Preface

This book grew out of courses given at the University of Wales Swansea to second- and third-year undergraduates. It is designed to provide enough material for a one-year course and splits naturally into a preliminary topology course (Chapters 1–6) and a follow-on course in algebraic topology (Chapters 7–11).

It is often said that topology is a subject which is poorly served for textbooks, and when preparing the lecture courses I found no book that was both accessible to our undergraduates and relevant to current research in the field. This book is an attempt to fill that gap. It is generally accepted that a one-year course on topology is not long enough to take a student to a level where she or he can begin to do research, but I have tried to achieve that as nearly as possible. By omitting some of the more traditional material such as metric spaces, this book takes a student from a discussion of continuity, through a study of some topological properties and constructions, to homotopy and homotopy groups, to simplicial and singular homology and finally to an introduction to fibre bundles with a view towards K-theory. These are subjects which are essential for research in algebraic topology, and desirable for students pursuing research in any branch of mathematics. In fact, if I may be so bold as to say so, the subjects covered by this book are those areas of topology which *all* mathematics undergraduates should ideally see. In that sense, the material is *essential* topology.

With this range of topics, and the low starting level, the coverage of each subject is, inevitably, not exhaustive. For example, there are many results about connectivity whose proofs could be understood by undergraduates at this level, but which do not appear in this book. Instead, a representative sample of such results is included, together with enough examples that the reader should fully understand the results presented. In an undergraduate course it seems better to present a brief account of several topics and give a feel for the overall shape of a subject, rather than an in-depth study of a small number of topics.

Some of the deeper results included are presented without proof, so that the student may meet an important theorem in the area even though the proof would lengthen the book unacceptably. In every such case references are given to books which do contain a complete proof.

Given the target audience, the book is designed to require as little prior knowledge as possible. Anyone who has some basic familiarity with functions, such as from a beginning course on calculus, should be able to follow the first four chapters. From Chapter 5 onwards, a little knowledge of algebra is required, in particular equivalence relations for Chapters 5 and 6, some familiarity with groups for Chapters 8 to 11, and with linear algebra and quotient groups for Chapters 9 and 10.

There is a short bibliography included, listing books where students can find details of the proofs which have been omitted. I have not included a list of further reading, as there are many books in topology and algebraic topology that should be intelligible to someone who has read through this book. The choice of which follow-on text to use is a matter of personal taste or, for students embarking on postgraduate study, is something that their supervisor will advise them about.

Acknowledgements

In writing on topology I must first thank John Hubbuck and Michael Crabb who taught me the subject as an undergraduate, and as a postgraduate.

I would also like to express my gratitude to my colleagues here at Swansea, particularly Francis Clarke who has passed on many helpful observations on communicating topology, and Geoff Wood who thought of Example 5.23. Thanks are also due to Simon Cowell and Nikki Burt who read preliminary versions of the manuscript and offered a number of helpful comments.

In the preparation of this book Karen Borthwick, Jenny Wolkowicki and Frank Ganz of Springer Verlag have been incredibly helpful, patiently leading me through all the different stages of the publishing process, for which I am very grateful. I also greatly appreciate the work of the anonymous reviewers who provided a wealth of helpful and perceptive comments. The mistakes and rough edges that remain are, of course, entirely my fault, but the reader should be truly thankful to these reviewers for greatly reducing their number.

Finally I wish to thank my family: Ailsa, Calum and Jenny, for being a constant source of encouragement and distraction.

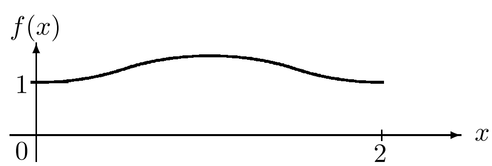
6
Homotopy

We said at the beginning of this book that topology is about the study of continuous functions and so the ultimate goal of topology should be to describe all the continuous maps between any given pair of topological spaces. Of course, with almost any pair of spaces, there are lots of continuous functions between them – far more than we can ever hope to list or understand. For example, it is not remotely feasible to list even the continuous functions from the interval $[0,1]$ to itself.

