

1. TOPOLOGICAL SPACES

§1.1. Is The Earth Flat Or Round?

“The world is flat, and the greatest hoax of history is the assertion that it’s round.” I remember hearing this many years ago when I was at university. The claim was made at a lunchtime lecture given by a representative of the Flat Earth Society.



Everyone in my physics class went along to heckle this ‘nut’. But we were stunned by the fact that he appeared to know far more physics than we did and every objection that we raised was answered by the most convincing and authoritative of explanations.

The belief that light travels in straight lines is the illusion, he said. Ships appear to disappear over the

horizon because the light is bending. And the fact that nobody has ever reached the edge of the world is because the closer you come to it, the smaller you become and the more slowly you travel while maintaining the illusion of constant speed. We'd heard of the Theory of Relativity and the lecturer's explanations seemed to be consistent with what very vague understanding we had of that theory.



We began to believe that he might just be right! No doubt this was partly due to the heavy atmosphere in the lecture theatre and to his charisma. As we walked out we felt that he was probably wrong but we were no longer sure we could prove it.

Now of course nobody who has ever walked up and down mountains believes that the world is *quite* flat. Nor is it as perfectly round as a mathematical sphere. It is, after all, slightly flattened at the poles and its surface is somewhat distorted by mountains and valleys. When the flat-earthist says that the world is flat he means that it's *essentially* a flat disk, but one that may be distorted in some way like a piece of rubber that has been stretched and rippled. We round-earthists likewise assert that the

surface of the earth is *essentially* a sphere but concede that it is actually somewhat distorted. The difference between a disk and a sphere is not simply one of shape or curvature. It's 'topological' – it has to do with the way the surface is connected.

The surface of a sphere can be distorted into many different shapes but without tearing it can't become a disk. A disk can be bent to form a hemisphere or even stretched till it becomes a sphere with a little round hole. But only by sewing up the hole (the reverse of tearing) could it become a complete sphere.

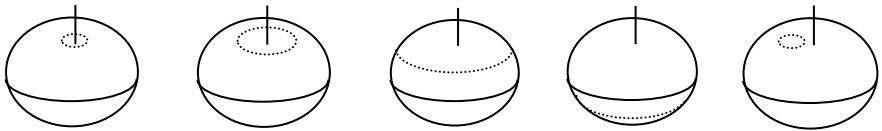
An even simpler experiment to prove that the earth is round is to go on a round the world cruise, counting the days. If you don't adjust your calendar as you cross the International Date Line you would be one day out when you returned, compared with those who stayed at home. If you have read Jules Verne's book *Around the World in 80 Days*, or seen the film, you'll know how important this extra day can be.

If the world is flat there's no way this phenomenon could occur. The fact that it does happen is proof that the world is round. However this proof is, strictly speaking, not a topological one. It relies on the existence of the sun and the assumptions about the movement of the earth relative to the sun. The flat-earthist would have all sorts of explanations about these movements that would appear to explain away the need for an International Date Line.

The difference between a 'flat' earth and a 'round' earth is fundamentally topological and the following is a

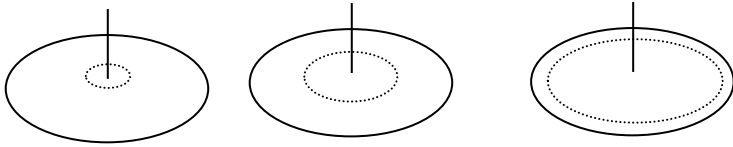
topological proof. Consider the following conceptual experiment. Place a rubber band around the base of the North Pole (assuming it to be an actual pole hammered into the ice, and assuming that global warming has left some ice there). Now imagine that this rubber band is enormously elastic and can be stretched as much as we want. Is it possible, by stretching the rubber band, but without breaking it, and keeping it at all times in contact with the earth’s surface, to free the band from the pole?

The answer depends on which topological model you accept for the earth’s surface. If it’s topologically a sphere, the answer is “yes”. All you have to do is to stretch the band over the surface until it runs right around the equator. Then continue moving it south, keeping it in contact with the surface of the sphere at all times, and let it shrink again as it moves towards the South Pole. Now back to its original size it can be slid back north till it lies right beside the North Pole – no longer enclosing it.

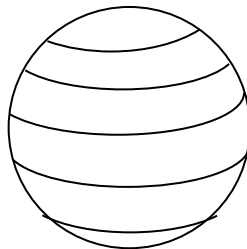


But if the earth is topologically flat then there’s no way it could be freed from the pole. No matter how much the band is stretched, the pole will remain ‘inside’. It’s tempting to say that we could stretch the band till it runs right around the boundary of the disk and then roll it onto

the other side. But remember that if the earth is really flat there is no other side, or at least it doesn't belong to the surface of the earth. So this is a topological way of distinguishing a sphere from a disk.

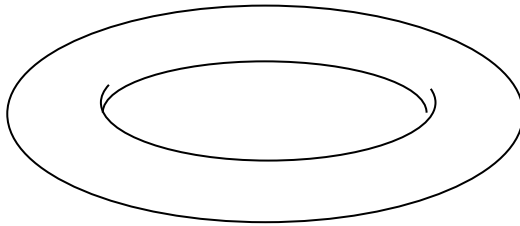


To decide which model fits the earth we just have to carry out this experiment. But there's no need to have an actual rubber band that can be stretched this much. The circles of latitude represent the successive positions of such an elastic band, moving continuously over the surface. Our flat-earthist might question the validity of the circles of latitude and so remain unconvinced. However the aim of this introduction is not to settle the geographic question but to ask topological ones.



On the other hand the earth might be neither a topological disk nor a sphere. Perhaps it's really a doughnut (or to use the more mathematical word, a 'torus'). Let's leave aside the objection that if so then one

part of the world would cast a shadow on the other. This depends on certain assumptions about light and leads us away from topology back into physics. If we lived on the surface of a torus and had no experience of anything above or below the surface, how could we tell that it wasn't the surface of a sphere? After all you can circumnavigate both a sphere and a torus by travelling in what appears to be a straight line.

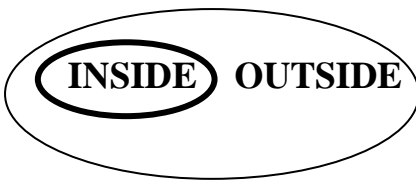


In other words we're asking for a topological difference between a torus and a sphere or a disk. The infinite elastic band experiment works for the sphere but not for the torus (remember that every part of the band must remain in contact with the surface at all times). But it can't distinguish a torus from a disk. This calls for a different conceptual experiment — the Great Wall Experiment.

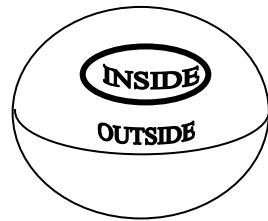
Build a great wall on the surface of the earth so that its two ends meet. This amounts to drawing a closed curve on the surface. Those inhabitants *inside* the wall are safe from the savage hordes *outside* or are they? What if the surface of the earth is a torus (doughnut shape) and the

wall is built around the smaller radius? The enemy is safely on the other side of the wall, until they wake up to the fact that all they have to do is to travel around the larger radius. Our wall is a closed curve that doesn't separate the surface into an inside and an outside. This can happen on a torus but it can't happen on a disk or a sphere.

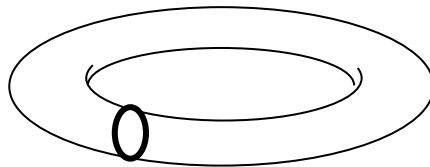
Which side of the wall is inside?



