The relation of path homotopy is an equivalence relation on the set $\mathsf{P}_{(X,\tau)}(x,y)$ of paths in X from x to y (as in (3.2.20)).*

Proof. We need to check that the relation is reflexive, symmetric, and transitive. Let X be a topological space and let α, β, γ be paths in X from x to y.

Reflexive: we must show $\alpha \simeq \alpha$. $F(s,t) = \alpha(s)$ is a homotopy from α to α so $\alpha \simeq \alpha$.

Symmetric: we must show $\alpha \simeq \beta \Rightarrow \beta \simeq \alpha$. If $\alpha \simeq \beta$ there is a path homotopy, F say, from α to β. Then $G : [0,1] \times [0,1] \to X$ given by

$$G(s,t) = F(s, 1-t)$$

is a path homotopy from β to α, so $\beta \simeq \alpha$.

Transitive: suppose $\alpha \simeq \beta$ and $\beta \simeq \gamma$. Then there are path homotopies F from α to β and G from β to γ. Consider the function $H : [0,1] \times [0,1] \to X$ given by

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

for all $s \in [0,1]$. We claim the function H is a path homotopy from α to γ. The restrictions of H to the closed subsets $[0,1] \times [0, \frac{1}{2}]$ and $[0,1] \times [\frac{1}{2}, 1]$ of $[0,1] \times [0,1]$ are both continuous. On the overlap region $[0,1] \times \{\frac{1}{2}\}$ we have that

$$F(s,2t) = \beta(s) = G(s, 2t-1).$$

Therefore H is continuous, by the glue lemma (Lemma 7.35). Clearly $H(s,0) = \alpha(s)$, $H(s,1) = \gamma(s)$, $H(0,t) = x$ and $H(1,t) = y$ so H is a path homotopy from α to γ and $\alpha \simeq \gamma$. □

12.1.3 Path 'algebra'

Definition 12.5. Let (X, τ) be a topological space. Let $x, y, z \in X$. Let α be a path from x to y and let β be a path from y to z.
(1) The **meld** of α and β is the path $\alpha * \beta : [0,1] \to X$ from x to z defined by

$$\alpha * \beta(s) = \begin{cases} \alpha(2s) & s \in [0, \frac{1}{2}] \\ \beta(2s-1) & s \in [\frac{1}{2}, 1]. \end{cases}$$

(2) The **reverse** of α is the path $\overline{\alpha} : [0,1] \to X$ from y to x defined by

$$\overline{\alpha}(s) = \alpha(1-s).$$

(3) For any point $x \in X$, the **constant path at** x is the path $e_x : [0,1] \to X$ from x to x defined by

$$e_x(s) = x \quad \forall s \in [0,1].$$

Exercise: check that $\alpha * \beta$, $\overline{\alpha}$ and e_x are paths (you will need the glue lemma 7.35 for $\alpha * \beta$).

Theorem 12.6. *The operations $*$ and $\alpha \mapsto \overline{\alpha}$ are well-defined on path-homotopy equivalence classes. Moreover they have the following properties:*

1. *Associativity: $[\alpha] * ([\beta] * [\gamma]) = ([\alpha] * [\beta]) * [\gamma]$ for all paths α, β, γ such that $\alpha(1) = \beta(0)$ and $\beta(1) = \gamma(0)$.*
2. *Identity: if α is a path from x to y then $[e_x] * [\alpha] = [\alpha]$ and $[\alpha] * [e_y] = [\alpha]$.*
3. *Inverse: if α is a path from x to y then $[\alpha] * [\overline{\alpha}] = [e_x]$ and $[\overline{\alpha}] * [\alpha] = [e_y]$.*

Proof. First we show that the operations are well-defined. Suppose that $\alpha, \alpha', \beta, \beta'$ are paths such that $\alpha(1) = \beta(0)$, $\alpha \simeq \alpha'$ and $\beta \simeq \beta'$. We need to check that $[\alpha * \beta] = [\alpha' * \beta']$, i.e. that $\alpha * \beta \simeq \alpha' * \beta'$. Let F be a path homotopy from α to α' and let G be a path homotopy from β to β'. We claim that the following function H is a path homotopy from $\alpha * \beta$ to $\alpha' * \beta'$:

$$H(s,t) = \begin{cases} F(2s, t) & (s,t) \in [0, \frac{1}{2}] \times [0,1] \\ G(2s-1, t) & (s,t) \in [\frac{1}{2}, 1] \times [0,1]. \end{cases}$$

This function is continuous by the glue lemma, because the sets $[0, \frac{1}{2}] \times [0,1]$ and $[\frac{1}{2}, 1] \times [0,1]$ are closed, and $F(2s,t) = G(2s-1,t)$ on their intersection. It is straightforward to check that $H(0,t) = \alpha(0)$ and $H(1,t) = \beta(1)$ $\forall t \in [0,1]$, and that $H(s,0) = \alpha * \beta(s)$ and $H(s,1) = \alpha' * \beta'(s)$, $\forall s \in [0,1]$.

We also need to show that $\overline{\alpha} \simeq \overline{\alpha}'$. It is straightforward to check that the function $(s,t) \mapsto F(1-s,t)$ is a path homotopy from $\overline{\alpha}$ to $\overline{\alpha}'$.

Next we prove that constant paths are identity elements. Let α be a path from x to y. We need to show that

$$e_x * \alpha(s) = \begin{cases} x & s \in [0, \frac{1}{2}] \\ \alpha(2s-1) & s \in [\frac{1}{2}, 1] \end{cases}$$

is path homotopic to α. We make an ansatz for a path homotopy of the form:

$$J(s,t) = \begin{cases} x & s \in [0, \frac{t}{2}] \\ \alpha(As + B) & s \in [\frac{t}{2}, 1] \end{cases}.$$

Notice that the ranges of the intervals depend on t! In order to satisfy the condition $J(1,t) = y$ we require that $\alpha(A+B) = y = \alpha(1)$ and hence that

$$A + B = 1.$$

In order to make this function J well-defined and continuous we require that $\alpha(At/2+B) = x$ and hence that

$$A\frac{t}{2} + B = 0.$$

These simultaneous equations are solved by $A = 2/(2-t)$, $B = -t/(2-t)$. We leave it as an exercise to check using the glue lemma J is continuous and moreover that J is a path homotopy from $e_x * \alpha$ to α. We also leave it as an exercise to find a path homotopy form $\alpha * e_y$ to α.

Next we prove that reverses are inverses. Let α be a path from x to y. We need to show that

$$\overline{\alpha} * \alpha(s) = \begin{cases} \alpha(1-2s) & s \in [0, \frac{1}{2}] \\ \alpha(2s-1) & s \in [\frac{1}{2}, 1] \end{cases}$$

is homotopic to the constant path e_y. We make an ansatz for a homotopy of the form

$$K(s,t) = \begin{cases} \alpha(As+B) & s \in [0, \frac{1}{2}] \\ \alpha(Cs+D) & s \in [\frac{1}{2}, 1] \end{cases}.$$

In order to satisfy the conditions $K(0,t) = y = K(1,t)$ we impose that

$$B = 1, \quad C + D = 1.$$

In order to make this a homotopy from $\overline{\alpha} * \alpha$ to e_y we also require that $K(1/2,t) = \alpha(t)$ (so that $K(1/2,0) = \overline{\alpha} * \alpha(1/2)$, $K(1/2,1) = e_y(1/2)$). This condition is met if

$$A/2 + B = t, \quad C/2 + D = t.$$

The above simultaneous equations are solved by

$$A = 2t-2,\ B = 1,\ C = 2-2t,\ D = 2t-1.$$

It is left as an exercise to check using the glue lemma that K is continuous and show that it satisfies the definition of a path homotopy. It is also left as an exercise to check that $\alpha * \overline{\alpha} \simeq e_x$.

Finally we prove associativity. We need to find a path homotopy from

$$\alpha * (\beta * \gamma)(s) = \begin{cases} \alpha(2s) & s \in [0, \frac{1}{2}] \\ \beta(4s-2) & s \in [\frac{1}{2}, \frac{3}{4}] \\ \gamma(4s-3) & s \in [\frac{3}{4}, 1] \end{cases}$$

to

$$(\alpha * \beta) * \gamma(s) = \begin{cases} \alpha(4s) & s \in [0, \frac{1}{4}] \\ \beta(4s-3) & s \in [\frac{1}{4}, \frac{1}{2}] \\ \gamma(2s-1) & s \in [\frac{1}{2}, 1]. \end{cases}$$

We seek a path homotopy of the form,

$$I(s,t) = \begin{cases} \alpha(As+B) & 0 \leq s \leq \frac{2-t}{4} \\ \beta(Cs+D) & \frac{2-t}{4} \leq s \leq \frac{3-t}{4} \\ \gamma(Es+F) & \frac{3-t}{4} \leq s \leq 1. \end{cases}$$

Notice how the endpoints of the intervals depend on t, such that when $t = 0, 1$ they agree with those for $\alpha * (\beta * \gamma)$ and $(\alpha * \beta) * \gamma$. We choose the constants A, B, C, D, E, F so that I satisfies the definition of a path homotopy. The conditions $I(0, t) = \alpha(0)$, $I(1, t) = \gamma(1)$ are satisfied if

$$B = 0, \quad E + F = 1.$$

Looking at the endpoints of the intervals, the function I is well-defined and continuous if

$$A\frac{2-t}{4} + B = 1, \quad C\frac{2-t}{4} + D = 0, \quad C\frac{3-t}{4} + D = 1, \quad E\frac{3-t}{4} + F = 0.$$

These simultaneous equations are solved by

$$A = \frac{4}{2-t},\ B = 0,\ C = 4,\ D = t - 2,\ E = \frac{4}{t+1}, F = \frac{t-3}{t+1}.$$

It is left as an exercise to check using the glue lemma that I is continuous and show that it satisfies the definition of a path homotopy. □

12.2 The fundamental group

The fundamental group of a topological space is a group defined using equivalence classes of paths in the space.

Definition 12.7. /Proposition. Let (X, τ) be a topological space and let $x \in X$. A path from x to x is called a **loop based at** x. The set of path homotopy equivalence classes of loops based at x together the operation $*$ is a group, called the **fundamental group of** X **relative to the base point** x. It is denoted $\pi_1(X, x)$:

$$\pi_1(X, x) = \mathsf{P}_{(X,\tau)}(x, x)/\simeq$$

(Caveat: note that our $*$ here is slightly different from the one we used on $\mathsf{P}'_{(X,\tau)}$ from (3.2.20). One of the effects of $\simeq$ is to effectively wash these kinds of differences out.)

Proof. This follows from the Theorem 12.6 proved above. □

Remark 12.8. The fundamental group is also known as the **first homotopy group**. There is also an n-th homotopy group $\pi_n(X, x)$ defined using maps from the n-dimensional disc to X. Fully understanding these groups remains an important open problem.

Example 12.9. Proposition. For any $x \in \mathbb{R}^n$, $\pi_1(\mathbb{R}^n, x)$ is trivial (i.e. it has only one element).

Proof. This is because every loop based at $x \in \mathbb{R}^n$ is path homotopic to the constant path at x (via a straight line homotopy as described in 12.2 above). □

12.2.1 The fundamental group $\pi_1(X)$ of a path-connected space

The fundamental group $\pi_1(X, x)$ involves a choice of base point. The next Theorem explains how the group depends on the base point:

Definition 12.10. Let G, H be two groups and let $\phi : G \to H$. Then ϕ is called a **homomorphism** if

$$\phi(g_1 \cdot g_2) = \phi(g_1) \cdot \phi(g_2) \quad \forall g_1, g_2 \in G.$$

ϕ is called an **isomorphism** if in addition it is a bijection; in this case G and H are said to be **isomorphic** and we write $G \cong H$.

Theorem 12.11. *Let X be a topological space, let x, y be points in X, and let α be a path from x to y. Then the map*

$$\begin{array}{rccc} \hat{\alpha}: & \pi_1(X,x) & \to & \pi_1(X,y) \\ & [\beta] & \mapsto & [\overline{\alpha}] * [\beta] * [\alpha] \end{array}$$

is an isomorphism of groups.

Proof. First we show that $\hat{\alpha}$ is a bijection. We do so by showing that $\hat{\alpha}$ has an inverse. The inverse is the map

$$\begin{array}{rccc} \hat{\overline{\alpha}}: & \pi_1(X,y) & \to & \pi_1(X,x) \\ & [\beta] & \mapsto & [\alpha] * [\beta] * [\overline{\alpha}]. \end{array}$$

That this is an inverse follows directly from the properties proved in Theorem 12.6, indeed, for any loop β based at x,

$$\begin{aligned} \hat{\overline{\alpha}} \circ \hat{\alpha}([\beta]) &= [\alpha] * ([\overline{\alpha}] * [\beta] * [\alpha]) * [\overline{\alpha}] \\ &= ([\alpha] * [\overline{\alpha}]) * [\beta] * ([\alpha] * [\overline{\alpha}]) \\ &= [e_x] * [\beta] * [e_x] \\ &= [\beta]. \end{aligned}$$

So $\hat{\overline{\alpha}} \circ \hat{\alpha} = \mathrm{id}_{\pi_1(X,x)}$. By a similar calculation, $\hat{\alpha} \circ \hat{\overline{\alpha}} = \mathrm{id}_{\pi_1(X,y)}$.