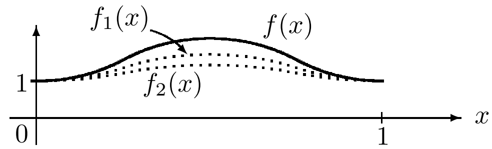
However, if we allow some leeway, then this difficulty can be avoided. The idea is that we should consider two functions to be equivalent, or "homotopic", if one can be deformed into the other.

6.1 Homotopy

For example, let $f : [0,2] \to \mathbf{R}$ be the function $f(x) = 1 + x^2(x-2)^2$, depicted below.



This is almost a constant function to 1, but with a small deviation around $x = 1$. If we take the function $f_1(x) = 1 + \frac{1}{2}x^2(x-2)^2$, then this has a similar shape, but with a smaller deviation. Similarly, $f_2(x) = 1 + \frac{1}{3}x^2(x-2)^2$ has the same shape but with an even smaller deviation.



Carrying on, for each $n \geq 1$, we can define $f_n(x) = 1 + \frac{1}{(n+1)}x^2(x-2)^2$, and thus obtain a family of functions interpolating between f and the constant function.

However, we need these interpolating functions to provide a *continuous* deformation of the one function into the other. To achieve this, we should not parametrize the interpolating functions f_1, f_2, etc. by integers, but, instead, we should index them by real numbers in some fixed range, say between 0 and 1. So we would then want a family of functions $\{f_t\}_{t\in[0,1]}$, such that $f_0 = f$, and f_1 is the constant function to 1.

In the above example, we can set $f_t(x) = 1 + (1-t)x^2(x-2)^2$ for each $t \in [0,1]$. Then $f_0(x) = 1 + x^2(x-2)^2 = f(x)$ and $f_1(x) = 1$ is the constant function.

Such a deformation then assigns a function to each point in $[0,1]$, so the deformation is a function from $[0,1]$ to the set of continuous maps $[0,2] \to \mathbf{R}$, which takes $t \in [0,1]$ to the function f_t. For the deformation to be continuous, we should obviously ask that this function be continuous. However, this would require us to put a topology on the set of continuous maps $[0,2] \to \mathbf{R}$ and, more generally, on the set of maps $S \to T$ for any topological spaces S and T. This can be done, and we will see how in Chapter 11, but for now we will use a simpler route to specify that the deformation be continuous.

Note that the family $\{f_t\}_{t\in[0,1]}$ assigns, to each point $t \in [0,1]$, a function $f_t : [0,2] \to \mathbf{R}$. This, in turn, assigns, to each point $x \in [0,2]$, a value $f_t(x) \in \mathbf{R}$. Thus we can think of this family as assigning to each pair $(x,t) \in [0,2] \times [0,1]$ the value $f_t(x) \in \mathbf{R}$. In other words, we have a function $[0,2] \times [0,1] \to \mathbf{R}$. Since we have a topology on $[0,2]$, and we know a topology on $[0,1]$, we can use the product topology to topologize $[0,2] \times [0,1]$, and therefore our interpolating family corresponds to a function between two topological spaces. Thus, we can define the family to be continuous if the corresponding function is continuous. Hence we arrive (finally!) at the following definition.

Definition: Two maps $f, g : S \to T$ are **homotopic** if there is a continuous function

$$F : S \times [0,1] \longrightarrow T$$

such that $F(s,0) = f(s)$ for all $s \in S$ and $F(s,1) = g(s)$ for all $s \in S$. In this case, F is a **homotopy** between f and g, and we write $f \simeq g$.

Example 6.1

In the preceding example, where $f : [0,2] \to \mathbf{R}$ is given by $f(x) = 1+x^2(x-2)^2$, the function $F : [0,2] \times [0,1] \to \mathbf{R}$ given by $F(x,t) = 1 + (1-t)x^2(x-2)^2$ is continuous, being a polynomial, and satisfies $F(x,0) = 1 + x^2(x-1)^2 = f(x)$ and $F(x,1) = 1$. Thus F is a homotopy from f to the constant function to 1.