DISK



SPHERE



TORUS

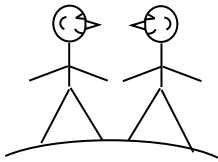
§1.2. If The World Were 2-dimensional

One aspect of topology that will help you to think topologically is the concept of dimension. We live in a 3-dimensional world, but what would life be like if we lived in a 2-dimensional one – a *flat* land? Many years ago

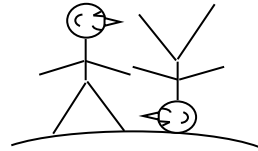
Edwin Abbott wrote a fantasy called *Flatland* in which he explored the difference between 2 and 3 dimensions.

Imagine living on a 1-dimensional earth in a 2-dimensional universe. One possible scenario is that you live on the circumference of a circle, just as we 3-dimensional people live on the surface of a spherical earth.

We'd probably only have one eye. Since it would require a third dimension to turn around we'd only ever



be able to look in one direction. So there'd be two types of people. There'd be the clockwise-facing-



people and the anti-clockwise-facing people. You could meet one of the opposite kind of person face-to-face but if you met one of your own kind you'd only be able to see the back of their head – unless one of you stands on his or her head!

Getting about wouldn't be easy. The only way you could pass somebody would be to climb over them. Being a pedestrian would be a nightmare! In one way, travelling by car would be simpler than in our 3-dimensional world. You wouldn't have to steer. And U-turns would be out of the question. There'd just be one road, around the circumference of the circle, and it would have to be one-way. Two-way traffic would only be possible if there was some way of driving over the top of an oncoming car.

There'd be no intersections. This would greatly simplify driving but the downside would be that if you wanted to duck down to the shops you'd have to reverse all the way home – or else continue driving right around the world!

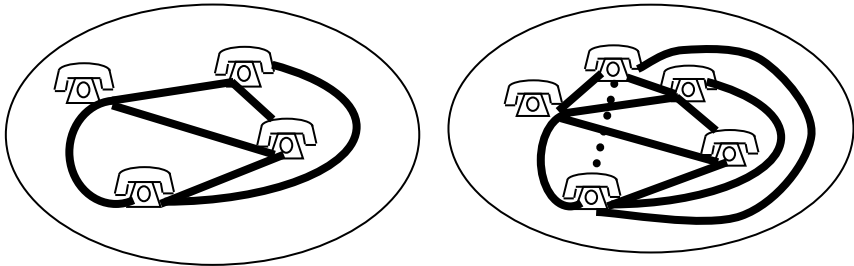
If you met a pedestrian you couldn't drive around him. You'd either have to drive over him or you'd have to stop so he could clamber over the top of the car! Or perhaps you could just drive along behind him. Forget the idea of a tunnel. As soon as the two ends of the tunnels joined it would all fall in!

For the same reason we'd have to evolve differently. The alimentary canal and intestines that enable food to be eaten, and waste products to be eliminated, would cause us to fall apart into two pieces. Things are connected differently in 2-dimensions.

Abbott's *Flatland* was 2-dimensional in a different way. There the world is an entire plane and the Flatlanders who inhabit it are geometric shapes that slide around. This variety of a 2-dimensional world avoids many of the problems that would arise if we lived on the circumference of a circle. You could now have roads with intersections, and 2-way traffic. But there'd be one very fundamental problem with communication. Forget radio waves, or even light. Any electromagnetic wave requires 3 dimensions.

Even the simpler technology of telephone cables would run into topological problems. To avoid crossed wires you'd have to have a wire from each subscriber to

every other subscriber in such a way that the wires don't cross.



You can do this with four subscribers on a flat earth, but not five. Yet in 3-dimensional space there's no problem with any number of subscribers. You just run the wires above or below each other.

§1.3. *n*-Dimensional Space

We can imagine what life would be like in a 2-dimensional world but it's much harder to imagine a 4-dimensional world, or a 17-dimensional one. However, mathematically, it's no harder to discuss spaces of any number of dimensions provided we use an algebraic formulation and don't attempt to visualise it.

Speaking algebraically, a point in \mathbb{R}^3 is simply a triple (x, y, z) of real numbers where x, y and z are called its coordinates. For any n , \mathbb{R}^n is simply the set of n -tuples (x_1, x_2, \dots, x_n) .

The distance between two points $X(x_1, x_2, x_3)$ and $Y(y_1, y_2, y_3)$ in \mathbb{R}^3 is:

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$

and it's very natural to extend this to \mathbb{R}^n by defining the distance between $X(x_1, x_2, \dots, x_n)$ and $Y(y_1, y_2, \dots, y_n)$ to be: $d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$.

It's easy to check that the familiar properties of distance, such as the triangle inequality, extend to n -dimensional distance. For example:

Triangle Inequality: For all points $X, Y, Z \in \mathbb{R}^n$:

$$d(X, Z) \leq d(X, Y) + d(Y, Z).$$

An **open sphere** (interior but no outer surface), with centre O and radius r , is the subset

$$N(O, r) = \{P \mid d(O, P) < r\} \text{ of } \mathbb{R}^3.$$

This definition extends unchanged to \mathbb{R}^n , where we call the 'sphere' an n -sphere. Such n -spheres give us a notion of closeness. If r is small we can think of points in the n -sphere $N(O, r)$ as being 'close' to O . Points in $N(O, s)$, where $s < r$, are even closer. If we're considering some subset S of \mathbb{R}^n we often only need to consider points that are close to a given point O .

We define a **neighbourhood** of a point P to be:

$$N(P, r) = \{Q \in S \mid d(P, Q) < r\}.$$

This will be the intersection of the n -sphere with S .

§1.4. Topological Spaces

You'll probably find this next section a bit hard going. This is because we're attempting to discuss topological spaces in general and it's hard to develop an intuition for them. But most of these notes deal with one particular type of topological space and here intuition can be very useful. Try not to get bogged down in the detail. It's provided here so that you can see that what we will do later is not just vague *hocus pocus*. It has quite firm foundations. But you'll feel a lot happier once we get on to surfaces as 'polygons with identified edges'.

We're leading up to a definition of a topological space – a space of points on which we can do topology. The fundamental concept is that of 'closeness'.

Actually closeness, by itself, is a relative term. How close is 'close'? If we were to mark two points on a balloon one millimetre apart we might reasonably say that they are close. But if we were to blow up the balloon (a hypothetical one, not a real one) so that it was as big as the earth, we'd no longer be justified in saying that they're close. Yet we want to be able to change scale without it affecting topological properties. So we introduce the idea of a *system* of neighbourhoods.

Suppose we have a set X and an element $P \in X$. To give this discussion a geometric flavour we'll call the set X a **space** and the element, P , a **point** in X .

A **system of neighbourhoods** about P is defined to be a collection of subsets, each containing P , with the property that the intersection of any pair of these subsets contains one of these subsets.

We'll call these subsets **neighbourhoods**. So every point is contained in every one of its neighbourhoods and the intersection of any two neighbourhoods must contain some neighbourhood.

We can express this in symbols as follows. Let X be a set and let $P \in X$. A system of neighbourhoods for P is a set of sets, \mathcal{N} , such that:

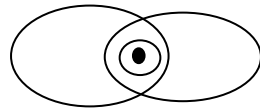
- (1) if $S \in \mathcal{N}$ then $S \subseteq X$;
- (2) if $S \in \mathcal{N}$ then $P \in S$;
- (3) if $S, T \in \mathcal{N}$ then there is some $R \in \mathcal{N}$ such that

$$R \subseteq S \cap T.$$

Property (1) says that neighbourhoods of P are subsets of the whole space X .