Next we show that $\hat{\alpha}$ is a homomorphism, i.e. that $\hat{\alpha}([\beta] * [\gamma]) = \hat{\alpha}([\beta]) * \hat{\alpha}([\gamma])$ for all loops β, γ based at x. By Theorem 12.6,

$$\begin{aligned} \hat{\alpha}([\beta]) * \hat{\alpha}([\gamma]) &= ([\overline{\alpha}] * [\beta] * [\alpha]) * ([\overline{\alpha}] * [\gamma] * [\alpha]) \\ &= [\overline{\alpha}] * [\beta] * ([\alpha] * [\overline{\alpha}]) * [\gamma] * [\alpha] \\ &= [\overline{\alpha}] * [\beta] * [\gamma] * [\alpha] \\ &= \hat{\alpha}([\beta] * [\gamma]) \end{aligned}$$

□

Remark 12.12. Note that different paths α from x to y may induce different isomorphisms. Also if there is no path from x to y, then there is no reason to expect that $\pi_1(X,x) \cong \pi_1(X,y)$. For this reason, we tend to study fundamental groups of path connected spaces only. Of course the above theorem tells us that if X is path connected then the fundamental group is independent of the choice of base point, and in this case we may well suppress the base point and write simply $\pi_1(X)$ to denote the fundamental group.

The fundamental group provides a means to test whether two paths are path homotopic:

Lemma 12.13. *Let X be a topological space, let $x, y \in X$ and let α, β be two paths from x to y. Then $\alpha \simeq \beta$ if and only if $[\alpha] * [\overline{\beta}]$ is equal to the identity element in $\pi_1(X,x)$.*

Proof. First suppose that $[\alpha] * [\overline{\beta}] = [e_x]$. Then by Theorem 12.6,

$$\begin{aligned} [\alpha] &= [\alpha] * [e_y] \\ &= [\alpha] * [\overline{\beta}] * [\beta] \\ &= [e_x] * [\beta] \\ &= [\beta]. \end{aligned}$$

Conversely, if $[\alpha] = [\beta]$ then

$$[\alpha] * [\overline{\beta}] = [\beta] * [\overline{\beta}] = [e_x].$$

□

Definition 12.14. A topological space X is called **simply connected** if it is path connected and the fundamental group of X is trivial (i.e. it is the group with only one element).

The latter property is sometimes denoted $\pi_1(X) \cong 0$.

Lemma 12.15. *In a simply connected topological space, any two paths having the same initial and final points are path homotopic.*

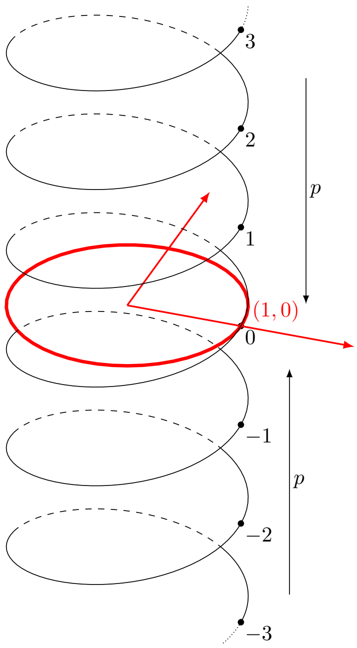
Proof. Direct consequence of the previous Lemma. □

12.3 The fundamental group of the circle

We have learnt a bit about the general properties of the fundamental group; and found some spaces with trivial fundamental group. Now we calculate the (non-trivial) fundamental group of the circle.

Fix the base point $b = (1, 0)$ in the circle $S^1 = \{x \in \mathbb{R}^2 : |x|^2 = 1\}$, and let α be a loop based at b. Roughly speaking, the winding number $N(\alpha)$ of α is the number of times the path goes around the circle anticlockwise, minus the number of times it goes around clockwise.

In order to work rigorously with this number we first need to see $\mathbb{R}$ as the **universal cover** of S^1.



12.3.1 Covering spaces

Have in mind here the homeomorphic copy of the real line arranged as a spiral in the picture, and the (continuous) geometric projection into a circle in the xy-plane.

(**12.3.1**) Let $f : X \to Y$ be a continuous function. Suppose that for each $y \in Y$ there is a neighbourhood V_y; and $f^{-1}(V_y)$ can be decomposed as a family of disjoint open subsets $\{U_a\}_a$ of X, such that the restriction to each part U_a is a homeomorphism from U_a to V_y. Then X is a **covering space** of Y.

In the picture, the spiral form of $\mathbb{R}$ is a covering space for S^1. Each little interval of S^1 has pre-image a tower of little intervals above and below it.

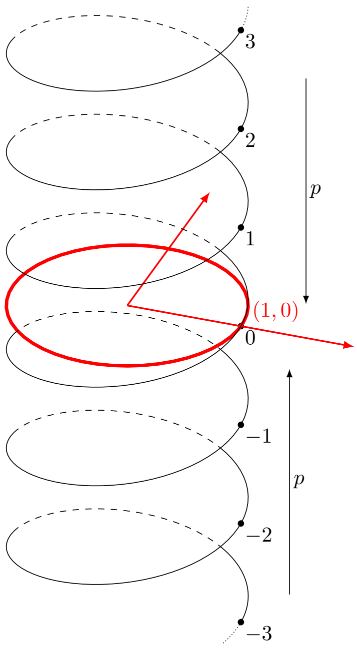
(**12.3.2**) A space Y can have many covering spaces. However it turns out that if covering space X of Y is simply connected as such, then it is unique up to homeomorphism. Such a covering space is called a **universal covering space** of Y.

Since $\mathbb{R}$ is simply connected, our spiral picture is a universal covering space for S^1. Universal covers are useful in computing fundamental groups, as we will see next.

12.3.2 The case of the circle in detail

Let $p : \mathbb{R} \to S^1$ be the map $p : u \mapsto (\cos(2\pi u), \sin(2\pi u))$. First of all notice that $p(u) = p(u+m)$ for all $u \in \mathbb{R}, m \in \mathbb{Z}$. Also, when we restrict p to any open interval of length one, i.e. in the form (u_0, u_0+1), we have that $p|_{(u_0,u_0+1)} : (u_0, u_0+1) \to S^1 \backslash \{p(u_0)\}$ is a homeomorphism onto its image: since p is a continuous bijection here with inverse $p|^{-1}_{(u_0,u_0+1)}(x) = u_0 + \frac{\theta(x,p(u_0))}{2\pi}$ where $\theta(x, p(u_0)) \in (0, 2\pi)$ is the anticlockwise angle between x and $p(u_0)$.[1]

You should be imagining $\mathbb{R}$ (arranged as the black coil) sitting over S^1 (in red) and winding around so that, for $x \in S^1$, $p^{-1}(\{x\}) \subset \mathbb{R}$ is the collection of points lying directly above x in the picture. We have drawn on $p^{-1}(\{(1,0)\}) = \mathbb{Z}$.



A **lift** of α is defined to be a path $\tilde{\alpha}$ in $\mathbb{R}$ such that $p \circ \tilde{\alpha} = \alpha$. The **winding number** of α is defined to be

$$N(\alpha) = \tilde{\alpha}(1) - \tilde{\alpha}(0). \qquad (*)$$

This number $N(\alpha)$ is in fact an integer, because $\tilde{\alpha}(0), \tilde{\alpha}(1) \in p^{-1}(\{b\})$ and $p^{-1}(\{b\}) = \mathbb{Z}$.

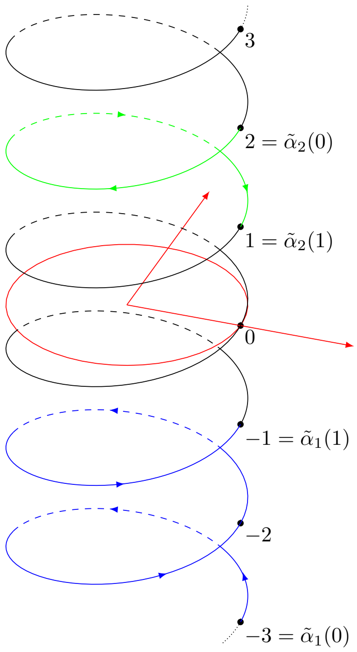
Of course there are infinitely many lifts: if $\tilde{\alpha}$ is a lift of α then $\hat{\alpha}(s) = \tilde{\alpha}(s) + m$ is another lift for any $m \in \mathbb{Z}$. But the point is that $N(\alpha)$ is independent of the lift that we use.

For example if $\alpha_1(s) = (\cos(2\pi \times 2s), \sin(2\pi \times 2s))$ then $\tilde{\alpha}_1 = -3 + 2s$ is a lift of α_1 (in blue below) and $N(\alpha_1) = 2$. Similarly if $\alpha_2(s) = (\cos(2\pi s), -\sin(2\pi s))$ then $\tilde{\alpha}_2(s) = 2 - s$ is a lift of α_2 (in green) and $N(\alpha_2) = -1$:

[1] If you want an expression for θ, first write the anticlockwise rotation of $p(u_0)$ by $\pi/2$ as $p(u_0)^\perp$. Then we have

$$\theta(x, p(u_0)) = \begin{cases} \arccos p(u_0) \cdot x & \text{if } p(u_0)^\perp \cdot x \geq 0 \\ 2\pi - \arccos p(u_0) \cdot x & \text{if } p(u_0)^\perp \cdot x \leq 0, \end{cases}$$

where $\arccos \in [0, \pi]$. You can check that this is continuous if you like.



Thankfully, we only need to specify the starting point of a lift in order that it is unique.

Lemma 12.16. *Let $\alpha : [0,1] \to S^1$ be any loop based at $b = (1,0)$ (so $\alpha(0) = \alpha(1) = (1,0)$). Then for each $q \in p^{-1}(\{b\}) = \mathbb{Z}$, there exits a unique lift $\tilde{\alpha} : [0,1] \to \mathbb{R}$ satisfying $\tilde{\alpha}(0) = q$ and $p \circ \tilde{\alpha} = \alpha$.*

See the end of this section for a proof (but this is non-examinable material).

Notice that this Lemma immediately tells us that N is well-defined. Since if $\tilde{\alpha}$ is a lift of α starting at $q_1 \in \mathbb{Z}$ (which is unique), and if $\hat{\alpha}$ is another lift starting at $q_2 \in \mathbb{Z}$, then setting $\tilde{\alpha}_1(s) = \hat{\alpha}(s) + q_1 - q_2$, we must have $\tilde{\alpha}_1(s) \equiv \tilde{\alpha}(s)$. Surely $\tilde{\alpha}_1$ defined in this way is another lift starting at q_1, but since lifts are unique when we fix the initial point, then $\tilde{\alpha} = \tilde{\alpha}_1$.

Thus

$$N(\alpha) = \tilde{\alpha}(1) - \tilde{\alpha}(0) = \hat{\alpha}(1) - \hat{\alpha}(0).$$

The next Lemma will guarantee that not only is N well-defined, but it only depends upon the homotopy class of α, i.e. $N(\alpha) = N([\alpha])$.

Lemma 12.17. *Suppose that $F : [0,1] \times [0,1] \to S^1$ is continuous and satisfies $F(0,t) = F(1,t) = b = (1,0)$ for all $t \in [0,1]$. Then for any $q \in p^{-1}(\{b\}) = \mathbb{Z}$, there exits a unique continuous map $\tilde{F} : [0,1] \times [0,1] \to \mathbb{R}$ satisfying $\tilde{F}(0,t) = q$ for all t and $p \circ \tilde{F} = F$.*

Once again, the proof is non-examinable but you can find it at the end of this section. A consequence of this result is crucial to us: it tells us that $N(\alpha)$ only depends on the homotopy class $[\alpha]$ of α.

Specifically, if $[\alpha] = [\beta] \in \pi_1(S^1, b)$ and F is a path homotopy between them, we can let $\tilde{F} : [0,1] \times [0,1] \to \mathbb{R}$ be the lift of F described by the previous Lemma for some fixed $q \in \mathbb{Z}$. Notice that, by definition $\tilde{F}(s,0) = \tilde{\alpha}(s)$ is a lift of α and $\tilde{F}(s,1) = \tilde{\beta}(s)$ is a lift of β. Furthermore, $p \circ \tilde{F}(1,t) = b = (1,0)$ for all t, so $\tilde{F}(1,t) \in \mathbb{Z}$ for all t. This implies that $\tilde{F}(1,t) = q_0 \in \mathbb{Z}$ is fixed for all t and in particular $\tilde{\alpha}(1) = \tilde{F}(1,0) = \tilde{F}(1,1) = \tilde{\beta}(1) = q_0$.

Of course we already know that $\tilde{F}(0,t) = q$ for all t which implies $\tilde{\alpha}(0) = \tilde{F}(0,0) = \tilde{F}(0,1) = \tilde{\beta}(0) = q$.

We have shown that $[\alpha] = [\beta] \implies N(\alpha) = N(\beta)$.

Theorem 12.18. *The map*

$$\phi : \pi_1(S^1, b) \to (\mathbb{Z}, +),$$
$$\phi : [\alpha] \mapsto N(\alpha)$$

is an isomorphism of groups.

Proof. Lemmata 12.16 and 12.17 guarantee that ϕ is well-defined (which is an important step). It remains to check that it is a bijective group homomorphism (i.e. an isomorphism).
ϕ is surjective: Let $n \in \mathbb{Z}$; we must find a loop α such that $\phi(\alpha) = n$. A suitable loop is given by

$$\alpha : t \mapsto (\cos(2\pi nt), \sin(2\pi nt)).$$

A lift of this map is given by $\tilde{\alpha} : [0,1] \to \mathbb{R}$, $\tilde{\alpha} : t \mapsto nt$; clearly $\tilde{\alpha}(1) - \tilde{\alpha}(0) = n$, as required.

ϕ is injective: Let α, β be two loops in S^1 such that $N(\alpha) = N(\beta)$; we must show that $[\alpha] = [\beta]$. Let $\tilde{\alpha}, \tilde{\beta} : [0,1] \to \mathbb{R}$ be lifts of these paths, both starting at the same point q. Since $N(\alpha) = N(\beta)$ they also have the same final point. Therefore there exists a path homotopy $\hat{F}$ from $\tilde{\alpha}$ to $\tilde{\beta}$ (fixing end points). The map $F = p \circ \hat{F} : [0,1] \times [0,1] \to S^1$ is a path homotopy from α to β, so $[\alpha] = [\beta]$.