Example 6.2

Let $f : S^1 \to \mathbf{R}^2$ be the natural inclusion map $f(x,y) = (x,y)$, and let $g : S^1 \to \mathbf{R}^2$ be the constant map $g(x,y) = (0,0)$ for all $(x,y) \in S^1$. These two maps are homotopic, for the function $F : S^1 \times I \to \mathbf{R}^2$ defined by

$$F((x,y),t) = (1-t)f(x,y)$$

is continuous, and has the property that $F((x,y),0) = (1-0)f(x,y) = f(x,y)$ and $F((x,y),1) = (1-1)f(x,y) = (0,0) = g(x,y)$.

Example 6.3

Let $f : [0,1] \to [0,1]$ be the identity map and let $g : [0,1] \to [0,1]$ be the constant map $g(x) = 0$ for all x. Then there is a homotopy $F : [0,1] \times I \to [0,1]$ between these maps given by

$$F(x,t) = (1-t)x.$$

Example 6.4

Let $f, g : \mathbf{R} \to \mathbf{R}$ be any two continuous functions. Define

$$F : \mathbf{R} \times [0,1] \to \mathbf{R}$$

by $F(x,t) = (1-t)f(x) + tg(x)$. Then F is continuous, being a composite of continuous functions, $F(x,0) = (1-0)f(x)+0 = f(x)$ and $F(x,1) = 0+1g(x) = g(x)$, so F is a homotopy between f and g. In other words, any two continuous functions on $\mathbf{R}$ are homotopic.

This idea can be used with any "convex" range space. A subspace T of $\mathbf{R}^n$ is said to be **convex** if, given any two points x, y in T, the straight line from x to y is contained in T. In other words, for any number $t \in [0,1]$, the point $tx + (1-t)y$ is in T.

Proposition 6.5

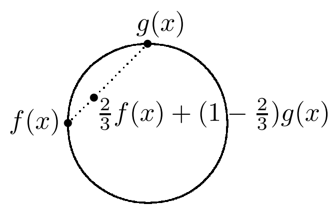
If T is convex, and S is any topological space, then any two maps $f, g : S \to T$ are homotopic.

Proof

Define the homotopy $F : S \times [0,1] \to T$ by

$$F(x,t) = tf(x) + (1-t)g(x). \qquad \square$$

On the other hand, we cannot use this argument for maps to S^1, for example, since if $f(x)$ and $g(x)$ are two distinct points in S^1, then $tf(x) + (1-t)g(x)$ will not usually be a point in S^1, as depicted below:



Now, we want to consider homotopic functions to be "the same". In other words, we want to form equivalence classes of homotopic functions, for which we need the following three lemmas.

Lemma 6.6

Let $f : S \to T$ be any continuous map. Then $f \simeq f$.

Proof

We can define a homotopy

$$F : S \times I \to T$$

by $F(x,t) = f(x)$ for all t. Then $F(x,0) = f(x)$ and $F(x,1) = f(x)$. $\square$

Lemma 6.7

Let $f, g : S \to T$ be two continuous maps. If F is a homotopy between f and g, then there is also a homotopy between g and f.

Proof

If $F(x,0) = f(x)$ and $F(x,1) = g(x)$, then define

$$G : S \times I \to T$$

by $G(x,t) = F(x, 1-t)$. Then $G(x,0) = g(x)$ and $G(x,1) = f(x)$. □

Lemma 6.8

Let $f, g, h : S \to T$ be three continuous maps. If f and g are homotopic and g and h are homotopic, then f and h are homotopic.

Proof

Let $F : S \times I \to T$ be a continuous map such that $F(x,0) = f(x)$ and $F(x,1) = g(x)$, and let $G : S \times I \to T$ be a continuous map such that $G(x,0) = g(x)$ and $G(x,1) = h(x)$. Define a function $H : S \times I \to T$ by

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

This is continuous by a suitable version of the gluing lemma, and has the property that $H(x,0) = F(x,0) = f(x)$, while $H(x,1) = G(x,1) = h(x)$. □

These results say that we can form a set of equivalence classes of homotopic functions between two given topological spaces. We write $[S,T]$ for the set of **homotopy classes** of maps $S \to T$. This is much more manageable than the complete set of continuous maps from S to T. For example, if $S = T = \mathbf{R}$, then, by Example 6.4, all functions $S \to T$ are homotopic, so $[\mathbf{R}, \mathbf{R}]$ consists of a single element. This is an extreme case. A more interesting example is that, as we shall see later, $[S^1, S^1]$ contains one element for each integer. This gives some reason to believe that homotopy classes of maps still contain some information about the topology of the spaces involved.