Property (2) says that all neighbourhoods of P must contain the point P .

Property (3) says that the intersection of any two neighbourhoods of P must contain a neighbourhood.



We'll call the system of neighbourhoods about the point P as the **system of P -neighbourhoods**.

A **topological space** is a set, X , together with a system of P -neighbourhoods for each $P \in X$. The **topology** is the structure that's imposed on the set and which consists of the systems of neighbourhoods.

Example 1: Let $X = \{1, 2, 3, 4\}$.

Then X is a topological space if the following subsets are considered to be neighbourhoods:

$$\{\{1, 2, 3\}, \{1, 2, 4\}, \{2, 3\}, \{1, 2\}, \{2\}\}.$$

The neighbourhoods of 1 are: $\{1, 2, 3\}$, $\{1, 2, 4\}$ and $\{1, 2\}$. The intersection of any two of these contains a neighbourhood of 1.

They are all neighbourhoods of 2 and the intersection of any two of these contains a neighbourhood of 1.

The neighbourhoods of 3 are $\{1, 2, 3\}$ and $\{2, 3\}$ and the only neighbourhood of 4 is $\{1, 2, 4\}$. In each case the intersection of any two of these contains one of them. (In the case of neighbourhoods of 4 the only possible intersection is $\{1, 2, 4\}$ intersected with itself.)

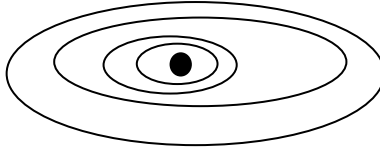
It's possible to put different topologies on the same set.

Example 2: $\{\{1\}, \{2\}, \{3\}, \{4\}\}$ and $\{\{1, 2, 3, 4\}\}$ are two other topologies on $\{1, 2, 3, 4\}$.

This finite example doesn't appear to be very geometric. In fact finite topological spaces aren't very interesting. A more natural example is the plane as a topological space with nested systems of neighbourhoods.

A **nested system of P-neighbourhoods** is one where the P-neighbourhoods form a descending chain:

$$N_1(x) \supseteq N_2(x) \supseteq \dots$$



Example 3: Let $X = \mathbb{R}^2$ and for each $x \in \mathbb{R}^2$ and for each $r > 0$ let $N_r(x) = \{y \mid |y - x| < r\}$. Geometrically each of the sets $N_r(x)$ is the interior of a circle with centre x and radius r . This is the normal topology on the plane.

Where we have a nested system of P-neighbourhoods we can say that the more P-neighbourhoods that a point belongs to, the closer it is to P. But these neighbourhoods aren't necessarily tagged with distances, so we mightn't be able to say how close a point is to P. We can say that one point, Q, is closer to P than another point, R, but we can't quantify how much closer.

Even when the neighbourhoods come from distances, as in Example 3, we throw away the actual distances and retain only the neighbourhoods themselves. Why do we do this? Because we don't want to focus on properties where actual distances matter. We want to study properties of connectedness which, while they

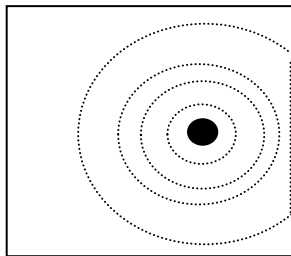
involve the notion of closeness, don't depend on a numerical measure of closeness. These properties will therefore remain fixed under transformations of scale such as enlargements or contractions.

We can extend Example 3 very easily to make n -dimensional space into a topological space. We do this by using distance to produce neighbourhoods, as in example 3, and then throwing away the distances and just retaining the systems of neighbourhoods. We can also make any subset of \mathbb{R}^n into a topological space by intersecting the n -dimensional spheres with the subset to get the neighbourhoods.

Example 4: Let X be the unit square
 $\{(x, y) \mid 0 \leq x, y \leq 1\}$.

For the neighbourhoods about a point $\mathbf{u} \in X$ take the sets
 $N_r(\mathbf{u}) \cap X$ where $N_r(\mathbf{u}) = \{\mathbf{v} \mid |\mathbf{v} - \mathbf{u}| < r\}$.

Here most of the neighbourhoods are open disks. But where the radius is too large for the open disk to fit inside X we take the neighbourhood to be the part of that disk that lies inside or on the square.



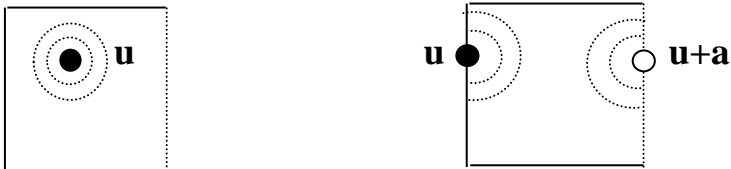
Example 5: Let X be the unit square

$$\{(x, y) \mid 0 \leq x < 1, 0 \leq y \leq 1\}.$$

We've removed the right boundary from the previous example. The neighbourhoods of points $\mathbf{u} = (x, y)$ where $x < 1$ are almost exactly as before: they are the intersections of circles about the point with the square. It's just that the right-hand boundary is omitted. But for a point \mathbf{u} on the left boundary we define the neighbourhoods as

$$(N_r(\mathbf{u}) \cup N_r(\mathbf{u} + \mathbf{a})) \cap X \text{ where } \mathbf{a} = (1, 0).$$

The consequence of this topology on the square is that points close to the right-hand end are close to the corresponding point on the left-hand end. The point $(0.99, 0.3)$ is very close to $(1, 0.3)$ which lies just outside of X and so, in the above topology, it is very close to the corresponding point $(0, 0.3)$ on the left-hand boundary.

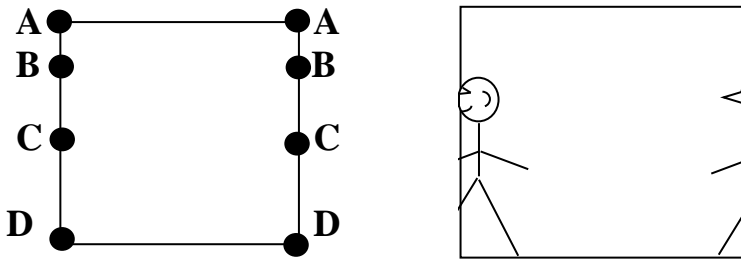


This modified topology might seem very artificial and hard to understand intuitively. But it isn't! We can make it seem much more natural if we reinstate the right-hand boundary, taking X again to be

$$\{(x, y) \mid 0 \leq x, y \leq 1\}$$

but *identifying* the points on the right-hand boundary with the corresponding points on the left-hand one.

Identifying points means treating them as the same point. So, for example, we consider the two points $(0, 0.3)$ and $(1, 0.3)$ as the *same* point. This makes the topology that we described above seem very natural. The point $(0.99, 0.3)$ is close to $(1, 0.3)$, in the ordinary sense of closeness, and so is close to $(0, 0.3)$ since $(0, 0.3)$ and $(1, 0.3)$ are being considered to be the same point.



If you lived on this square, with the left and right boundaries identified you'd have the experience of moving towards the left-hand boundary and, as you reached it you'd find that you'd suddenly been transported to the corresponding point at the right.

This phenomenon has become commonplace in the world of many computer games. It's called 'wraparound'. A little figure runs off the screen at the left-hand end and suddenly reappears at the right-hand end, still moving to the left.

Another way of thinking about this space is to roll up the square and glue the left-hand end to the right hand end. We can then see that we've constructed a cylinder.