ϕ is a homomorphism: Let α, β be any two loops based at b; we must show that

$$N(\alpha * \beta) = N(\alpha) + N(\beta).$$

Let $\tilde{\alpha}, \tilde{\beta} : [0,1] \to \mathbb{R}$ be lifts of these paths, but this time suppose that $\tilde{\alpha}$ is a lift starting at q and $\tilde{\beta}$ a lift starting at $\tilde{\alpha}(1)$, i.e. $\tilde{\beta}(0) = \tilde{\alpha}(1)$. The join $\tilde{\alpha} * \tilde{\beta}$ is a lift of $\alpha * \beta$, because

$$p \circ \tilde{\alpha} * \tilde{\beta} : s \mapsto \begin{cases} p \circ \tilde{\alpha}(2s) & 0 \le s \le \frac{1}{2} \\ p \circ \tilde{\beta}(2s-1) & \frac{1}{2} \le s \le 1 \end{cases} = \begin{cases} \alpha(2s) & 0 \le s \le \frac{1}{2} \\ \beta(2s-1) & \frac{1}{2} \le s \le 1 \end{cases}.$$

So

$$N(\alpha * \beta) = \tilde{\alpha} * \tilde{\beta}(1) - \tilde{\alpha} * \tilde{\beta}(0) = \tilde{\beta}(1) - \tilde{\beta}(0) + \tilde{\alpha}(1) - \tilde{\alpha}(0) = N(\beta) + N(\alpha)$$

as required. □

Proof of Lemma 12.16. Here we will need to use the notion of a Lebesgue number that we met at the end of the previous chapter. Consider $U = S^1 \backslash \{b\}$ and $V = S^1 \backslash \{-b\}$ which is an open cover of S^1. Furthermore, $p^{-1}(U) = \mathbb{R} \backslash \mathbb{Z}$ and $p^{-1}(V) = \mathbb{R} \backslash (\mathbb{Z} + 1/2)$.

Given any $m \in \mathbb{Z}$ and an interval of the form $(m, m+1)$ then $p_m = p|_{(m,m+1)} : (m, m+1) \to S^1 \backslash \{b\}$ is a homeomorphism onto its image. Similarly $p_{m/2} = p|_{(m-1/2,m+1/2)} : (m - 1/2, m + 1/2) \to S^1 \backslash \{-b\}$ is a homeomorphism onto its image.

Since α is continuous then $\alpha^{-1}(U)$ and $\alpha^{-1}(V)$ form an open cover of $[0,1]$ so let $\varepsilon > 0$ be the Lebesgue number of this cover (which exists because $[0,1]$ is compact cf. Proposition 11.69). In particular we can find $0 = t_0 < t_0 < \cdots < t_{n-1} < t_n = 1$ so that $\alpha([t_i, t_{i+1}]) \subset U$ **or** $\alpha([t_i, t_{i+1}]) \subset V$ for all $i = 0, \ldots, n-1$ (e.g. as long as $t_{i+1} - t_i < \varepsilon/2$ this is guaranteed by the definition of a Lebesgue number). Now by fixing $q \in \mathbb{Z}$ we define $\tilde{\alpha} : [t_0, t_1] \to \mathbb{R}$ by $\tilde{\alpha} = p_{q/2}^{-1} \circ \alpha$ which is well defined since on this interval $\alpha([t_0, t_1]) \subset V$ (it cannot be a subset of U since $\alpha(0) \notin U$). Notice that $\tilde{\alpha}([t_0, t_1]) \subset (q - 1/2, q + 1/2)$ by definition.

If $\alpha([t_1, t_2]) \subset V$ as well then simply extend $\tilde{\alpha}$ in exactly the same fashion to $[t_0, t_2]$. : i.e. set $\tilde{\alpha} = p_{q/2}^{-1} \circ \alpha$ on $[t_0, t_2]$ which is well defined since $\alpha([t_0, t_2]) \subset V$.

If, on the other hand, if $\alpha([t_1, t_2]) \not\subset V$ then we must have $\alpha([t_1, t_2]) \subset U$. In particular $\tilde{\alpha}(t_1) \neq q$. We first ask: is $\tilde{\alpha}(t_1) \in (q, q + 1/2)$ or $\tilde{\alpha}(t_1) \in (q - 1/2, q)$?[2] It must be one or the other, so if it's the first we define $\tilde{\alpha} : [t_1, t_2] = p_q^{-1} \circ \alpha$ and if it's the second we define $\tilde{\alpha} : [t_1, t_2] = p_{q-1}^{-1} \circ \alpha$.

[2] i.e. did the curve start winding anticlockwise, or clockwise?

Notice that in either case $\tilde{\alpha} : [t_0, t_2] \to \mathbb{R}$ is well-defined and continuous (e.g. by the glue Lemma): it is also uniquely determined in each interval $[t_0, t_1]$ and $[t_1, t_2]$[3] since we have sufficiently restricted p, $p|$ so that $p|^{-1}$ has a specified target and is a homeomorphism onto that target.

Now, we can inductively keep extending the definition of $\tilde{\alpha}$ in this way and obtain a path $\tilde{\alpha} : [0, 1] \to \mathbb{R}$ with $\tilde{\alpha}(0) = q$. Furthermore, $\rho \circ \tilde{\alpha} = \alpha$ by construction, and since it was uniquely determined on each interval $[t_i, t_{i+1}]$ it must be uniquely determined by the initial choice of $q \in \mathbb{Z}$.

□

Proof of Lemma 12.17. This is really the same proof as the previous Lemma 12.16 in spirit, so I will only sketch the details. Once again, let $\varepsilon > 0$ be a Lebesgue number for the cover $\{F^{-1}(U), F^{-1}(V)\}$ of $[0, 1] \times [0, 1]$ (guaranteed to exists since $[0, 1] \times [0, 1]$ is compact cf. Proposition 11.69).

In this case, we can find $0 = t_0 < t_1 < \cdots < t_n = 1$ so that each sub-cube $[t_i, t_{i+1}] \times [t_j, t_{j+1}] \subset F^{-1}(U)$ or $[t_i, t_{i+1}] \times [t_j, t_{j+1}] \subset F^{-1}(V)$ for all $i, j \in \{0, \ldots, n\}$.

Again, we can construct $\tilde{F}$ uniquely on each sub-cube, under the starting condition that $\tilde{F}(0, 0) = q$: start with $[t_0, t_1] \times [t_0, t_1]$ and notice that $F([t_0, t_1] \times [t_0, t_1]) \subset V$. Then move onto $[t_1, t_2] \times [t_0, t_1]$ and notice that $\tilde{F}(\{t_1\} \times [t_0, t_1])$ has already been determined by the previous step. Again we can uniquely extend $\tilde{F}$ to be defined in $[t_0, t_2] \times [t_0, t_1]$.

Inductively we keep going along the bottom row of sub-cubes, exactly as before, until $\tilde{F}$ is defined on $[0, 1] \times [t_0, t_1]$. Now we start at $[t_0, t_1] \times [t_1, t_2]$, notice that $\tilde{F}$ has already been uniquely determined by $\tilde{F}([t_0, t_1] \times \{t_1\})$ so we can uniquely extend it into this cube. Again work horizontally so that we end up having uniquely determined $\tilde{F}$ on $[0, 1] \times [t_0, t_2]$. At this point we start on $[t_0, t_1] \times [t_2, t_3]$... etc etc.

At each stage of the extension, at least one edge of the next cube (and always a connected component of it) has been uniquely determined and we can continue extending the definition of $\tilde{F}$ in this way until we have exhausted all sub-cubes. □

[3] you might we wondering what happens if $\alpha([t_i, t_{i+1}]) \subset U$ **and** $\alpha([t_i, t_{i+1}]) \subset V$: in this case we must have $\alpha([t_i, t_{i+1}])$ (since it is connected) is contained in one of the connected components of $U \cap V = S^1 \backslash \{\pm b\}$: thus for all $r \in \mathbb{Z}$ we have $p_r^{-1} = p_{r/2}^{-1}$ and $p_{r-1}^{-1} = p_{r/2}^{-1}$ in either of these regions

12.4 Deformation retracts

In this section we will show that π_1 is a topological invariant. So as usual for an invariant, if π_1 of two spaces differs then they are definitely not homeomorphic. But also as usual for us, π_1 being the same does not necessarily mean that spaces *are* homeomorphic. Here we turn this to our advantage by considering non-homeomorphism relations between spaces that 'preserve' π_1.

Some spaces have π_1 much easier to compute than others. So knowing when they must have the same π_1 allows us to work always with the easier ones.

As a very simple example, the trivial space $\{0\}$ and the space $\mathbb{I} = [0,1]$ have the same π_1. Computing π_1 for the former is completely trivial. For the latter it is less so. So if we can establish a relation this is powerful.

12.4.1 Topological invariant property of π_1 and 'functoriality'

Definition 12.19. Proposition. Let X and Y be topological spaces, and let

$$h : X \to Y$$

be a continuous function. Fix any $x \in X$. The **homomorphism induced by** h is the map

$$\begin{array}{rccc} h_* : & \pi_1(X,x) & \to & \pi_1(Y,h(x)) \\ & [\alpha] & \mapsto & [h \circ \alpha]. \end{array}$$

Proof. For this definition/Proposition to make sense, we need to confirm two things: that the map h_* is well-defined, and that it is a homomorphism.

Checking that the map is well-defined means confirming that if α, α' are two loops based at x such that $\alpha \simeq \alpha'$, then $h \circ \alpha \simeq h \circ \alpha'$. Suppose then that F is a path homotopy from α to α'. We claim that $h \circ F$ is a path homotopy from $h \circ \alpha$ to $h \circ \alpha'$. This function is continuous because it is the composition of two continuous functions, and it is straightforward to check that

$$h \circ F(s,0) = h \circ \alpha(s), \quad h \circ F(s,1) = h \circ \alpha'(s), \quad h \circ F(0,t) = h \circ F(1,t) = h(x)$$

$\forall s,t \in [0,1]$. Therefore $h \circ \alpha \simeq h \circ \alpha'$.

Checking that h_* is a homomorphism means confirming that $h_*([\alpha] * [\beta]) = h_*([\alpha]) * h_*([\beta])$ for all loops α, β based at x. This property follows from the following calculation:

$$\begin{aligned} (h \circ \alpha) * (h \circ \beta)(s) &= \begin{cases} h(\alpha(2s)) & 0 \le s \le \frac{1}{2} \\ h(\beta(2s-1)) & \frac{1}{2} \le s \le 1 \end{cases} \\ &= h(\alpha * \beta(s)). \end{aligned}$$

□

Theorem 12.20. *Inducing homomorphisms is 'functorial', that is*

(a) If X, Y, Z are topological spaces, $x \in X$, and $f : X \to Y$, $g : Y \to Z$ are continuous functions then the induced homomorphisms obey $(g \circ f)_* = g_* \circ f_*$.

(b) If X is a topological space, $x \in X$ and $\mathrm{id}_X : X \to X$ is the identity function on X, then $(\mathrm{id}_X)_*$ is the identity function on $\pi_1(X,x)$.

Proof. Exercise. □

Corollary 12.21. *If X, Y are topological spaces and $f : X \to Y$ is a homeomorphism then f_* is an isomorphism. In other words, the isomorphism class of the fundamental group is a topological invariant.*

Proof. As f^{-1} is continuous, we have that $(f^{-1})_* \circ f_* \cong \mathrm{id}_{\pi_1(X,x)}$ and $(f)_* \circ (f^{-1})_* \cong \mathrm{id}_{\pi_1(Y,f(x))}$, so $(f^{-1})_*$ is the inverse of f_* and f_* is a bijection. □

12.4.2 Deformation retracts

Definition 12.22. Let X be a topological space and let $A \subset X$ be a subspace. A is said to be a **strong deformation retract** of X if there exists a continuous map $H : X \times [0,1] \to X$ such that

$$\begin{aligned} H(x,0) &= x \quad \forall x \in X \\ H(x,1) &\in A \quad \forall x \in X \\ H(a,t) &= a \quad \forall a \in A,\, t \in [0,1]. \end{aligned}$$

The map H is called a **strong deformation retraction** in this case.

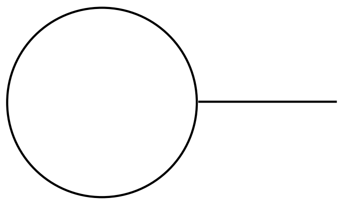
Example 12.23. $\{0\}$ is a strong deformation retract of $[0,1]$.

Let $H : [0,1] \times [0,1] \to [0,1]$ be the function $H(x,t) = (1-t)x$. Then H is continuous (by the calculus of limits). Clearly $H(x,0) = x$ and $H(x,1) \in \{0\}$ $\forall x \in [0,1]$. Also, $H(0,t) = 0$ $\forall t \in [0,1]$. So H is a strong deformation retraction from $[0,1]$ to $\{0\}$.

Example 12.24. S^{n-1} is a strong deformation retract of $\mathbb{R}^n \setminus \{0\}$.

Let $H : \mathbb{R}^n \setminus \{0\} \times [0,1] \to \mathbb{R}^n \setminus \{0\}$ be the function $H(x,t) = (1-t)x + t\frac{x}{|x|}$. Then H is continuous by the calculus of limits. Clearly $H(x,0) = x$ and $H(x,1) \in S^{n-1}$ for any $x \in \mathbb{R}^n \setminus \{0\}$. And if $x \in S^{n-1}$, $H(x,t) = x$ for any $t \in [0,1]$. So H is a strong deformation retraction.