Of course, for homotopy classes to be a useful tool in studying continuous functions, they must respect the most basic operation on functions, namely composition. Fortunately, they do:

Proposition 6.9

If $f \simeq g : S \to T$ and $h \simeq j : T \to U$, then $(h \circ f) \simeq (j \circ g) : S \to U$.

Proof

Let $F : S \times [0,1] \to T$ be a homotopy from f to g, and let $H : T \times [0,1] \to U$ be a homotopy from h to j. Define a homotopy $G : S \times [0,1] \to U$ by

$$G(s,t) = H(F(s,t),t).$$

It is straightforward to check that $G(s,0) = h(f(s))$ and $G(s,1) = j(g(s))$, and the map G is continuous since it is a composite of continuous maps. $\square$

6.2 Homotopy Equivalence

If we are going to consider two functions to be equivalent when they are homotopic, then we should modify the definition of homeomorphism, replacing the $=$ signs by homotopies. This leads to the following notion of "homotopy equivalence":

Definition: Two topological spaces S, T are **homotopy equivalent** if there are continuous maps $f : S \to T$ and $g : T \to S$ such that $g \circ f$ is homotopic to the identity on S and $f \circ g$ is homotopic to the identity on T. If S and T are homotopy equivalent, then we write $S \simeq T$.

Lemma 6.10

If $S \simeq T$ and Q is any topological space, then $[S,Q] = [T,Q]$ and $[Q,S] = [Q,T]$.

Proof

If $S \simeq T$, then there are maps $f : S \to T$, $g : T \to S$ whose composites are homotopic to the respective identity maps. Now if $h : S \to Q$, then we can compose with g to obtain a map $(h \circ g) : T \to Q$, and if $j : T \to Q$, then we can compose with f to obtain a map $(j \circ f) : S \to Q$. And, up to homotopy, these two operations are mutually inverse: $(h \circ g) \circ f = h \circ (g \circ f) \simeq h \circ 1_S = h$, while $(j \circ f) \circ g = j \circ (f \circ g) \simeq j \circ 1_T = j$.

In a similar way, by composing with f or with g we can get correspondences between $[Q,S]$ and $[Q,T]$. $\square$

At the time of writing, it is becoming increasingly common to say that two spaces are **homotopic** rather than "homotopy equivalent".

Lemma 6.11

If S and T are homeomorphic, then they are also homotopy equivalent.

Proof

If we have homeomorphisms $f : S \to T$ and $g : T \to S$, then $f \circ g$ and $g \circ f$ are the respective identity maps, so these composites are homotopic to the respective identity maps by Lemma 6.6. □

Of course, there are many pairs of spaces which are homotopy equivalent but not homeomorphic.

Example 6.12

If S is a space containing a single point, then S and $\mathbf{R}$ are homotopy equivalent. To see this, define $f : \mathbf{R} \to S$ to be the constant function (there is no choice as to how to define f), and let $g : S \to \mathbf{R}$ be the function which takes the single point in S to 0 in $\mathbf{R}$. The composite $f \circ g : S \to S$ is the identity map, while the composite $g \circ f : \mathbf{R} \to \mathbf{R}$ is the constant function to 0. Since all functions $\mathbf{R} \to \mathbf{R}$ are homotopic, by Example 6.4, so $g \circ f$ is homotopic to the identity.

By Lemma 6.10, this tells us that $[\mathbf{R}, \mathbf{R}] = [\{0\}, \{0\}]$. Since there is only one continuous function $\{0\} \to \{0\}$, there can be only one homotopy class of maps $\{0\} \to \{0\}$. Thus $[\{0\}, \{0\}]$ contains only one element and, consequently, so does $[\mathbf{R}, \mathbf{R}]$, confirming Example 6.4.