Why go to the trouble of dressing it up with all this stuff about identifying points? There are two reasons:

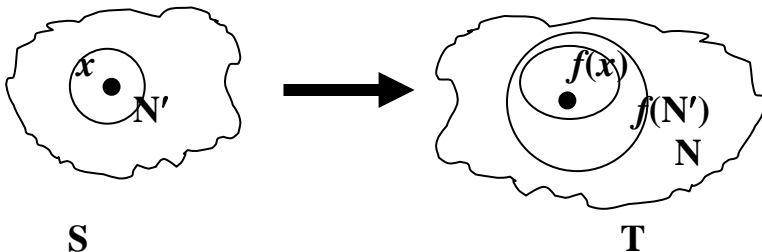
(1) It's usually easier to deal with an unwrapped space, using identified edges. We can more easily draw on a flat surface than to make a 3-dimensional drawing.

(2) There are many spaces that can be described quite easily as polygons with certain edges identified that don't exist in 3-dimensional space. The only way we can draw pictures of them is using identified edges.

§1.5. Continuous Functions and Homeomorphisms

Every mathematical structure (group, vector space etc) is a set with some form of structure upon it. Functions that preserve that structure play a pivotal role in the theory of that type of structure. In the case of topological spaces we have continuous functions and homeomorphisms.

If S, T are topological spaces, a function $F:S \rightarrow T$ is **continuous at the point** $P \in S$ if every neighbourhood, N , of $F(P)$ contains the image of some neighbourhood of P . This is a generalisation of the concept of a continuous function on the set of real numbers.



In elementary calculus we say that a function $f: \mathbb{R} \rightarrow \mathbb{R}$ is continuous at $x = a$ if for all $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $|x - a| < \delta$ we have $|f(x) - f(a)| < \varepsilon$. Here we're using the usual topology on \mathbb{R} where the neighbourhoods of a are the open intervals $(a - r, a + r)$ for various $r > 0$.

The above ε - δ definition can be rephrased by saying that for all $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $x \in N_\delta(a)$ then $f(x) \in N_\varepsilon(f(a))$. In other words, every neighbourhood of $f(a)$ contains the image of some neighbourhood of a .

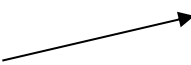


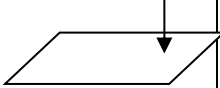

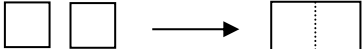
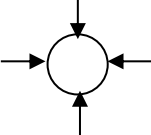
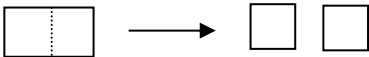
A function is **continuous** if it's continuous at every point. Examples of continuous functions on the real line are polynomials and functions like $\sin x$, e^x etc. In fact nearly all naturally occurring functions of a real variable are continuous. The function $f(x) = [x]$, the integer part of x , is discontinuous at integer values of x .

We can easily construct a function from \mathbb{R} to \mathbb{R} which is discontinuous at some point by artificially constructing a discontinuity, such as: $f(x) = \begin{cases} x & \text{if } x \neq 0 \\ 2 & \text{if } x = 0 \end{cases}$.

A very rough geometric description of continuity is to say that under a continuous function “points which are close together stay close together”. But closeness is a relative thing and under the function $f(x) = 1000x$ you'd be hard put to say that $f(1)$ and $f(1.01)$ are close. A more accurate description is to say that under a continuous function “you can make the images as close as you like by making the original points sufficiently close”. For the

function $f(x) = 1000x$ on the real line you can make $f(x)$ within a distance 0.01 of $f(1) = 1000$ by making x sufficiently close to 1. In fact you'd achieve this if x was within a distance of 0.00001 of 1.

Geometric Examples of Continuous Functions:

<p>Translations</p> 	<p>Twists and Bends</p> 
<p>Reflections</p> 	<p>Projections</p> 
<p>Enlargements</p> 	<p>Joining</p> 
<p>Contractions</p> 	<p>But tearing is not continuous</p> 

Tearing is not a continuous operation. Points that are close together before the tearing are no longer close together, if the two pieces are pulled apart after the tear.

The inverse of tearing is joining and in this direction it is continuous. But for many purposes we need functions that are continuous in *both* directions. And, of course, before we can talk about *both* directions we need to have a function that has an inverse — namely one that is 1-1 and onto.

A function $f:S \rightarrow T$ is a **homeomorphism** if it is 1-1 and onto and if both f and $f^{-1}:T \rightarrow S$ are continuous. We use the concept of a homeomorphism to define the concept of topologically equivalent (or homeomorphic) topological spaces. Topological spaces S and T are **homeomorphic** (we write $\mathbf{S} \approx \mathbf{T}$) if there exists a homeomorphism from S to T .

WARNING: Don't confuse 'homeomorphism' of a topological space with 'homomorphism' in group theory or vector space theory. Quite apart from the fact that they are spelt slightly differently and relate to different types of mathematical structures, a homeomorphism must be 1-1 and onto while homomorphisms need not be. The true counterpart of a homeomorphism in topology is an *isomorphism* of groups or vector spaces. Homeomorphic topological spaces can be considered to be the same, as can isomorphic groups or isomorphic vector spaces.

In \mathbb{R}^3 with the usual topology, translations, reflections, rotations, enlargements, contractions and twisting and bending are homeomorphisms. Projections are not, because they don't have inverses. Tearing and joining have inverses but, because tearing isn't continuous, neither of them are homeomorphisms.

Theorem 1: The relation of being homeomorphic is an equivalence relation.

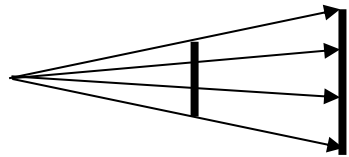
Proof: *Reflexive:* The identity function $1:S \rightarrow S$ is clearly a homeomorphism.

Symmetric: If $f: S \rightarrow T$ is a homeomorphism then so is $f^{-1}: T \rightarrow S$.

Transitive: The composition of continuous functions is continuous and so the same is true of homeomorphisms. So if $f: S \rightarrow T$ and $g: T \rightarrow U$ are homeomorphisms then so is $fg:S \rightarrow U$ where $(fg)(s) = g(f(s))$.

We omit the proofs of these assertions. 🙌😊

Example 6: All finite open intervals are homeomorphic to one another.



We can prove this using neighbourhoods, but we can see this intuitively by imagining a short finite interval being stretched continuously until it fits the larger. Or we can imagine one interval being projected onto another.

However a more rigorous explanation comes from constructing the function $f(x) = \frac{x - a}{b - a}$ which maps the open interval (a, b) onto the open interval $(0, 1)$. Being a linear map it's continuous. Moreover its inverse exists: $f^{-1}(x) = (b - a)x + a$, and this is likewise continuous. So all finite open intervals are homeomorphic to $(0, 1)$ and hence, by transitivity, they are homeomorphic to one another.

Example 7: The real line \mathbb{R} is homeomorphic to every finite open interval.

We might try to imagine a finite open interval being stretched to infinity in both directions, but this might be stretching our geometric intuition a little too far! Instead we can construct an explicit homeomorphism from the open interval $(-\pi/2, \pi/2)$ to \mathbb{R} and this settles the question. For example the function $f(x) = \tan x$ is such a homeomorphism. [Note that $f(x) = \tan x$ is not continuous on \mathbb{R} , but it is on the interval $(-\pi/2, \pi/2)$.]

There are three levels at which we can prove that two subsets of \mathbb{R}^3 are homeomorphic.