Example 12.25. There is a strong deformation retraction from the following subset of $\mathbb{R}^2$ to the circle:



Appealing to 'geometric intuition': the strong deformation retraction shrinks the straight line down to a point (just as in the first example the interval $[0,1]$ was shrunk to a point).

Proposition 12.26. *Let A be a strong deformation retract of X and let $a \in A$. Then the inclusion $i : A \to X$ induces an isomorphism*

$$i_* : \pi_1(A,a) \to \pi_1(X,a)$$

of fundamental groups.

Proof. Let $H : X \times [0,1] \to X$ be a strong deformation retraction and let $h : X \to A$ be the map $h(x) = H(x,1)$. We claim that h_* is the inverse of i_*; it follows that i_* is an isomorphism.

To prove the claim we must show that $h_* \circ i_*$ and $i_* \circ h_*$ are the identity maps on $\pi_1(A,a)$ and $\pi_1(X,a)$ respectively. First note that $h \circ i$ is equal to the identity map on A. So by parts (a) and (b) of Theorem 12.20 $h_* \circ i_* = (h \circ i)_*$ is equal to the identity map on $\pi_1(A,a)$.

Now consider the map $i_* \circ h_*$, which by Theorem 12.20 is equal to $(i \circ h)_*$. Let α be any loop in X based at a. We claim that the function $F : [0,1] \times [0,1]$ defined by

$$F(s,t) = H(\alpha(s),t)$$

is a path homotopy from α to $i \circ h \circ \alpha$; from this it follows that $(i \circ h)_*([\alpha]) = [\alpha]$. F is continuous by Theorem 7.34. Since $\alpha(0) = \alpha(1) = a \in A$ and H is a deformation retraction, $F(0,t) = \alpha(0)$ and $F(1,t) = \alpha(1)$ $\forall t \in [0,1]$. $F(s,0) = \alpha(s)$ $\forall s \in [0,1]$ because H is a deformation retraction, and $F(s,1) = i \circ h \circ \alpha(s)$ $\forall s \in [0,1]$ by the definition of h. □

Corollary 12.27. *Let $n \geq 1$, then $\pi_1(\mathbb{R}^{n+1} \setminus \{(0)\}) \cong \pi_1(S^n)$. In particular we have*

$$\pi_1(\mathbb{R}^2 \backslash \{0\}) \cong \mathbb{Z}.$$

Remark 12.28. We will see later that $\pi_1(S^n) \cong 0$ when $n \geq 2$.

Remark 12.29. Using deformation retracts it is relatively straightforward to show that the fundamental groups of the cylinder and Möbius band are isomorphic to $\mathbb{Z}$ (they can both be deformation retracted to S^1).

12.4.3 The Van Kampen Theorem

The Seifert–Van Kampen Theorem is a powerful tool for calculating fundamental groups.

This theorem is efficiently stated using "presentations of groups".

Remark. Indeed the ideas around Van Kampen and presentations support each other to a considerable extent - which is notable, considering one of these is essentially geometric, while the other is essentially combinatorial. For further references on this connection we can suggest W Magnus, A Karrass and D Solitar, Combinatorial Group Theory, Dover.
B Knudsen, Configuration spaces in algebraic topology,

https://scholar.harvard.edu/files/knudsen/files/lecture_4.pdf

Free groups and presentations of groups

Here we return to the notation introduced in §10.3.2.

Let $A = \{a_1, a_2, \ldots, a_n\}$ be a finite non-empty set. A sequence of length m in A is an ordered m-tuple

$$w = (w_1, w_2, ..., w_m) \qquad\qquad w_i \in A$$

We sometimes write $A^{\times m}$ for the set of all sequences of length m in set A. We write

$$A^* \;=\; \bigcup_{m \in \mathbb{N}_0} A^{\times m}$$

Note that this includes the empty sequence (). That is, $A^{\times 0} = \{()\}$, for any A.

Notationally, where no ambiguity arises, we often drop all the bracket and comma 'furniture' from a sequence and simply write

$$(w_1, w_2, ..., w_m) \;\rightsquigarrow\; w_1 w_2 ... w_m$$

And then we sometimes call these sequences 'words'.

We define a binary operation on A^* by

$$(w_1, w_2, ...w_m).(v_1, v_2, ..., v_{m'}) = (w_1, w_2, ...w_m, v_1, v_2, ..., v_{m'})$$

Note that this product operation is associative and unital. But only the sequence () has an inverse.

You might like to think of the elements of A as representing some set of basic (duration-1) paths from point x to x in a space (X, τ). That is, as elements of the set $\mathsf{P}_{(X,\tau)}(x, x)$ from (3.2.20). Then $a_i a_j$ represents the combined journey made of a_i followed by a_j (an element of $\mathsf{P}'_{(X,\tau)}(x, x)$ of duration 2).
Remark: If we replace A with a graph then there is a natural generalisation of our 'word algebra' A^* above to a magmoid (that is indeed a category) as in §3.2. Words become paths on the graph (and we need an 'empty path' at each graph vertex). With this we could bring in the whole of the duration-1 path set $\mathsf{P}_{(X,\tau)}$ for the paths-in-(X, τ) analogy.

Remembering how we made 'reverses' of given paths in §3.2, then given a set A we might like to have a 'doubled-up' version

$$A^{\pm} = A \cup \{\overline{a_1}, \overline{a_2}, ..., \overline{a_n}\}$$

The over-bar $\overline{a_i}$ is really just a way of making a disjoint ('reverse') copy of the element a_i. We could perhaps have done this with a prime: a', or perhaps (*caveat emptor*) with a^{-1}.

We can summarize as follows. A **word** in $A^{\pm}$ (or in the 'group-specific' setting that follows shortly we may say just 'a word in A', automatically including the reverses) is a finite sequence of symbols a_i and $\overline{a_i}$. For example, the following are all words in the set $\{a_1, a_2\}$:

$$a_1, \quad a_1\overline{a_1}, \quad a_1a_2\overline{a_1}a_2a_2.$$

The **empty word** is the word " " in which no symbols occur – this can be denoted 1 or e (or we can go back to (), the empty sequence) in cases where this avoids confusion.
The **product** ww' of two words w and w' is defined by joining them together in order as shown. Thus

$$(a_1)(a_1\overline{a_1}) = a_1a_1\overline{a_1}$$

The **reverse** $\overline{w}$ of a word w is defined by reversing the order of the elements and replacing a_i with $\overline{a_i}$ and $\overline{a_i}$ with a_i.
Thus:

$$\overline{(a_1a_2\overline{a_1}a_2a_2)} = \overline{a_2}\,\overline{a_2}a_1\overline{a_2}\,\overline{a_1}.$$

Now let $R = \{r_1, r_2, \ldots, r_m\}$ be a finite set of words in A. We define an equivalence relation $\overset{R}{\sim}$ on the set $(A^{\pm})^*$ of all words by saying that two words w_1 and w_2 are equivalent if w_1 can be obtained from w_2 by finitely many operations of the following type:

1. inserting $a_i\overline{a_i}$ or $\overline{a_i}a_i$, where $a_i \in A$, anywhere into the word,
2. removing $a_i\overline{a_i}$ or $\overline{a_i}a_i$, where $a_i \in A$, anywhere from the word,
3. inserting r or $\overline{r}$, where $r \in R$,
4. removing r or $\overline{r}$, where $r \in R$.

It is an exercise to check that $\overset{R}{\sim}$ is an equivalence relation.

Proposition. Let G be the quotient of the set of all words by the $\overset{R}{\sim}$ equivalence relation. The set G, with the product $[a].[b] = [ab]$ and inverse $[a]^{-1} = [\overline{a}]$, is a group.

Proof. Optional exercise. □

Notationally, we tend to write just a for $[a]$ in G, and a^{-1} for $[\overline{a}]$, where no ambiguity arises.
We write

$$G = \langle a_1,\, a_2,\, \ldots,\, a_n;\; r_1,\, r_2,\, \ldots,\, r_m \rangle$$

The elements of A are called **generators** for G and the elements of R are called **relations**; together the sets A and R are called a **presentation** for G.

This quotient by an equivalence relation is the formal version of our quotient by path-homotopy in the algebra of paths. (In particular the reverse passes to the inverse again.) Thus we can connect groups described by generators and fundamental groups considered with certain generating paths (or classes of paths).

Next we will give some examples of presented groups. The example of no generators and no relations is allowed: it gives a version of the trivial group. The next simplest is 'one generator and no relations', meaning one generator and its reverse (and only the relations that are always built into the equivalence relation from the getgo).

Example 12.30. The infinite cyclic group.

Let $A = \{a\}$ be a set with one element and let $R = \emptyset$. Then any word is equivalent to one of the following words:

$$\ldots, a^{-2}, a^{-1}, e = a^0, a, a^2, \ldots.$$

For example, the word $aa^{-1}aa$ is equivalent to $aa = a^2$. Moreover, it can be shown that no two words on this list are equivalent to each other. So

$$G = \langle a; \ \rangle = \{a^m \ : \ m \in \mathbb{Z}\}.$$

This group is isomorphic to $(\mathbb{Z}, +)$. The map $\phi : \mathbb{Z} \to G$, $\phi(m) = a^m$ is an isomorphism (ϕ is clearly a bijection, and $\phi(m + n) = a^{m+n} = a^m a^n$).

Remarks:

What's in a name? Not much:

$$\langle a; \ \rangle \cong \langle b; \ \rangle$$

The fundamental group comparison to think of here is $\pi_1(S^1)$ as above. A 'generating path' is any path in the class of paths that goes exactly once around the circle (in either direction — we can't really pick a direction *extrinsically*, so we can choose one to get started). As we have already shown, there is no way to change the 'winding number' by path-homotopy. So the only kind of relation is that making the reverse as the inverse (which relation is 'built in' to $\overset{R}{\sim}$ anyway).

Example 12.31. Cyclic groups of order n.

Let $A = \{a\}$ as before and let $R = \{a^n\}$ for some $n \in \mathbb{Z}^+$. We claim that any word is equivalent to one of the following:

$$e, a, \ldots, a^{n-1}.$$

For example, $a^n \sim e$, and $a^{-1} \sim a^n a^{-1} = aa \ldots a(aa^{-1}) \sim a^{n-1}$. Moreover, it can be shown that no two words on this list are equivalent to each other. So the group G has n elements. This group is called the cyclic group of order n and is sometimes denoted $(\mathbb{Z}_n, +_n)$. This is the group $\{0, 1, \ldots, n-1\}$ under addition modulo n.

Example 12.32. Free group on n generators.

Let $A = \{a_1, a_2, \ldots, a_n\}$ and let $R = \emptyset$. The group

$$G = \langle a_1, \ldots, a_n; \rangle$$

is once again infinite. Unlike previous examples, this group is **not** abelian.

What about a fundamental group realisation of $G = \langle a_1, \ldots, a_n; \rangle$ for $n = 2$, say? Intuitively at least, we need a space X which suitably 'doubles up' the circle S^1 (recalling that $\pi_1(S^1)$ is free on one generator). Of course it must be connected, so a very simple thing to try is two copies of S^1 glued together at a single point. Path a_1 can then be a path once-around circle 1 — essentially ignoring circle 2; and a_2 can go once around circle 2. Several things are not obvious at this point. It is not obvious that these two paths are enough to generate all paths up to path-homotopy in X. And it is not obvious what relations there might be. We will come back to these later. But in fact

$$\pi_1(X) \cong \langle a_1, a_2; \rangle$$

The next example again has two generators. But this time with relations. It will be interesting to see how relations (and in particular the relation that we are about to write) can be seen as arising naturally in the fundamental group context (i.e. as path-homotopies). And the next example is a good paradigm for this. But to see how it works we will need to get back to the topology!

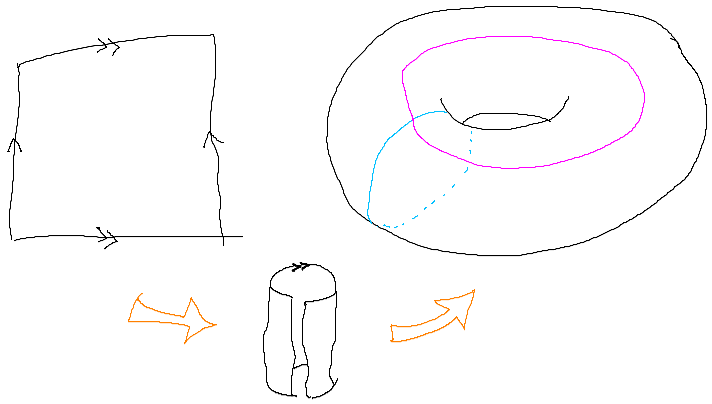
Example 12.33. Let $A = \{a, b\}$ and let $R = \{aba^{-1}b^{-1}\}$ so

$$G = \langle a, b; aba^{-1}b^{-1}\rangle.$$

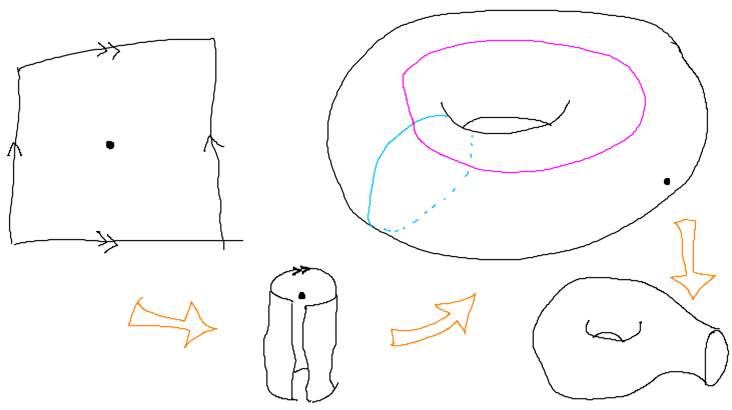
Thus, in G we have $aba^{-1}b^{-1} = e$ and so $ab = ba$. In particular G must be abelian and any element can be written in the form $a^m b^n$ for some m, n. It can now easily be shown that the group is isomorphic to $(\mathbb{Z}^2, +)$ (where $(a, b) + (c, d) = (a + c, b + d)$ as in $(\mathbb{R}^2, +)$).