A space which, like $\mathbf{R}$, is homotopy equivalent to a one-point space is said to be **contractible**.

Example 6.13

The interval $[0, 1]$ is homotopy equivalent to a one-point space $\{0\}$: Define $f : [0, 1] \to \{0\}$ by $f(x) = 0$, and define $g : \{0\} \to [0, 1]$ by $g(0) = 0$. Then $(f \circ g) : \{0\} \to \{0\}$ is the identity map, and so this is certainly homotopic to the identity map.

Conversely, $(g \circ f)(x) = 0$ for all x. This is homotopic to the identity map by Example 6.3.

Example 6.14

In the same way we can show that the open interval $(0,1)$ is homotopy to $\{0\}$. Of course, we need to define $g : \{0\} \to (0,1)$ differently, for example by $g(0) = 1/2$. With this choice of g, a homotopy from $(g \circ f)$ to the identity map of $(0,1)$ is given by $H : (0,1) \times [0,1] \to (0,1)$ defined by

$$H(x,t) = \frac{1 - t + tx}{2}.$$

The image $H(x,t)$ is certainly contained in $(0,1)$ if $x, t \in (0,1) \times [0,1]$, and H is continuous, being a composite of multiplications and additions.

Example 6.15

Consequently, any open interval (a,b) is homotopy equivalent to $\{0\}$, since (a,b) is homeomorphic with $(0,1)$. This applies even to infinite intervals (a,∞) and $(-\infty,b)$.

Proposition 6.16

If S is contractible and T is any topological space, then any two continuous functions $f, g : T \to S$ are homotopic. In particular, any continuous function to a contractible space is homotopic to a constant map.

Proof

Let $f, g : T \to S$ be two continuous maps. If S is contractible, then there are continuous maps $h : S \to \{0\}$ and $j : \{0\} \to S$ such that $h \circ j \simeq 1$ and $j \circ h \simeq 1$. In particular,

$$f = (1 \circ f) \simeq (j \circ h \circ f) \quad \text{and} \quad g = (1 \circ g) \simeq (j \circ h \circ g).$$

Since $h \circ f : T \to \{0\}$, so $j \circ h \circ f : T \to S$ must be the constant map $t \mapsto j(0)$ for all $t \in T$. Similarly, $j \circ h \circ g$ is this same constant map, and so $f \simeq g$. $\square$

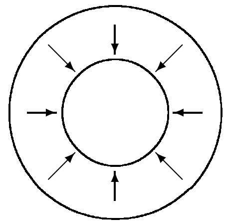
Of course, there are many pairs of spaces which are homotopy equivalent without being contractible.

Example 6.17

Let A be the annulus

$$A = \{(x,y) \in \mathbf{R}^2 : 1 \leq \sqrt{x^2 + y^2} \leq 2\}.$$

Then $A \simeq S^1$ as follows. Define $f : S^1 \to A$ to be the natural inclusion $f(x,y) = (x,y)$, and $g : A \to S^1$ to be the radial projection inwards



which can be described algebraically as

$$g(x,y) = \frac{1}{\sqrt{x^2+y^2}}(x,y).$$

Now, $g \circ f$ is the identity on S^1 because, if $(x,y) \in S^1$, then $g(x,y) = (x,y)$. This is certainly homotopic to the identity, by Lemma 6.6.

And $(f \circ g)(x,y) = \left(1/\sqrt{(x^2+y^2)}\right)(x,y)$. This is homotopic to the identity on A by the homotopy $F : A \times [0,1] \to A$ defined by

$$F((x,y),t) = \frac{t\sqrt{x^2+y^2} + (1-t)}{\sqrt{x^2+y^2}}(x,y).$$

Check: This is continuous, being a composite of continuous maps; $F((x,y),0) = (f \circ g)(x,y)$ and $F((x,y),1) = (x,y)$.

Hence f and g form a homotopy equivalence between A and S^1. We will see in the next section that these spaces are not contractible.

Example 6.18

Similarly, the space $\mathbf{C}^\times = \mathbf{R}^2 - \{(0,0)\}$ is homotopy equivalent to S^1.