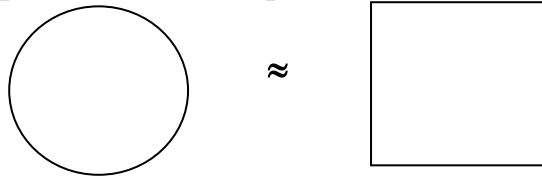
Level 1: We can use our geometric intuition to imagine one subset being continuously deformed into the other. (Strictly speaking this is a little stronger than homeomorphism in that the converse doesn't hold. If you can't continuously deform one set into another, in \mathbb{R}^3 , this doesn't prove that they are not homeomorphic. We'll explain this stronger concept shortly. (Watch out for the word 'homotopic'.)

Level 2: We can describe geometrically an explicit function from one set to the other and rely on our geometric intuition to convince ourselves that it's a homeomorphism.

Level 3: We can describe a suitable 1-1 and onto map and *prove* that it's continuous in both directions.

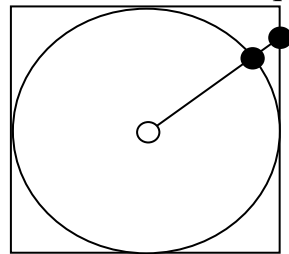
Obviously the ideal would be to operate at level 3 at all times, since relying on geometric intuition is less rigorous. The problem is that it can get us bogged down in very technical and fiddly detail.

Example 8: A closed disk (interior plus boundary) is homeomorphic to a closed square.



Level 1: One can imagine a sufficiently elastic circle being stretched in such a way so that it becomes a square.

Level 2: Place a unit disk, centred at O , the origin, inside the square bounded by the lines $x = -1$, $x = 1$, $y = -1$ and $y = 1$. Given a point P on the disk, but not the origin, represented by the vector \mathbf{v} and project the ray OP until it meets the square. If the distance from the origin of the point of intersection of the ray and the square is k define $f(\mathbf{v}) = k\mathbf{v}$.



Finally define $f(\mathbf{0}) = \mathbf{0}$. It is intuitively obvious that this function is continuous. The inverse is $f^{-1}(\mathbf{v}) = (1/k)\mathbf{v}$, which, intuitively, is also continuous.

But note that the k here is not constant. It depends on the direction of the point from the origin. We're relying on the fact that small changes in \mathbf{v} will give rise to small changes in the angle and hence small changes to k .

Level 3: Take the disk with radius 2 and square with sides 2 units, both centred on the origin.

For $0 \leq \theta < 2\pi$ let P_θ be the point P:

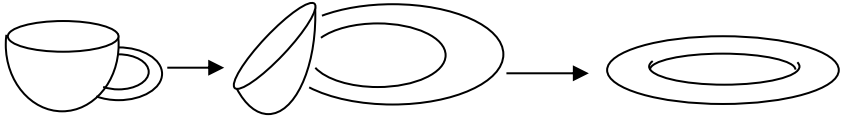
$$\begin{cases} (1, \tan \theta) & \text{if } 0 \leq \theta < \pi/4 \\ (\tan(\pi/2 - \theta), 1) & \text{if } \pi/4 \leq \theta < 3\pi/4 \\ (-1, \tan \theta) & \text{if } 3\pi/4 \leq \theta < 5\pi/4 \\ (\tan(3\pi/2 - \theta), -1) & \text{if } 5\pi/4 \leq \theta < 7\pi/4 \\ (1, \tan \theta) & \text{if } 7\pi/4 \leq \theta < 2\pi. \end{cases}$$

As θ moves from 0 to 2π the point P_θ moves around the square. The fact that $\tan x$ is continuous on $[-\pi/4, \pi/4]$, and a check on the values $\theta = 0, \pi/4, 3\pi/4, 5\pi/4$ and $7\pi/4$ shows that $\theta \rightarrow P_\theta$ is continuous.

Example 9: A teacup is homeomorphic to a doughnut.

Proof: You can imagine a teacup made of elastic plasticine being continuously deformed into a doughnut shape. First the bowl is flattened out. Then the handle is enlarged while what was the bowl is contracted. Finally the object becomes all handle as the material that once formed the cup is incorporated into the handle. The new

object is suitably rounded and finally takes on the shape of a doughnut.



That's a proof at level 1. To go to level 2 we'd need to find an equation that would define the tea cup and then construct an explicit homeomorphism from the tea cup to a doughnut shape. While you should be convinced that this could be done, you should also be convinced that it would be incredibly messy. To go to level 3 and prove rigorously that this highly complicated function is a homeomorphism would be a nightmare!

In what follows you'll hardly ever confront a situation so simple that a level 3 proof of homeomorphism would be appropriate. In most cases you should attempt a level 2 proof, that is, explicitly describe a homeomorphism but omit the proof that it is. In many cases you'll need to describe in words and pictures a continuous deformation from one to the other. In fact often you'll be expected only to give a level 0 proof (that is, no proof at all). In other words we'll simply ask whether or not two spaces are homeomorphic.

But how do you prove that two spaces are *not* homeomorphic? How can you, for example, prove that a

sphere *isn't* homeomorphic to a torus? You could imagine moulding a sphere made of plasticine until it became fairly flat – a sort of fat disk – and then pushing down in the middle of the top and up in the middle of the bottom. But at some stage you'd need to make the hole if you wanted to end up with a doughnut shape. This would involve a discontinuity.

This might seem to you fairly convincing, but we can actually do better. We can construct properties that are preserved by homeomorphisms. These are called **topological invariants**. If one of our spaces has such a property and the other one doesn't then quite clearly they aren't homeomorphic.

That's what we did with the flat earth and the round earth. The property that one point can be removed (the North Pole), and that a closed curve surrounding this point can be continuously deformed so that it no longer surrounds the point, is a topological invariant because it can be expressed in terms of continuity. But the surface of a sphere has this property while the surface of a disk does not. Therefore these spaces are not homeomorphic.

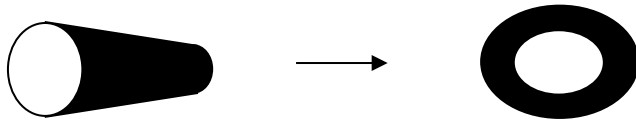
Another invariant that we used is the property that a closed curve always separates the surface into two distinct regions, where there is no path from a point in one region to a point in the other without crossing this closed curve.

A disk and a sphere have this property while the surface of a torus does not. But this property is clearly

preserved by homeomorphisms since it's expressible entirely in terms of continuity. So the torus is not topologically equivalent (homeomorphic) to either the surface of a sphere or a disk.

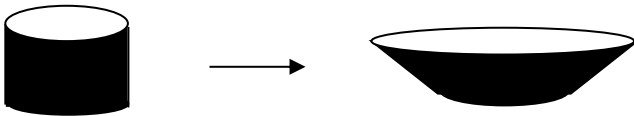
Example 10: An annulus is homeomorphic to a cylinder that's open at both ends.

You can imagine the bottom of the cylinder to be continuously enlarged so that the surface becomes the surface of a truncated cone (with the top sliced off). You can then project this truncated cone down onto a flat surface (yet another homeomorphism – 1-1 projections are homeomorphisms) and so end up with an annulus.