Can we come up with a space X that might have this group as its fundamental group? Firstly we need a space that has at least a couple of paths which are not 'generated by each other'. From this perspective, something to try is the torus — the surface of the donut — since we have paths that go around the donut and paths that go through the donut.
But again it is distinctly un-obvious what relations there might be. (Possibly it is not even clear that every path is in a class generated by these two generators. We will sort out all these questions later. The role of π_1 here is just to provide a bit of an intuitive framework.)

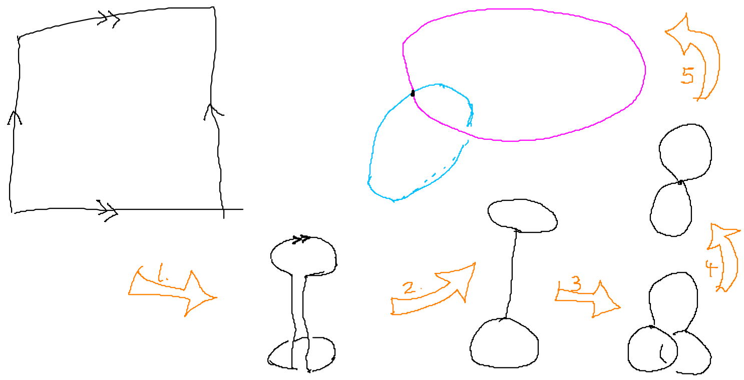
Perhaps oddly, a torus with a puncture in it is actually a bit simpler to work with than a perfect torus. So from the torus — which we built by 'identifying' (i.e. equivalencing under an equivalence relation $\sim$) opposite sides in a copy of $[0, 1]^2$ (with identified sides indicated by matched arrows):



we can build a punctured torus:



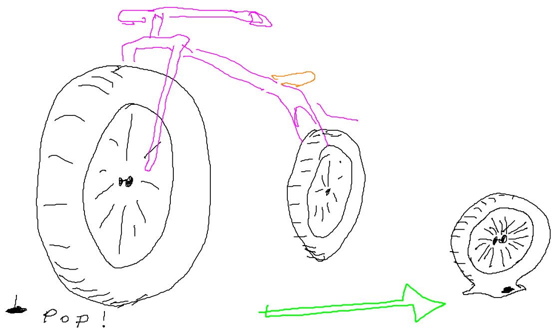
Just in case the puncture is hard to spot, in the last picture we have made the hole a bit bigger. It probably looks like the puncture has made things worse. But remember how there was a deformation retract that expanded a puncture and thus made the punctured disk into a topological circle? If we do something analogous here then we can expand the puncture until all that remains is the 'frame' of our square. Under the equivalence, realised by the folding, this frame folds up until it is just two circles joined at a point.



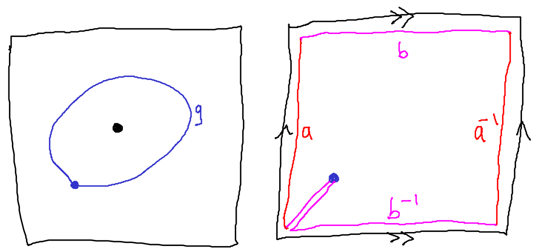
So in other words with the puncture the blue and purple paths exactly generate π_1, and there are no more relations. We have, writing T for the torus $[0,1]^2/\sim$ and $\bullet \in T$ for the puncture point:

$$\pi_1(T \setminus \{\bullet\}) \cong \langle a_1, a_2; \rangle$$

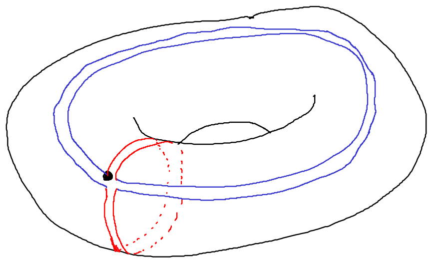
Now to get back to the torus what we need is a puncture repair kit.



The puncture repair kit consists of a patch without a hole — let's say a copy P of $(0,1)^2$, and a ring of glue applied to $(T \setminus \{\bullet\}) \cap P = P \setminus \{\bullet\}$.
On the $(T \setminus \{\bullet\})$ side the ring of glue g, say, is not path-homotopic to the trivial path, because of the puncture. But on the P side it is trivial (P is simply connected). So the 'patch' induces a relation that says any such path g should be made equivalent to the identity. The glue ring g is, when considered in the punctured torus, path-homotopic to a path written in terms of the generators a and b (say, as indicated in the picture) of the group for the punctured torus:



Thus $g \rightsquigarrow aba^{-1}b^{-1} \cong 1$.
Drawing this path back on a traditional torus perhaps you can see directly that it is indeed homotopic to the trivial path, after a lot of unwrapping:



So the patch introduces a relation $aba^{-1}b^{-1}$. Altogether then, we have after patching back to the torus:

$$\pi_1(T) = \langle a, b; aba^{-1}b^{-1} \rangle$$

In summary we took the π_1 of the punctured tyre; and the π_1 of the puncture repair patch; and then added in relations matching the two structures together across the intersection region.

Shortly we will lift this informal construction to a formal Theorem.

Example 12.34. Let $G = \langle a, b; abab^{-1} \rangle$. This group is *not* abelian, and we claim that $G \cong \langle c, d; c^2d^2 \rangle$. Without writing a formal proof (which isn't difficult), letting $c = ab$ and $d = b^{-1}$ notice that any word in $\{a, b\}$ can be written as a word in $\{c, d\}$ (and vice versa) since $a = cd$ and $b = d^{-1}$. Furthermore the relation $abab^{-1} = e$ is equivalent to $c^2d^2 = e$.

Note that if A is empty then the only word is the empty word e – so $\langle\ ;\ \rangle$ is the trivial group $\{e\}$.

The Theorem

Theorem 12.35 (Seifert-Van Kampen)**.** *Suppose that U, V are path-connected open subsets of a topological space X such that $X = U \cup V$ and $U \cap V$ is nonempty and path connected. Let $x \in U \cap V$. Suppose that the fundamental groups of $U, V, U \cap V$ have the following presentations:*

$$\begin{aligned} \pi_1(U, x) &\cong \langle a_1, \ldots, a_l;\ r_1, \ldots, r_i \rangle \\ \pi_1(V, x) &\cong \langle b_1, \ldots, b_m;\ s_1, \ldots, s_j \rangle \\ \pi_1(U \cap V, x) &\cong \langle c_1, \ldots, c_n;\ t_1, \ldots, t_k \rangle. \end{aligned}$$

Let $\phi : U \cap V \to U$, $\psi : U \cap V \to V$ be the inclusions. Then

$$\pi_1(X, x) \cong \langle a_1, \ldots, a_l, b_1, \ldots, b_m\ ;\ \ r_1, \ldots, r_i, s_1, \ldots, s_j, \phi_*(c_1)\psi_*(c_1)^{-1}, \ldots, \phi_*(c_n)\psi_*(c_n)^{-1} \rangle.$$

Here is that final long construction again with some braces added as a guide to the eye:

$$\pi_1(X, x) \cong \langle \underbrace{a_1, \ldots, a_l}_{\text{from } U}, \underbrace{b_1, \ldots, b_m}_{\text{from } V};\ \underbrace{r_1, \ldots, r_i}_{U}, \underbrace{s_1, \ldots, s_j}_{V}, \underbrace{\overbrace{\phi_*(c_1)\psi_*(c_1)^{-1}}^{\text{consistency}}, \ldots, \phi_*(c_n)\psi_*(c_n)^{-1}}_{U \cap V} \rangle.$$

So — making reference to the puncture repair example above — U is the punctured tyre; V is the puncture repair patch (or the roles can be reversed); and $U \cap V$ is the shape of the glue joining them together.

Remark 12.36. Each generator corresponds to some homotopy class of loops based at x, $a_i = [\alpha_i]$, say.

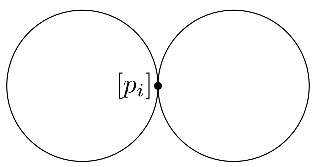
Notice that $\phi_*(c_i) \in \pi_1(U, x)$ and $\psi_*(c_i) \in \pi_1(V, x)$, thus we respectively have that $\phi_*(c_i)$ is a word in $\{a_1, \ldots, a_l\}$ and $\psi_*(c_i)$ is a word in $\{b_1, \ldots, b_m\}$. The new relations are therefore well-defined and tell us how the homotopy classes of the fundamental groups $\pi_1(U, x)$, $\pi_1(V, x)$ interact with one-another.

A careful proof of this Theorem is quite long — formalising the construction from the puncture repair example — and will not be included here (it can be found in many textbooks, such as Armstrong). Here we will only *use* the Theorem to calculate some more fundamental groups.

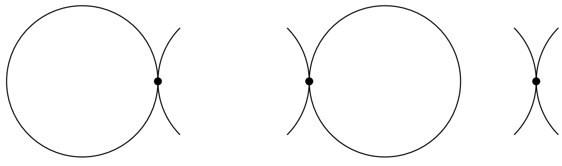
Examples

Given X and Y two topological spaces, then we can form the **disjoint union** of these spaces to get a new topological space $Z = X \coprod Y$. The topology on Z is formed by considering sets $W = U \coprod V$ for $U \subset X$ open and $V \subset Y$ open. Inductively we can do this for any finite number of topological spaces.

Example 12.37. Let $X = S^1_1 \coprod S^1_2$ be the disjoint union of two copies of $S^1 \subset \mathbb{R}^2$. Pick points $p_i \in S^1_i$ and define $\sim$ on X via $p_1 \sim p_2$ and if $x, y \notin \{p_1, p_2\}$ then $x \sim y \iff x = y$. We end up with ${}^{X}/_{\sim}$ looking like:



Pick $U \subset {}^{X}/_{\sim}$ and $V \subset {}^{X}/_{\sim}$ both open and path connected sets as in the next picture (with $U \cap V$ shown at the end)



Both U and V deformation retract into a circle S^1_i, and $U \cap V$ deformation retracts to a point $[p_i]$. So we have $\pi_1(U,[p_i]) \cong \langle a;\cdot\rangle \cong \mathbb{Z}$, $\pi_1(V,[p_i]) \cong \langle b;\cdot\rangle \cong \mathbb{Z}$ and $\pi_1(U\cap V,[p_i]) \cong \langle\cdot;\cdot\rangle \cong 0$.

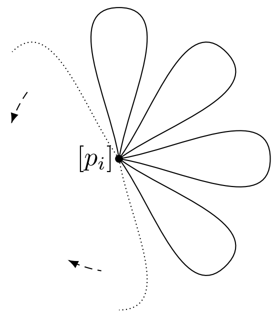
The Van Kampen Theorem tells us, therefore, that

$$\pi_1(X/\sim) \cong \langle a,b;\cdot\rangle$$

which is a free group on two generators (and is *not* Abelian). Notice that a is the homotopy class of a loop winding once around the first S^1 (say anti-clockwise), and b is the homotopy class of a loop winding once around the second S^1 (clockwise): Van Kampen's Theorem is thus telling us that $\pi_1(X/\sim)$ is completely determined by a and b (and there is indeed no relation between these two loops). Notice that we do not really need to be precise about which "directions" a and b are winding, as long as they go around once and once only.

Example 12.38. Let $X = \coprod_{i=1}^n S^1_i$, the disjoint union of n copies of $S^1 \subset \mathbb{R}^2$. Pick a point $p_i \in S^1_i$ for each i, and define $\sim$ on X via $p_i \sim p_j$ for all i,j, otherwise $x \sim y \iff x = y$. We end up with $X/\sim$ being n copies of S^1 joined together at a single point $[p_i] = \{p_j : j = 1,\dots,n\}$.

Pictorially:



By induction on the previous example we end up with $\pi_1(X/\sim) \cong \langle a_1,\dots a_n;\cdot\rangle$ which is a free group on n generators. Once again, each a_i is a homotopy class of a loop winding once around S^1_i.

Example 12.39. The spheres $S^n \subset \mathbb{R}^{n+1}$ for $n \geq 2$. (If we think of S^2, say, then intuitively every closed loop can be contracted to a point without obstruction. But can we verify this formally?)

Recalling Example 9.10 (the stereographic projection) we know that $U = S^n\backslash\{N\}$ and $V = S^n\backslash\{S\}$ are both homeomorphic to $\mathbb{R}^n$. In particular they are simply connected. Furthermore $S^n = U \cup V$ and

$$U \cap V = S^n\backslash\{N,S\} \cong \mathbb{R}^n\backslash\{0\}$$

is path connected (since $n \geq 2$).

Whatever the generators for $\pi_1(U\cap V,x)$ might be, their images in the trivial groups $\pi_1(U)$ and $\pi_1(V)$ must be trivial. So any glue relation must actually be trivial. So it almost follows from the Van Kampen Theorem already that

$$\pi_1(S^n) = \pi_1(S^n,x) \cong \langle\ ;\ \rangle$$

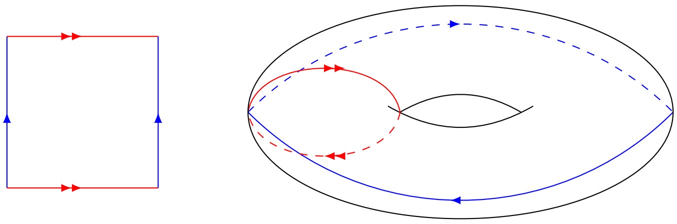
is trivial. (This is correct, but ...)