Proving that two spaces are not homotopy equivalent is hard, just as it was hard to prove directly that two spaces are not homeomorphic. One case where we can do this is when we are dealing with finite discrete spaces, such as S^0.

Proposition 6.19

The 2-point space S^0 is not contractible.

Proof

Suppose that S^0 is contractible, with homotopy equivalences $f : S^0 \to \{0\}$ and $g : \{0\} \to S^0$. Then $f \circ g : \{0\} \to \{0\}$ has to be the identity, and $g \circ f : S^0 \to S^0$

is homotopic to the identity. So there is a homotopy

$$F : S^0 \times I \to S^0$$

with $F(x,0) = x$ and $F(x,1) = g(f(x)) = g(0)$.

Now define a map $h : I \to S^0$ by

$$h(t) = F(-g(0), t).$$

This will be a continuous map, with $h(0) = F(-g(0), 0) = -g(0)$ and $h(1) = F(-g(0), 1) = g(0)$. Since S^0 only has two points, h must be surjective. But Lemma 4.3 and Example 4.5 show that this cannot happen. Hence there cannot have been a homotopy equivalence between S^0 and $\{0\}$. □

This idea can be developed to show that a space consisting of m points is homotopy equivalent to a space consisting of n points *only* if $m = n$. It can also be developed to show that a connected space cannot be homotopy equivalent to a disconnected space.

Proposition 6.20

If S is connected and T is disconnected, then S and T are not homotopy equivalent.

Proof

Suppose that S and T are homotopy equivalent, with maps $f : S \to T$ and $g : T \to S$ whose composites are homotopic to the identity. In particular, there is a homotopy $F : T \times [0,1] \to T$ such that $F(t,0) = f(g(0))$ and $F(t,1) = t$ for all $t \in T$.

If T is disconnected, then it can be expressed as a disjoint union $T = U \amalg V$ where both U and V are open and non-empty, and so there is a continuous surjection $p : T \to S^0$ with $p(t) = 1$ if $t \in U$, $p(t) = -1$ if $t \in V$.

Since S is connected, the map f has image contained in one of these components, say $\operatorname{Im} f \subset U$. Since V is not empty, there is at least one point $v \in V$, and we can define a map $h : [0,1] \to S^0$ by

$$h(x) = p(F(v, x)).$$

Since $F(v,0) = f(g(0)) \in U$ and $F(v,1) = v \in V$, so $h(0) = 1$ and $h(1) = -1$. Thus h is a surjection $[0,1] \to S^0$, and h is continuous as it is a composite of continuous maps. As in the preceding proposition, this is not possible, so S and T cannot be homotopy equivalent. □

Thus we can use connectivity to distinguish homotopy inequivalent spaces.

Example 6.21

The circle S^1 is not homotopy equivalent to the 0-sphere S^0.

Unfortunately, compactness cannot be used in this way since Examples 6.13 and 6.14 exhibit two spaces which are both contractible and, hence, homotopy equivalent, but one is compact and the other is not. Similarly, there are Hausdorff spaces which are homotopy equivalent to non-Hausdorff spaces.

Example 6.22

Let $S = \{1, 2\}$ with the indiscrete topology, as in Example 4.36. Let T be a one-point space, say $T = \{0\}$. We can define $f : T \to S$ by $f(0) = 1$ and this is continuous since every function to an indiscrete space is continuous (Proposition 3.9). We can define $g : S \to T$ by $g(s) = 0$, this being continuous as T is also indiscrete. Then $g \circ f : T \to T$ is the identity map, and $f \circ g : S \to S$ is the constant map to 1. This is not the identity, but is homotopic to it, as we will now show. Such a homotopy will be a function $F : S \times [0, 1] \to S$. Since S is indiscrete, any such function is continuous, i.e., we can define F any way we choose and it will be continuous. In particular, we can define F by

$$F(s,t) = \begin{cases} s & \text{if} \quad t \leq \frac{1}{2}, \\ 1 & \text{if} \quad t > \frac{1}{2}. \end{cases}$$

Hence $F(s, 0) = s$ is the identity on S, and $F(s, 1) = 1 = (f \circ g)(s)$. Thus we have a homotopy from $f \circ g$ to the identity, so S is homotopy equivalent to T. Example 4.36 showed that S is not Hausdorff, whereas T is Hausdorff, being a subspace of $\mathbf{R}$.