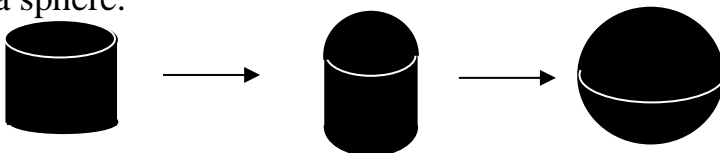


Example 11:

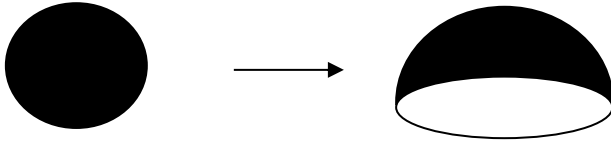
(i) A cylinder that's open at the top but closed at the bottom is homeomorphic to a disk.



(b) A cylinder that's closed at both ends is homeomorphic to a sphere.



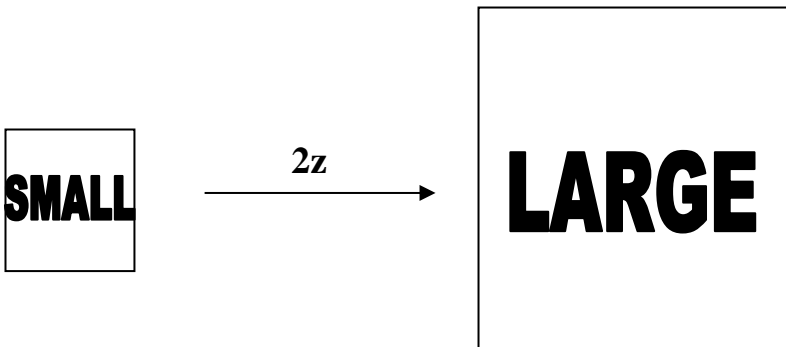
(c) A disk is homeomorphic to a hemisphere.



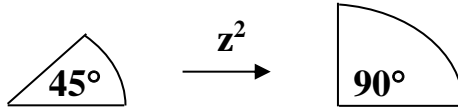
§1.6. Topological Invariants

A topological invariant is something that's preserved by homeomorphisms. It might be a property, or a number, or even a polynomial or a group. Whatever it is, if two topological spaces have different values of the topological invariant that proves that they aren't homeomorphic. (But beware – if two spaces have the same value of some invariant this does not prove that they are homeomorphic.)

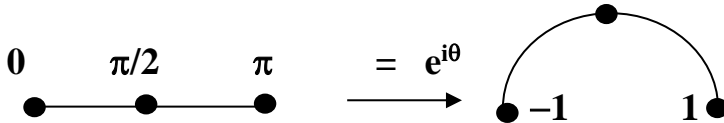
Size: isn't preserved by homeomorphisms. For example the homeomorphism $f(z) = 2z$ takes a small square in the complex plane to a bigger one.



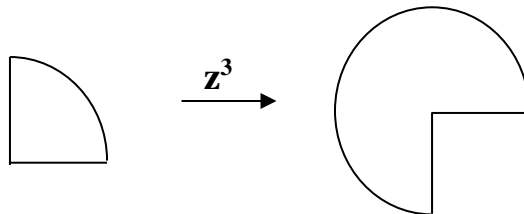
Angles: aren't preserved by homeomorphisms. For example the homeomorphism $f(z) = z^2$ doubles angles in the complex plane.



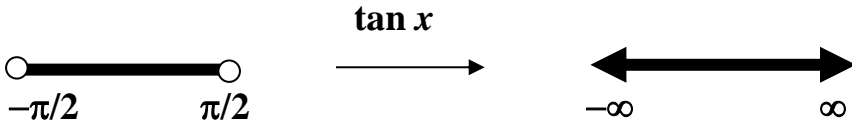
Collinearity: isn't preserved by homeomorphisms. Homeomorphisms can take straight lines into curved ones. For example the map $f(\theta) = e^{i\theta}$ is a homeomorphism from the (straight) interval $[0, \pi]$ to the (curved) upper half of the unit circle in the complex plane.



Convexity: isn't preserved by homeomorphisms. For example the homeomorphism $f(z) = z^3$ maps the first quadrant of the unit circle (a convex set) to three quarters of the unit circle (non convex).



Bounded-ness: isn't preserved by homeomorphisms. For example the map $f(x) = \tan x$ takes the bounded open interval $(-\pi/2, \pi/2)$ to the whole real line (unbounded).



The concepts of size, angle, collinearity, convexity and boundedness are important concepts in geometry but they have no significance in topology. We have to look further afield to find topological invariants.

Certainly none of the above properties have any relevance to topology, but you may be left with the impression a property that can be changed by a homeomorphism is outside the realm of topology. However there's a second level to topology, *homotopy*, and invariants under homotopies, rather than homeomorphisms, are often important. For example a knotted string that's joined at both ends, is homeomorphic to an un-knotted one so at the level of homeomorphisms knotted-ness is irrelevant. However at the level of homotopy it is quite important. Soon we'll be investigating knots in some depth, but now we need to familiarise ourselves with the concept of homotopy.

§1.7. Paths

We all know what is meant by a 'path' (or 'curve') from one point to another in the plane. It's something you draw, starting at the first point and ending up at the other. Of course it has to be continuous – there must be no

breaks – we must keep our pencil on the paper the whole time.

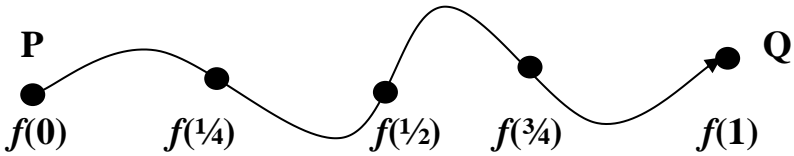
But if we're going to become topologists, even amateur ones, we need to do better than that. We need a more precise definition.

Think of a real path that goes from point A to point B in a park. It takes a certain length of time to traverse it. If you made a chalk mark after every minute of your walk, you'd have recorded a description of your walk. Instead of using chalk you could have used your GPS navigation tool to record the latitude and longitude of your position at one minute intervals.

Someone reading your data could reconstruct your walk. They'd not only know *where* you walked, and in which direction. They could also work out your speed. You mightn't have walked at a uniform speed the whole way. You might have started out at a fair pace, stopped at a park bench for a few minutes to chat to someone, walked on, but more slowly because you were still talking while walking with your new friend and so on. You might even have backtracked at some point because you'd dropped something.

Someone else might have walked the same path but their pattern of speed might have been completely different. Of course their starting and finishing times might be different too, but we can standardise that by scaling time so that time $t = 0$ is the time at the start and $t = 1$ is the time at the end. This is essentially what we mean by a 'path' in topology.

A **path** from P to Q in a topological space X is a continuous function $f: [0, 1] \rightarrow X$, where the closed interval has the usual topology, such that $f(0) = P$ and $f(1) = Q$. The points P and Q are called the **endpoints** of the path.



Example 12: Consider the functions $f(t) = t + 1$ and $g(t) = t^2 + 1$. Both of these are paths from 1 to 2 on the real line. While the set of points traversed, as t goes from 0 to 1, is the interval $[1, 2]$ in each case these functions are *different paths*. When we use the word ‘path’ in ordinary language we mean the concrete strip along which we walk. But in topology it means rather more. The path includes information about where you are at such and such a time, which direction you were going (forward or backward) and how fast you were going at any given moment. A more appropriate word might have been ‘walk’ because this can convey the dynamic nature of the thing, as something happening in time.