...Note that this is cutting a logical corner. We should at least check that $\pi_1(U\cap V,x)$ is of the form $\langle c_1,\dots,c_n;t_1,\dots t_k\rangle$. *A-priori* there is no guarantee of this. There are several ways of getting around this. One is that there is a more general version of the Van Kampen Theorem that applies in this situation. Another is the following:

Theorem 12.40. *Suppose that X is a topological space and $U, V \subset X$ are open and simply connected. Suppose that $X = U \cup V$ and $U \cap V \neq \emptyset$ is path connected. Then X is simply connected.*

A proof of this is possible for example using Lebesgue numbers. We will not need the details here.

Example 12.41. The torus T^2. Recall that this topological space is a quotient space obtained via the square $X = [0,1] \times [0,1] \subset \mathbb{R}^2$ and the equivalence relation $\sim$ as shown in the diagram below:



We will use Van Kampen's Theorem to prove that $\pi_1(T^2) = \pi_1(T^2, x) = (\mathbb{Z}^2, +)$.
Let $y = (1/2, 1/2) \in [0,1]^2$. Then let $\tilde{U} = [0,1]^2 \backslash \{y\}$ and

$$\tilde{V} = [0,1]^2 \backslash \partial [0,1]^2 \quad = \quad (0,1) \times (0,1).$$

where $\partial[0,1]^2$ is the boundary of the square. We set

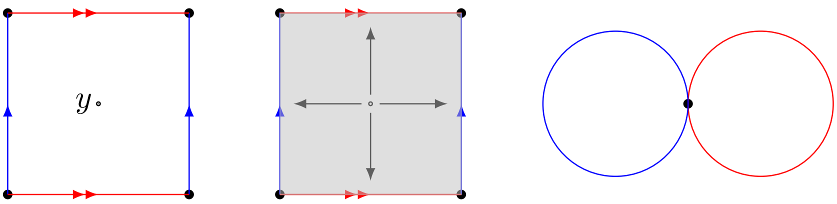
$$U = \tilde{U} /_{\sim}$$

(thus a torus with a point missing — an open subset) and

$$V = \tilde{V} /_{\sim} = \tilde{V},$$

giving $U \cap V = V \backslash \{[y]\}$.

For U notice that it deformation retracts onto $\partial X /_{\sim}$, the boundary of the square (under an equivalence relation), as the following picture indicates.
Notice that $\partial X /_{\sim}$ is two circles joined together at a point (it was the point $(0,0) \in [0,1]^2$, or any one of the corners, before the equivalence) as in Example 12.37:



Let $\tilde{a}$ denote the homotopy class of a curve going once around the blue loop (thus from $[(0,0)]$ to $[(0,0)]$) and $\tilde{b}$ denote the homotopy class of a curve going once around the red loop (in some prescribed directions). Then we have

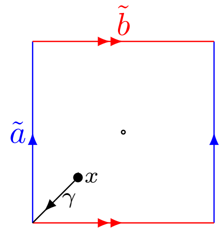
$$\pi_1(U, [(0,0)]) \cong \pi_1\left(\partial X /_{\sim}, [(0,0)]\right) \cong \langle \tilde{a}, \tilde{b}; \cdot \rangle$$

by Proposition 12.26.

The point $[(0,0)]$ is not strictly in V. So we want to compute $\pi_1(U,x)$ (U is our punctured tyre from earlier), $\pi_1(V,x)$ (V is our puncture repair patch) and $\pi_1(U\cap V,x)$ (our glue shape) for $x\in U\cap V$. To that end, let $x=[(1/4,1/4)]$, say. By Theorem 12.11 we have

$$\pi_1(U,x)\cong\pi_1(U,[(0,0)])$$

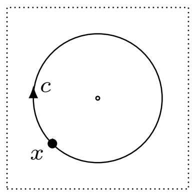
via the curve γ:



Thus if we set $a=[\gamma]*\tilde{a}*[\overline{\gamma}]$ and $b=[\gamma]*\tilde{b}*[\overline{\gamma}]$ then we have

$$\pi_1(U,x)\cong\langle a,b;\cdot\rangle. \tag{12.1}$$

Space V is clearly simply connected so $\pi_1(V,x)\cong\langle\cdot;\cdot\rangle$. It remains to study $\pi_1(U\cap V,x)$ which is $(0,1)\times(0,1)\backslash\{[y]\}$ as in the diagram below.



We can see (e.g. via (12.24)) that $U\cap V$ deformation retracts onto a copy of S^1 (a circle of radius $1/2$ say). Thus

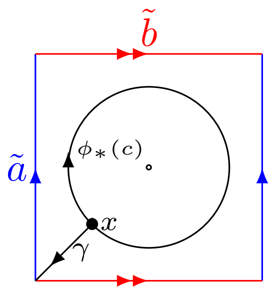
$$\pi_1(U\cap V,x)=\langle c;\cdot\rangle\cong\mathbb{Z}.$$

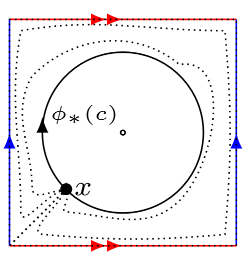
The homotopy class c is represented in the above diagram and corresponds to winding once around $[y]$ in a clockwise direction.

Now we have computed all the fundamental groups, we next use the Van Kampen Theorem to knit them together and give the fundamental group $\pi_1(T^2)$.

So what does $\phi_*(c)$ look like in U? We need to express the image of path c from $U\cap V$ (the image is just the 'same' path, but in a different subspace) in terms of the a,b generators.

Consider the pictures below. The dashed curves indicate a homotopy from a representative of $[\gamma]*\tilde{a}\tilde{b}\tilde{a}^{-1}\tilde{b}^{-1}*[\overline{\gamma}]$ to a representative of $\phi_*(c)$.





Thus via (12.1)

$$\phi_*(c)=aba^{-1}b^{-1}=[\gamma]*\tilde{a}\tilde{b}\tilde{a}^{-1}\tilde{b}^{-1}*[\overline{\gamma}].$$

Since $\psi_*(c)=e$ (V is simple connected), then Van Kampen's Theorem gives us

$$\pi_1(T^2)\cong\langle a,b;aba^{-1}b^{-1}\rangle\cong(\mathbb{Z}^2,+).$$

Example 12.42. The real projective spaces $\mathbb{RP}^n$. Define $\mathbb{RP}^n := \mathbb{R}^{n+1}\backslash\{0\} \big/ \sim$ where $x \sim y$ if and only if there exists $\lambda \in \mathbb{R}_*$ so that $x = \lambda y$. In other words $\mathbb{RP}^n$ is the topological space of lines passing through the origin in $\mathbb{R}^n$.

Consider the map $F : \mathbb{R}^{n+1}\backslash\{0\} \to S^n$ given by $F(x) = x/|x|$ and notice that $x \sim y$ if any only if $F(x) = \pm F(y)$ and so by Proposition 7.50 F induces a continuous map $\tilde{F} : \mathbb{RP}^n \to S^n \big/ \sim_1$ where $x \sim_1 y \iff x = \pm y$. The map $\tilde{F}$ is in fact a homeomorphism with inverse $\tilde{F}^{-1}([z]_1) = [z]$. Notice that this immediately tells us that $\mathbb{RP}^n$ is compact, connected and Hausdorff.

We now trivially have that

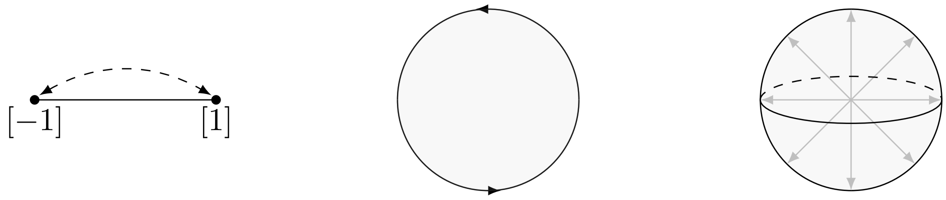
$$G : S^n \big/ \sim_1 \to S^n_+ \big/ \sim_1, \qquad G([x]_1) = [x]_1$$

is a homeomorphism. Recall from example 9.11 that $S^n_+ \cong D^n$ via $f : D^n \to S^n_+$. Remember that $D^n = \{x \in \mathbb{R}^n : |x| \leq 1\}$ and we now let $x \sim_2 y \iff x = y$, or $|x| = |y| = 1$ **and** $x = \pm y$. Notice that $x \sim_2 y$ if and only if $f(x) \sim_1 f(y)$ so in fact $\tilde{f} : D^n \big/ \sim_2 \to S^n_+ \big/ \sim_1$ is a homeomorphism. In particular

$$\tilde{f}^{-1} \circ G \circ \tilde{F} : \mathbb{RP}^n \to D^n \big/ \sim_2$$

is a homeomorphism, i.e. $\mathbb{RP}^n \cong D^n \big/ \sim_2$.

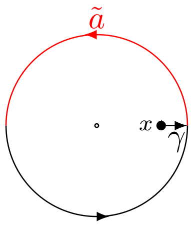
Below are some representations of $\mathbb{RP}^n \cong D^n \big/ \sim_2$ for $n = 1$, $n = 2$ and $n = 3$



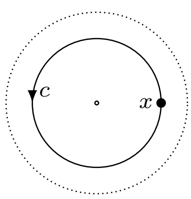
We can now see that $\mathbb{RP}^1 \cong S^1$ (using something similar to examples 7.40, 11.56), giving $\pi_1(\mathbb{RP}^1) \cong \mathbb{Z}$.

Similarly $\mathbb{RP}^2$ is exactly as it appears in Example 7.47 and we will prove that $\pi_1(\mathbb{RP}^2) \cong \mathbb{Z}_2$. Letting $\tilde{U} = D^2\backslash\{0\}$ and $\tilde{V} = B^2_1(0)$ we let $U = \tilde{U} \big/ \sim_2$ and $V = \tilde{V} \big/ \sim_2 = \tilde{V}$. Thus $U \cap V = B^2_1\backslash\{0\}$.

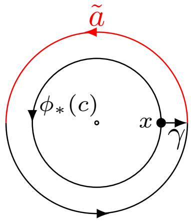
As in the torus case, U deformation-retracts onto $S^1 \big/ \sim_1 \cong S^1_+ \big/ \sim_1 \cong D^1 \big/ \sim_2 \cong \mathbb{RP}^1 \cong S^1$. Letting $\tilde{a}$ represent a generator of $\pi_1(S^1)$ we have that $\pi_1(U) \cong \langle \tilde{a}; \rangle$. Once again, we want to realise $\pi_1(U, x)$, this time for $x = [(0, 0.75)]$ (say) and we do this via γ as below, and set $a = [\gamma] * \tilde{a} * [\overline{\gamma}]$. Thus we have $\pi_1(U, x) \cong \langle a; \rangle$.



We trivially have $\pi_1(V, x) \cong \langle ; \rangle$, and almost as trivially (by now) we see that $\pi_1(U \cap V, x) \cong \langle c; \rangle$.



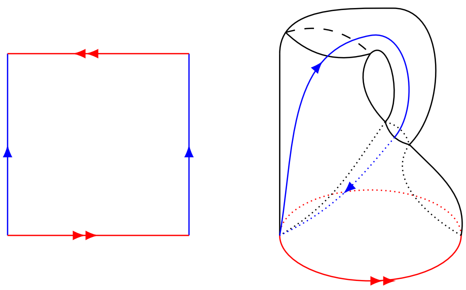
Again we must consider what $\phi_*(c) \in \pi_1(U,x)$ is in terms of a. The picture below suggests that we have $\phi_*(c) = a^2$ (which you can check directly, if you need to).



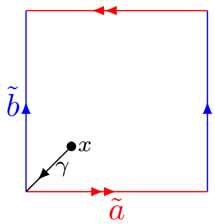
Thus Van Kampen's Theorem tells us that $\pi_1(\mathbb{RP}^2) \cong \langle a; a^2\rangle \cong \mathbb{Z}_2$.

In fact, one can prove that $\pi_1(\mathbb{RP}^n) \cong \mathbb{Z}_2$ for all $n \geq 2$ using Van Kampen's Theorem and the same method as above. By induction we may assume that we know this to be true for $\mathbb{RP}^{n-1}$ and we aim to prove it for $\mathbb{RP}^n \cong D^n/\sim_2$ when $n \geq 3$. In exactly the same way as above we set $U = D^n\backslash\{0\}/\sim_2$ and $V = B_1^n(0)$. The crux of the proof is that U deformation retracts onto $S^{n-1}/\sim_1 \cong \mathbb{RP}^{n-1}$ and thus $\pi_1(U) \cong \mathbb{Z}_2$. Now, since $n \geq 3$ we have $\pi_1(V) \cong \pi_1(U \cap V) \cong \{e\}$. Thus $\pi_1(\mathbb{RP}^n) \cong \mathbb{Z}_2$.

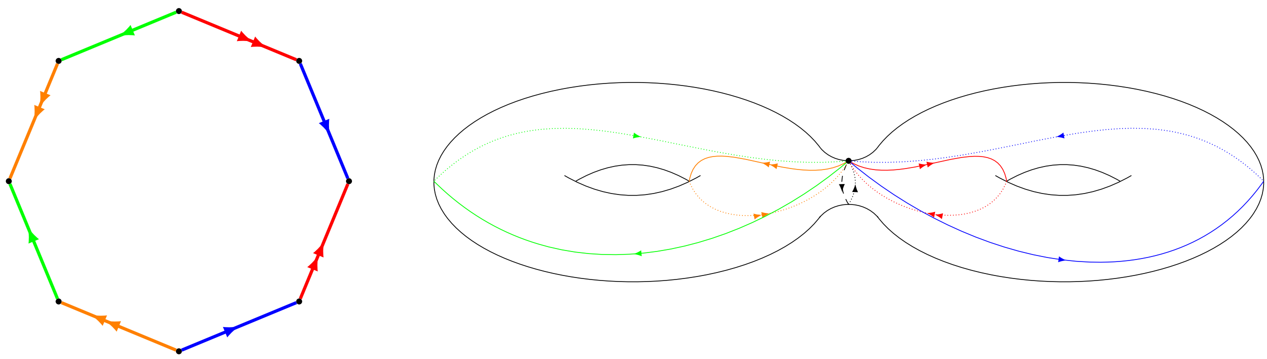
Example 12.43. The Klein bottle, K^2. Recall that this is the topological space obtained as a quotient of the square $X = [0,1] \times [0,1]$ as suggested below.