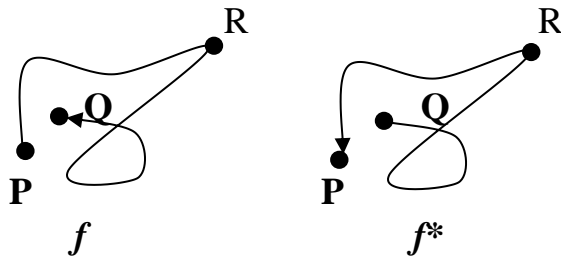
In this example the path $f(t) = t + 1$ represents the journey of someone moving along the interval at a uniform speed while $g(t) = t^2 + 1$ has someone moving slowly at first and then speeding up. In both cases they

start and finish together but at half-time the f path is at the point 1.5 while the g path has only reached 1.25.

Another path is $h(t) = 2 - t$. This represents someone walking back from the point 2 to the point 1 at a constant speed. In all three cases the set of points reached is the same – the interval $[1, 2]$. But all three paths are different.

The **inverse** of a path f from P to Q is the path f^* from Q to P defined by $f^*(t) = f(1 - t)$. It's like turning time backwards. If you recorded a path by taking a video then running the video backwards at the same speed would be a record of the inverse path.

Note that we don't write the inverse as f^{-1} . This is because it isn't the inverse function in the usual sense. While f is a map from the unit interval to the topological space f^* is not a map from the space back to the unit interval. However f^* is an inverse in a different sense.



$$P = f(0) = f^*(1)$$

$$Q = f(1) = f^*(0)$$

$$R = f(1/4) = f^*(3/4)$$

NOTE: R isn't equal to $f(1/4)$ because it's a quarter of the way along the curve (it isn't) but because we choose it to be there. In terms of the moving point analogy the fact that R appears to be more like half-way in terms of distance means that in defining f we've chosen to have the point move faster during the first quarter of its journey.

Example 13: The continuous function $f(t) = e^{\pi i t}$ is a path from the point 1 on the real axis to the point -1 . The movement is at a uniform speed around the top half of the unit circle.

The inverse is $f^*(t) = e^{\pi i(1-t)}$.

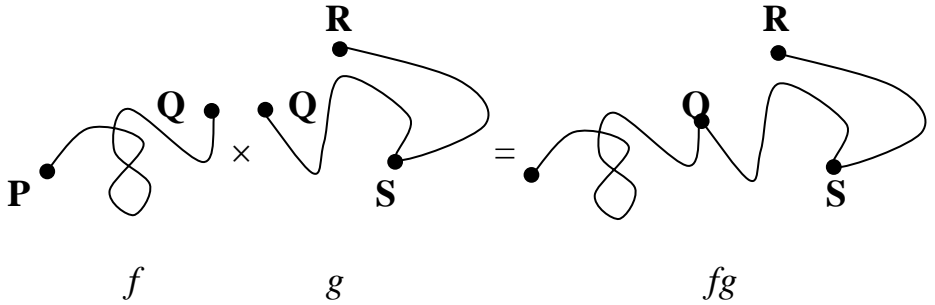


The **product** of two paths f and g , where $f(1) = g(0)$, is the path fg defined by:

$$(fg)(t) = \begin{cases} f(2t) & \text{if } 0 \leq t \leq 1/2 \\ g(2t - 1) & \text{if } 1/2 \leq t \leq 1 \end{cases}$$

Notice that it's not the same as the product in the usual senses of $f(t)g(t)$ or composition $g(f(t))$.

Travelling the path fg you'd travel the path f at twice the speed, so that at time $t = 1/2$ you've reached the endpoint of f . Why twice the speed? That's so you have time left over to travel the g path. This you also do at twice the speed so that at time $t = 1$ you've reached the endpoint of g .

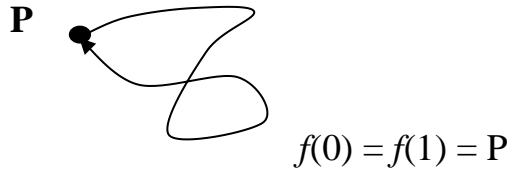


$$\begin{aligned}
 P &= f(0) = (fg)(0) \\
 Q &= f(1) = g(0) = (fg)(1/2) \\
 R &= g(1) = (fg)(1) \\
 S &= g(1/2) = (fg)(3/4)
 \end{aligned}$$

The **identity path** from a point P to itself is the constant path $1: [0, 1] \rightarrow X$ where $1(t) = P$ for all t .

Now you might think that $ff^* = 1$ but it isn't quite as simple as that. The path ff^* doesn't stay fixed at P. It wanders off, just like f , only at twice the speed, and then returns to P. But ff^* and 1 are equivalent in a certain sense. We'll have to wait until we've discussed the concept of 'homotopy' before we can understand this.

A path is a **closed path** if the endpoints are equal.

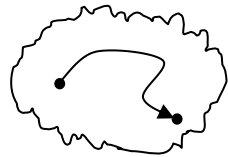


Example 14: If f is the path $f(t) = e^{\pi i t}$ and g is the path $g(t) = 2t - 1$ then fg is the closed path:

$$(fg)(t) = \begin{cases} e^{2\pi i t} & \text{if } 0 \leq t \leq \frac{1}{2} \\ 4t - 3 & \text{if } \frac{1}{2} \leq t \leq 1 \end{cases}$$

[Note that $4t - 3 = 2(2t - 1) - 1$.]

A topological space X is **pathwise connected** if every two points are endpoints of some path in X . Note that this is a topological property. A pathwise connected space cannot be homeomorphic to one which is not.



PATHWISE CONNECTED



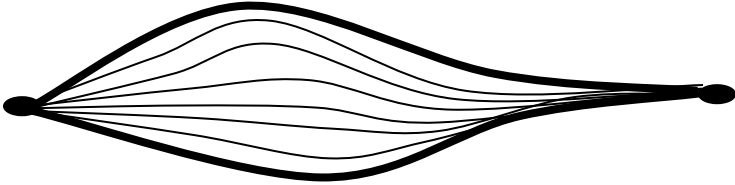
NOT PATHWISE CONNECTED

The concept of path-wise connectedness is more or less the ordinary sense of ‘connected’. There is another sort of connectedness that we’ll discuss in the next section.

§1.8. Homotopy

Having discussed connectedness of points we move on to discuss connectedness of paths. What do we mean when we say that one path with endpoints A, B can be continuously deformed into another? We have the image of a piece of elastic fixed at these endpoints and running along the first path. Then, as we continuously deform it we imagine the elastic being gradually moved,

still anchored at both ends, until it runs along the second path. That's a fairly accurate picture except that we have to remember that the path is not just the set of points but sets of points and associated times.



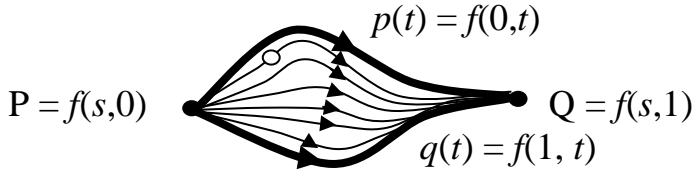
Let S denote the unit square $[0, 1] \times [0, 1]$ with the usual topology. A **homotopy** on a topological space X is a continuous function $f: S \rightarrow X$ such that $f(s, 0)$ and $f(s, 1)$ are each constant.

Here the set $[0, 1] \times [0, 1]$ is just the unit square, so essentially we have a function of two real variables s, t which each take values between 0 and 1, with $f(s, t)$ being a point in our space. Both the variables s and t represent time, but in different ways. The first variable, s , represents time as the path is being continuously deformed. The second variable, t , is the one used to describe a path.