Computing the fundamental group here is very much like the torus above, so we will be very brief. Consider the curves $a = [\gamma] * \tilde{a} * [\overline{\gamma}]$ and $b = [\gamma] * \tilde{b} * [\overline{\gamma}]$ - noting from the picture below that this is subtlety different to the torus case. I will simply point out that we have $a * b * a * b^{-1} = [e_x]$ and leave it to the reader to prove (using Van Kampen and the same method for the torus) that $\pi_1(K^2, x) \cong \langle a, b; abab^{-1}\rangle \cong \langle \hat{a}, \hat{b}; \hat{a}^2\hat{b}^2\rangle$ (see example 4.11).



Example 12.44. A genus two surface Σ_2^2. Recall that this is obtained from an octagon via the gluing suggested by the colours and arrows of the edges.

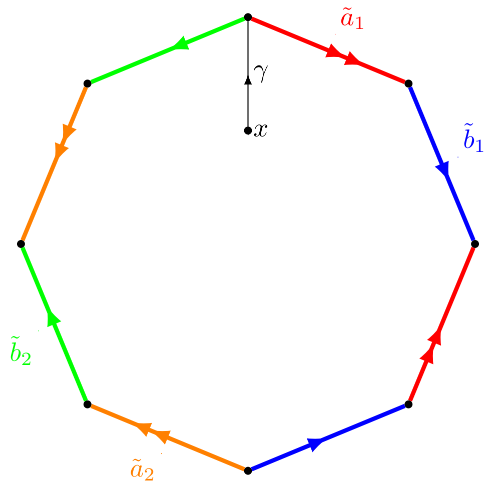


Define $\tilde{a}_i$ and $\tilde{b}_i$ as suggested below, and use γ to define $a_i = [\gamma] * \tilde{a}_i * [\overline{\gamma}]$, $b_i = [\gamma] * \tilde{b}_i * [\overline{\gamma}]$. First of all letting $[a_i, b_i] = a_i b_i a_i^{-1} b_i^{-1}$ you can check that

$$e_x \simeq [a_1, b_1][a_2, b_2].$$

Thus, using Van Kampen's Theorem (analogously as to the torus and Klein bottle) we have

$$\pi_1(\Sigma_2^2, x) \cong \langle a_1, b_1, a_2, b_2; [a_1, b_1][a_2, b_2] \rangle \cong \langle a_1, b_1, a_2, b_2; a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \rangle.$$



Let's review what we have learned. The fundamental group of $S^1 = \mathbb{RP}^1$ is $\mathbb{Z}$ under addition. The fundamental groups of the sphere S^n is trivial when $n \geq 2$. The fundamental group of the projective spaces $\mathbb{RP}^n$ is $\mathbb{Z}_2$, so is non-trivial but finite when $n \geq 2$. The fundamental group of the torus is infinite and abelian. The fundamental group of the Klein bottle is infinite and non-abelian. Finally the fundamental group of a genus two surface is infinite, non-abelian, and is generated by 4 generators. When we set $n = 2$, none of these groups are isomorphic to each other, thus S^2, $\mathbb{RP}^2$, T^2, K^2 and Σ_2^2 are topologically distinct from one-another (because the fundamental group is a topological invariant). So the fundamental group is a powerful tool!

12.5 Applications of the fundamental group

The last section! Here we will prove some interesting theorems using our knowledge of the fundamental group.

Theorem 12.45 (Brouwer)**.** *Let $D^2 = \{x \in \mathbb{R}^2 : |x| \leq 1\}$. Any continuous map $f : D^2 \to D^2$ has a fixed point.*

Proof. First note that S^1 is not a deformation retract of D^2 since $\pi_1(S^1, (1,0)) \cong \mathbb{Z}$ and $\pi_1(D^2, (1,0)) \cong \{e\}$.

Now suppose for contradiction that there exists a continuous map $f : D^2 \to D^2$ with no fixed point, i.e. such that $f(x) \neq x$ for all $x \in D^2$. For any $x \in D^2$, let L_x be the line starting at $f(x)$ and passing through x. If we continue this line far enough beyond x it will hit the boundary. Thus we can find $t > 0$ such that $L_x = \{(1-t)f(x) + tx : t > 0\}$ hits S^1, i.e. solve

$$(1-t)^2|f(x)|^2 + 2(1-t)t\langle x, f(x)\rangle + t^2|x|^2 = 0$$

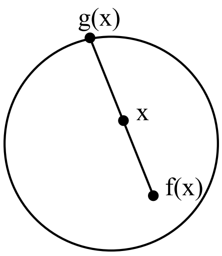
for $t > 0$. You can check that we have

$$t(x, f(x)) = \frac{-\langle f(x), x - f(x)\rangle + \sqrt{\langle f(x), x - f(x)\rangle^2 + |x - f(x)|^2(1 - |f(x)|^2)}}{|x - f(x)|^2} > 0,$$

which is a continuous function of x since $|x - f(x)| > 0$ for all x.

Let $g(x)$ be the point of intersection of L_x with the boundary S^1 of D^2 that is furthest from $f(x)$, so that

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



Then $g : D^2 \to S^1$ is a continuous map with $g(x) = x$ when $x \in S^1$. But then $H(x,t) = (1-t)x + tg(x)$ is a deformation retract of D^2 to S^1 which is our contradiction.

□

Theorem 12.46 (Borsuk-Ulam)**.** *There does not exist a continuous map $f : S^2 \to S^1$ such that $f(-x) = -f(x)$ for all $x \in S^2$.*

This theorem can be proved using only the fundamental group of S^1. Before presenting the proof, we will investigate some of its consequences.

Corollary 12.47. *Let $f : S^2 \to \mathbb{R}^2$ be a continuous map. Then there exists a point $x \in S^2$ such that $f(-x) = f(x)$.*

This corollary implies that at any given moment in time there exist two antipodal points on the surface of the Earth that have the same temperature and pressure.

Proof. Suppose for contradiction that there exists a continuous map $f : S^2 \to \mathbb{R}^2$ such that $f(-x) \neq f(x)$ $\forall x \in S^2$. Then we may define $g : S^2 \to S^1$ by

$$g(x) = \frac{f(x) - f(-x)}{|f(x) - f(-x)|}.$$

This definition is valid because $|f(x) - f(-x)| \neq 0$, moreover g is obviously continuous. However, g satisfies

$$g(-x) = \frac{f(-x) - f(x)}{|f(x) - f(-x)|} = -g(x).$$

But this contradicts the Borsuk-Ulam Theorem, which states that no continuous map g with this property exists. Therefore the map f cannot exist. $\square$

Corollary 12.48. *No subset of* $\mathbb{R}^2$ *is homeomorphic to* S^2

Proof. Suppose that $f : S^2 \to U$ is a homeomorphism from S^2 to a subset U of $\mathbb{R}^2$. Let $i : U \to \mathbb{R}^2$ be the inclusion; then $i \circ f : S^2 \to \mathbb{R}^2$ is continuous and injective. By the previous corollary there exists a point $x \in S^2$ such that $i \circ f(x) = i \circ f(-x)$, but this contradicts the fact that $i \circ f$ is injective. Therefore no such homeomorphism f exists. $\square$

Corollary 12.49. $\mathbb{R}^2$ *is not homeomorphic to* $\mathbb{R}^3$*.*

Proof. If $f : \mathbb{R}^3 \to \mathbb{R}^2$ is a homeomorphism, then when we restrict $f|_{S^2} : S^2 \to f(S^2) \subset \mathbb{R}^2$, it is a homeomorphism onto its image. Corollary 12.48 rules this out so no such f can exist. $\square$

Now for the proof of the Borsuk-Ulam Theorem:

Proof of Borsuk-Ulam. We start with an important

Lemma 12.50. *Suppose that* $h : S^1 \to S^1$ *is continuous and satisfies* $h(-x) = -h(x)$ *for all* x*. Defining* $\alpha(s) = h(\cos(2\pi s), \sin(2\pi s))$ *so that* $\alpha : [0,1] \to S^1$ *is a loop (based at some point* $y = h((1,0)) \in S^1$*), we have* $N(\alpha) = 2k+1$ *for some* $k \in \mathbb{Z}$*, thus* $[\alpha] \neq [e_y]$ *is homotopically non-trivial.*

Proof of Lemma 12.50. The condition that $h(-x) = -h(x)$ implies that for all $s \in [0,1/2]$, $\alpha(s) = -\alpha(s+1/2)$. Letting $\tilde{\alpha}$ be a lift of α, we have that $\tilde{\alpha}(s+1/2)$ differs from $\tilde{\alpha}(s)$ by $k(s) + 1/2$ for some $k(s) \in \mathbb{Z}$, and for all $s \in [0,1/2]$. In other words, for all $s \in [0,1/2]$ there exists $k(s) \in \mathbb{Z}$ such that

$$\tilde{\alpha}(s+1/2) - \tilde{\alpha}(s) = k(s) + 1/2.$$

Since the left hand side of the above expression is continuous for $s \in [0,1/2]$, it implies that $k(s)$ is continuous and thus $k(s) \equiv k$ is constant (since it must always be an integer). In particular,

$$\tilde{\alpha}(1) = \tilde{\alpha}(1/2) + k + 1/2 = \tilde{\alpha}(0) + 2k + 1.$$

We have proved that $N(\alpha) = \tilde{\alpha}(1) - \tilde{\alpha}(0) = 2k+1$. $\square$

Now, assuming that $f : S^2 \to S^1$ is continuous and satisfies $f(-x) = -f(x)$ we will derive a contradiction. We restrict f to the upper hemisphere $f|_{S^2_+} : S^2_+ \to S^1$. By Example 9.11, $S^2_+ \cong D^2 = \{x \in \mathbb{R}^2 : |x| \leq 1\}$ via $G(x_1,x_2,x_3) = (x_1,x_2)$ and we therefore have a continuous function $g = f|_{S^2_+} \circ G^{-1} : D^2 \to S^1$. Letting $h = g|_{S^1} : S^1 \to S^1$, which is continuous, we also have, for all $x \in S^1$, $h(-x) = h(-x_1,-x_2) = f(-x_1,-x_2,0) = -f(x_1,x_2,0) = -h(x)$.

Defining α as in Lemma 12.50, $H(s,t) = g((1-t)\cos(2\pi s)+t, (1-t)\sin(2\pi s))$ is a homotopy from α to e_y, i.e. $[e_y] = [\alpha] \neq [e_y]$ which is our contradiction, so no such f can exist.

$\square$

12.5.1 Fundamental theorem of algebra (and more)

Theorem 12.51 (Fundamental Theorem of algebra)**.** *Every non-constant complex polynomial has a root.*

Corollary 12.52. *Every complex polynomial of degree $k \in \mathbb{Z}_+$ has k roots, counted with multiplicities.*

Proof. Suppose for contradiction that there exists a non-constant polynomial P with no root. We may assume without loss of generality that the polynomial takes the form

$$P(z) = a_0 + a_1 z + \ldots + a_{k-1} z^{k-1} + z^k$$

for some $k \geq 1$. If $a_0 = 0$ then $z = 0$ is a root, so it must be that $a_0 \neq 0$.

Choose $R > 0$ large enough that

$$R^k > |a_0| + |a_1|R + \ldots + |a_{k-1}|R^{k-1}. \qquad (*)$$

Let $\alpha : [0,1] \to \mathbb{C} \setminus \{0\}$ be the following loop based at 1:

$$\alpha(s) = \frac{P(Re^{2\pi i s})}{P(R)}.$$

We claim that:

- α is path homotopic to the constant loop e_1
- α is path homotopic to the loop $\beta : s \mapsto e^{2\pi i k s}$.

Together these lead to a contradiction, because e_1 has winding number 0 and β has winding number k.

To prove the first claim, let $F : [0,1] \times [0,1] \to \mathbb{C} \setminus \{0\}$ be the following function:

$$F(s,t) = \frac{P(Rte^{2\pi i s})}{P(Rt)}.$$

The denominator in this expression is never zero because P has no root. The image of this function is contained in $\mathbb{C} \setminus \{0\}$ for the same reason. F is continuous with $F(0,t) = F(1,t) = 1$ $\forall t \in [0,1]$, and $F(s,0) = 1$, $F(s,1) = \alpha(s)$ $\forall s \in [0,1]$. So F is a path homotopy from e_1 to α.

To prove the second claim, let $G : [0,1] \times [0,1] \to \mathbb{C} \setminus \{0\}$ be the following function:

$$G(s,t) = \frac{(1-t)(a_0 + a_1 Re^{2\pi i s} + \ldots + a_{k-1}(Re^{2\pi i s})^{k-1}) + (Re^{2\pi i s})^k}{(1-t)(a_0 + a_1 R + \ldots + a_{k-1} R^{k-1}) + R^k}.$$

The inequality $(*)$ implies that the numerator and denominator in this expression are nonzero, because

$$\begin{aligned}
&|(1-t)(a_0 + a_1 Re^{2\pi i s} + \ldots + a_{k-1}(Re^{2\pi i s})^{k-1}) + (Re^{2\pi i s})^k| \\
&\qquad \geq R^k - (1-t)\left(|a_0| + |a_1|R + \ldots + |a_{k-1}|R^{k-1}\right) \\
&\qquad > 0,
\end{aligned}$$

and thus G is continuous. We have that $G(0,t) = G(1,t) = 1$ $\forall t \in [0,1]$ and that $G(s,0) = \alpha(s)$ and $G(s,1) = \beta(s)$ $\forall s \in [0,1]$. Therefore G is a path homotopy from α to β. □

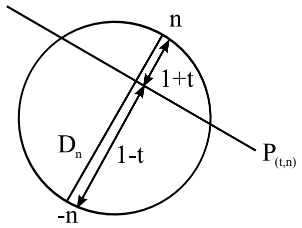
Theorem 12.53 (Ham Sandwich Theorem)**.** *Let A, B, C be bounded subsets of $\mathbb{R}^3$. Then there is a plane in $\mathbb{R}^3$ which simultaneously divides each region exactly in half by volume.*

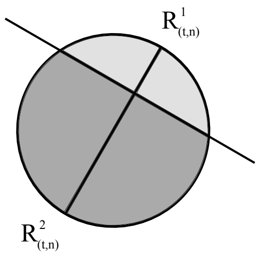
Remark 12.54. How would you adapt this theorem to prove an analogous statement in $\mathbb{R}^2$? You would consider two bounded subsets A, B of $\mathbb{R}^2$ and prove that there was a line which simultaneously divides both subsets in half by area. After you understand the proof below you should try to prove this (easier) version!

Do you think there is an analogous statement in $\mathbb{R}^n$? (You should!)

Do you think the above theorem works for **four** bounded subsets A, B, C, D of $\mathbb{R}^3$? You should not, but why?

Proof. We may re-scale our space such that A, B and C are contained in the closed unit ball. For any $n \in S^2$ let D_n denote the diameter of the ball through n, and for any $t \in [-1, 1]$ let $P_{(t,n)}$ be the plane orthogonal to D_n which passes through the point at distance $1 + t$ from n. The plane $P_{(t,n)}$ divides the unit ball into two regions $R^+_{(t,n)}$ and $R^-_{(t,n)}$ such that $n \in R^+_{(t,n)}$ and $-n \in R^-_{(t,n)}$.





For any fixed $n \in S^2$, let

$$V(t) = \text{Vol}(A \cap R^+_{(t,n)}) - \text{Vol}(A \cap R^-_{(t,n)}).$$

Then $V : [-1, 1] \to \mathbb{R}$ is continuous (this is intuitively clear, but requires a bit of measure theory to prove rigorously). Moreover, $V(-1) = -\text{Vol}(A)$ and $V(1) = \text{Vol}(A)$. We wish to choose a number α such that $V(\alpha) = 0$. Such a number exists by the intermediate value theorem. Since $V(t)$ is monotonically increasing, the set of points t such that $V(t) = 0$ is either a singleton $\{b\}$ or a closed interval $[c, d]$. In the former case we define $\alpha(n) = b$, and in the latter case we define $\alpha(n) = (c + d)/2$.

By repeating this process for every $n \in S^2$ we obtain a function $\alpha : S^2 \to [-1, 1]$. This function is continuous (this is intuitively clear, but again requires a bit of measure theory to make rigorous). It satisfies $\alpha(-n) = -\alpha(n)$, because $P_{(-t,n)} = P_{(t,-n)}$.

Now let

$$\begin{aligned} f(n) &= \text{Vol}(B \cap R^+_{(\alpha(n),n)}) \\ g(n) &= \text{Vol}(C \cap R^+_{(\alpha(n),n)}). \end{aligned}$$

The function $F : S^2 \to \mathbb{R}^2$ defined by $F(n) = (f(n), g(n))$ is again continuous (apply intuition and measure theory!). By corollary 12.47 there exists a point $n \in S^2$ such that $F(-n) = F(n)$. Thus,

$$\begin{aligned} \text{Vol}(B \cap R^+_{(\alpha(n),n)}) &= \text{Vol}(B \cap R^+_{(\alpha(-n),-n))}) \\ &= \text{Vol}(B \cap R^+_{(-\alpha(n),-n)}) \\ &= \text{Vol}(B \cap R^-_{(\alpha(n),n)}), \end{aligned}$$

and we have that $P_{(\alpha(n),n)}$ divides B in half by volume. By a similar calculation $P_{(\alpha(n),n)}$ divides C in half by volume, and by construction $P_{(\alpha(n),n)}$ divides A in half by volume.

□

Appendix A

Set Theory

A.1 Sets

A **set** X is a collection of 'objects'.

If an object x belongs to the set X we write $x \in X$ or $X \ni x$. If x does not belong to the set X we write $x \notin X$.

A **subset** A of a set X is a set A such that $x \in A \Rightarrow x \in X$. A shorthand for this is $A \subset X$. Note that $X \subset X$ (the notation $\subseteq$ may also be used).

You may be familiar with the following sets:

- $\mathbb{R}$ the set of real numbers
- $\mathbb{Z}$ the set of integers
- $\mathbb{Z}^+$ the set of positive integers (note $0 \notin \mathbb{Z}^+$)
- $\mathbb{N}$ the set of non-negative integers, or natural numbers (note $0 \in \mathbb{N}$)
- $\mathbb{Q}$ the set of rational numbers
- $\emptyset$ the empty set (the unique set containing no elements).

A.2 Set arithmetic

Given two sets X and Y,

- their **intersection** is the set $X \cap Y = \{z : z \in X \text{ and } z \in Y\}$;
- their **union** is the set $X \cup Y = \{z : z \in X \text{ or } z \in Y\}$;
- and the **difference** of X and Y is the set $X \setminus Y = \{z : z \in X \text{ and } z \notin Y\}$.

If A is a subset of X then $X \setminus A$ is sometimes called the **complement** of A. If $X \cap Y = \emptyset$ then X and Y are called **disjoint**.

The **Cartesian product** of two sets X and Y is the set

$$X \times Y = \{(x, y) : x \in X \text{ and } y \in Y\}.$$

The **power set** of a set X is the set of all subsets of X.

A.3 Relations and Functions

Given two sets X and Y, a **function** $f : X \to Y$ is a rule that assigns to every element $x \in X$ and element $f(x) \in Y$. In this case the set X is called the **domain** of f and the set Y is called the **codomain** of f.

If $A \subset X$, the **image** of A under f is the subset $f(A) \subset Y$ defined by $f(A) := \{f(a) : a \in A\}$. The **range** of f is the set $f(X)$. If $B \subset Y$, the **preimage** of B under f is the subset $f^{-1}(B) \subset X$ defined by $f^{-1}(B) = \{x \in X \ : \ f(x) \in B\}$. Note that the preimage $f^{-1}(B)$ is well-defined regardless of whether f is invertible. The preimage behaves well with respect to set theoretic operations, as opposed to the image.

A function $f : X \to Y$ is called **injective** if $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$, and is called **surjective** if $f(X) = Y$. A function which is both injective and surjective is called **bijective**. One can rephrase these definitions in terms of the preimage of sets. Indeed, let $A \subset X$ and $B \subset Y$. Then

- $A \subset f^{-1}(f(A))$, with equality if f is injective.
- $f(f^{-1}(B)) \subset B$, with equality if f is surjective.

Let $f : X \longrightarrow Y$ and $g : Y \longrightarrow Z$ be functions. The **composition** $g \circ f$ is defined by $(g \circ f)(x) = g(f(x))$.

A.4 Collections of sets

The remainder of this section will explain how to deal with collections of sets, and in particular how to take the intersection and union of these.

Definition A.1. Let X and Λ be two sets such that $\Lambda \neq \emptyset$. For each $\lambda \in \Lambda$, let A_λ be a subset of X. Then the collection of all sets A_λ with $\lambda \in \Lambda$ is called an *indexed family of subsets of* X, and is denoted $\{A_\lambda\}_{\lambda\in\Lambda}$ or $\{A_\lambda \ : \ \lambda \in \Lambda\}$. The set Λ is called the **indexing set**.

Definition A.2. Let $\{A_\lambda\}_{\lambda\in\Lambda}$ be an indexed family of subsets of X. The **union** of this family is the set

$$\bigcup_{\lambda\in\Lambda} A_\lambda := \{x \in X \ : \ \text{there exists } \lambda \in \Lambda \text{ such that } x \in A_\lambda\}.$$

The **intersection** of this family is the set

$$\bigcap_{\lambda\in\Lambda} A_\lambda := \{x \in X \ : \ \text{for every } \lambda \in \Lambda, \ x \in A_\lambda\}.$$

Remark A.3. Actually, if Λ is uncountable, we need the axiom of choice to even make sense of $\bigcap_{\lambda\in\Lambda} A_\lambda$, but let's ignore this subtle point.

Remark A.4. The following remarks follow immediately from the definition:

- If Λ is a finite set, say $\Lambda = \{1, 2, \ldots, n\}$, then

$$\bigcup_{\lambda\in\Lambda} A_\lambda = \bigcup_{\lambda=1}^{n} A_\lambda = A_1 \cup A_2 \cup \cdots \cup A_n$$

$$\bigcap_{\lambda\in\Lambda} A_\lambda = \bigcap_{\lambda=1}^{n} A_\lambda = A_1 \cap A_2 \cap \cdots \cap A_n.$$

- $\bigcup_{\lambda\in\Lambda} A_\lambda$ and $\bigcap_{\lambda\in\Lambda} A_\lambda$ are both subsets of X.

- For every $\mu \in \Lambda$,

$$U_\mu \subset \bigcup_{\lambda \in \Lambda} U_\lambda, \quad \text{and} \quad \bigcap_{\lambda \in \Lambda} U_\lambda \subset U_\mu.$$

Example A.5. $X = \mathbb{R}$, $\Lambda = \mathbb{Z}^+ = \{1, 2, 3, \ldots\}$, $A_n = \big(-\frac{1}{n}, \frac{1}{n}\big)$

$$\bigcup_{n \in \mathbb{Z}^+} A_n = (-1, 1), \quad \bigcap_{n \in \mathbb{Z}^+} A_n = \{0\}.$$

Example A.6. $X = \mathbb{R}^2$, $\Lambda = (0, \infty)$, $A_\lambda = \{(x, y) \in \mathbb{R}^2 \; : \; x^2 + y^2 = \lambda^2\}$

$$\bigcup_{\lambda \in \Lambda} A_\lambda = \mathbb{R}^2 \backslash \{(0,0)\}, \quad \bigcap_{\lambda \in \Lambda} A_\lambda = \emptyset$$

Proposition A.7 (De Morgan's Laws)**.** *Let A_λ be any indexed family of subsets of X. Then*

$$X \setminus (\bigcup_{\lambda \in \Lambda} A_\lambda) = \bigcap_{\lambda \in \Lambda} (X \setminus A_\lambda) \tag{DM1}$$

$$X \setminus (\bigcap_{\lambda \in \Lambda} A_\lambda) = \bigcup_{\lambda \in \Lambda} (X \setminus A_\lambda). \tag{DM2}$$

Proof. The proof of (DM1) follows from elementary operations in logic:

$$\begin{aligned} x \in X \setminus (\bigcup_{\lambda \in \Lambda} A_\lambda) &\Leftrightarrow x \in X \text{ and } x \notin \bigcup_{\lambda \in \Lambda} A_\lambda \\ &\Leftrightarrow x \in X \text{ and } \forall \lambda \in \Lambda \, x \notin A_\lambda \\ &\Leftrightarrow \forall \lambda \in \Lambda \, (x \in X \text{ and } x \notin A_\lambda) \\ &\Leftrightarrow \forall \lambda \in \Lambda \, x \in X \setminus A_\lambda \\ &\Leftrightarrow x \in \bigcap_{\lambda \in \Lambda} X \setminus A_\lambda \end{aligned}$$

The proof of (DM2) is left as an exercise. □

Proposition A.8. *Let $f : X \to Y$ be a function, $\{A_\lambda \; : \; \lambda \in \Lambda\}$ be an indexed family of subsets of X and $\{B_\gamma \; : \; \gamma \in \Gamma\}$ be an indexed family of subsets of Y. Let $S \subset X$, $T \subset Y$. Then*

(a)	$f^{-1}(\bigcup_{\gamma \in \Gamma} B_\gamma) = \bigcup_{\gamma \in \Gamma} f^{-1}(B_\gamma)$,	*(b)*	$f^{-1}(\bigcap_{\gamma \in \Gamma} B_\gamma) = \bigcap_{\gamma \in \Gamma} f^{-1}(B_\gamma)$,
(c)	$f(\bigcup_{\lambda \in \Lambda} A_\lambda) = \bigcup_{\lambda \in \Lambda} f(A_\lambda)$,	*(d)*	$f(\bigcap_{\lambda \in \Lambda} A_\lambda) \subset \bigcap_{\lambda \in \Lambda} f(A_\lambda)$,
(e)	$f^{-1}(Y \setminus T) = f^{-1}(Y) \setminus f^{-1}(T)$,	*(f)*	$f(X \setminus S) \supset f(X) \setminus f(S)$.

Proof. See exercise sheet 1. □

Bibliography

[Arm79] M A Armstrong. *Basic Topology.* McGraw Hill, 1979.

[Bir48] G Birkhoff. *Lattice Theory.* AMS, 1948.

[Bor69] A Borel. *Linear Algebraic Groups.* Benjamin, 1969.

[BS81] S Burris and H P Sankappanavar. *A course in universal algebra.* GTM. Springer-Verlag, 1981. online Millenium Edition available.

[Har77] R. Hartshorne. *Algebraic Geometry.* Graduate Texts in Mathematics 52. Springer, 1977.

[Men62] B Mendelson. *Introduction to topology.* Blackie, 1962.

Index