When $s = 0$ we have the continuous function $f(0, t)$. This is in fact a path and it's the **initial path**. When $s = 1$ we have the continuous function $f(1, t)$ which is the **final path**. For a value of s between 0 and 1 the function $f(s, t)$, as a function of t , is the path that occurs at time s . As s runs from 0 to 1 these paths represent the successive positions of the deforming path.

Why should $f(s, 0)$ and $f(s, 1)$ be constant? These represent the endpoints of a typical intermediate path

throughout the deforming process. Since we're anchoring the ends, and merely moving the path between them, we want all of the intermediate paths to have the same two endpoints.



This represents a homotopy that takes the path p to the path q , both connecting the same two points P and Q . Think of the path p being continuously deformed into the path q with the other six paths from P to Q representing some of the intermediate paths. They might, for example, represent the paths:

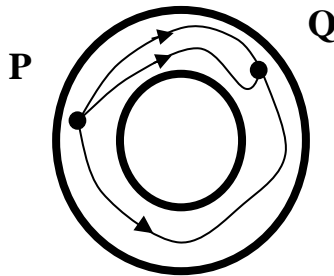
$f(0.2, t), f(0.4, t), f(0.5, t), f(0.6, t), f(0.75, t)$ and $f(0.8, t)$

respectively. So the parameter s represents time in the sense of one path being deformed into the other, over time. But, because each path itself represents a movement in time, the movement of a point from P to Q , we need the second parameter t .

The point indicated by \circ in the above diagram might represent $f(0.2, 0.4)$. It lies on the path $f(0.2, t)$, which we're taking to be the path that p has become at time $s = 0.2$. This path represents the motion of a point moving from P to Q and we might be taking the indicated point to be the position of this moving point at time $t = 0.4$.

Two paths p and q from points P to Q in a topological space are **homotopic** if there exists a homotopy f such that $f(s, 0) = P$ and $f(s, 1) = Q$ for all s and $f(0, t) = p(t)$ and $f(1, t) = q(t)$ for all values of t .

Example 15: Let X be an annulus in the x - y plane and let P, Q be any two points in X . A path from P to Q can be deformed into many others. These will all be homotopic with the original path. But there will be other paths that join P to Q by going the other way around the central hole. There's no way that such a path can be continuously deformed into the original one. It would mean, at some stage, crossing over the hole which is not allowed because the inside of the hole lies outside of the space.



Clearly the top two paths are homotopic, but neither is homotopic to the bottom one if the space is just the annulus. There's no way one of the top paths can be continuously deformed into the bottom one without moving out of the annular space.

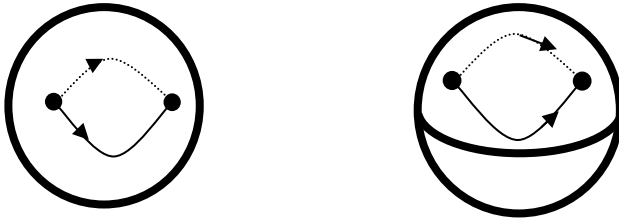
You can see the difference between the notions of being 'homeomorphic' and 'homotopic' illustrated here.

Both the upper path and the lower path are *homeomorphic* to each other. But because of the hole they're not *homotopic*.

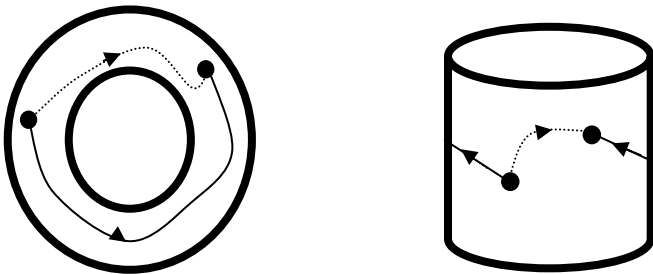
If we say that two spaces are homotopic we're not just talking about the spaces themselves. It depends on the space in which they lie. The property of being homeomorphic however, is just a property of the spaces in isolation.

A topological space is **simply connected** if for all points P and Q in the space any two paths from P to Q are homotopic.

Example 16:



The disk and the sphere are simply connected.



The annulus and the cylinder are not simply connected.

But all four of these topological spaces are pathwise connected.

Pathwise connectedness and simple connectedness are topological invariants.

EXERCISES FOR CHAPTER 1

Exercise 1: Sort the digits 0-9, represented as follows, into homeomorphism classes, treating them as 1-dimensional.

0 1 2 3 4 5 6 7 8 9

Exercise 2: Sort the digits 0-9, represented as follows, into homeomorphism classes, treating them as 2-dimensional shapes:

0 1 2 3 4 5 6 7 8 9

Exercise 3:

Sort the following 2-dimensional figures into homeomorphism classes:

A B C D E F

Exercise 4:

(a) Find an explicit homeomorphism (i.e. write down a formula for the function) from the open interval $(1, 3)$ to the set of all real numbers.

(b) Find an explicit homeomorphism from \mathbb{R} to \mathbb{R}^+ , the set of positive real numbers.

(c) By composing the homeomorphisms obtained in (a) and (b), or otherwise, find an explicit homeomorphism from $(1, 3)$ to \mathbb{R}^+ .

Exercise 5:

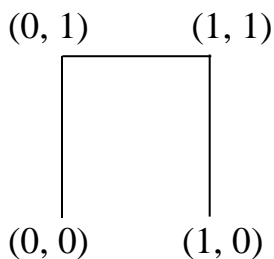
Write down a homeomorphism from \mathbb{R}^+ , the set of positive reals, to the open interval $(-\pi/2, \pi/2)$.

Exercise 6:

Write down a path in \mathbb{R}^2 (as a function $p(t)$ from $[0, 1]$ to \mathbb{R}^2) which runs along the x -axis from $(0, 0)$ to $(1, 0)$ and then around the unit circle to $(0, 1)$.

Exercise 7:

(a) Find a path $p(t)$ whose image forms the 3 sides of the unit square indicated:



(b) Find a homotopy from $p(t)$ to $q(t) = (t, 0)$ whose trace is the fourth side of the square.

SOLUTIONS FOR CHAPTER 1

Exercise 1:

1, 2, 3, 5, 7 are homeomorphic to _____

0 is homeomorphic to 


6, 9 are homeomorphic to 

4 is homeomorphic to 

8 is homeomorphic to 

Exercise 2:


1, 2, 3, 5, 7 are homeomorphic to  (disk with no holes)

0, 4, 6, 9 are homeomorphic to  (disk with one hole)

8 is homeomorphic to  (disk with two holes)

Exercise 3:

C, E, F are homeomorphic to  (disk).

A, D are homeomorphic to  (disk with one hole).

B is homeomorphic to  (disk with two holes).

- Exercise 4:** (a) $f(x) = \tan(\pi x/2)$ is probably the simplest.
 (b) $f(x) = e^x$ is the simplest.
 (c) $f(x) = e^{\tan(\pi x/2)}$.

Exercise 5: $f(x) = \tan^{-1}(\log x)$

Exercise 6:

$$f(t) = \begin{cases} (0, 2t) & \text{if } 0 \leq t \leq 1/2 \\ (\cos\pi(t - 1/2), \sin\pi(t - 1/2)) & \text{if } 1/2 \leq t \leq 1 \end{cases}$$

Exercise 7: (a) $p(t) = \begin{cases} (0, 3t) & \text{for } 0 \leq t \leq 1/3 \\ (3t - 1, 1) & \text{for } 1/3 \leq t \leq 2/3 \\ (1, 3 - 3t) & \text{for } 2/3 \leq t \leq 1 \end{cases}$

(b) $h(s, t) = sq(t) + (1 - s)p(t)$

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