So the properties developed in Chapter 4 are of limited use in a homotopy context. In particular, there are many interesting spaces which share all of these properties, while being quite distinct. For example the circle, S^1, is connected, compact, and Hausdorff, i.e., it looks just like a point as far as Chapter 4 is concerned. Instinctively, we can see that S^1 is not homotopy equivalent to a point, but our instinct can sometimes be wrong (our instinct would tell us that $\{1, 2\}$ could not be homotopy equivalent to a one-point space, in contrast to Example 6.22) so we need a rigorous proof before we can be entirely confident. The next section contains such a proof.

6.3 The Circle

We wish to prove that the circle is not contractible. However, for only a little extra effort, we can perform a more impressive calculation which will enable us to list all homotopy classes of maps $S^1 \to S^1$. This section is devoted to that calculation.

The trick is to open the circle out, and consider maps $[0,1] \to S^1$ instead of $S^1 \to S^1$. Since S^1 can be obtained as a quotient of $[0,1]$ by gluing the endpoints together (see Example 5.51), there is a continuous surjection $\pi : [0,1] \to S^1$, and we will study maps $f : S^1 \to S^1$ by looking at the composites $f \circ \pi : [0,1] \to S^1$.

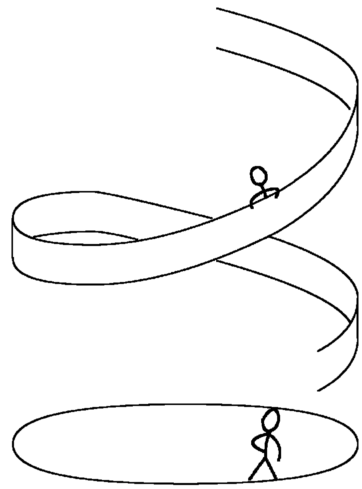
Having opened the circle out in this way, it turns out that we can "lift" any map $[0,1] \to S^1$ to a map $[0,1] \to \mathbf{R}$ which, when we compose with the exponential map $e : \mathbf{R} \to S^1$ of Example 3.47, gives back the original map $[0,1] \to S^1$. More precisely:

Proposition 6.23 (Path Lifting)

If $g : [0,1] \to S^1$ is a continuous function and $x \in \mathbf{R}$ is any point such that $e(x) = g(0)$, then there is a unique continuous function $\tilde{g} : [0,1] \to \mathbf{R}$ such that $e\tilde{g}(t) = g(t)$ for all $t \in [0,1]$ and $\tilde{g}(0) = x$. So the following triangle commutes:

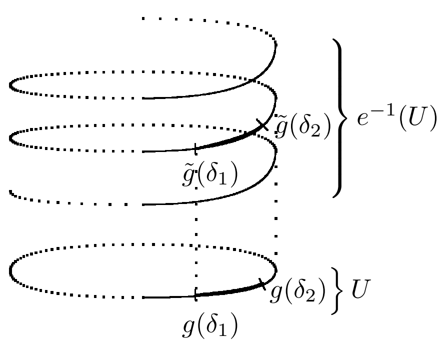
$$\begin{tikzcd} & \mathbf{R} \arrow[d, "e"] \\ {[0,1]} \arrow[ur, "\tilde{g}"] \arrow[r, "g"'] & S^1 \end{tikzcd}$$

One way to think of this statement is to consider the parameter $t \in [0,1]$ as specifying a moment in time. As t runs from 0 to 1, so g traces out a path in S^1. The condition $e\tilde{g}(t) = g(t)$ specifies that $\tilde{g}(t)$ must always be above $g(t)$ in the spiral picture of Example 3.47. It is as if one person is walking around a circle, and someone else is on a spiral staircase and determined always to be directly above the first person. Clearly they can always do that if they move fast enough (this is the existence part of the proposition), but there is no choice about where they move (this is the uniqueness part).



Proof

We will construct $\tilde{g}$ bit by bit. The key to this is that if we take any **proper subset** U of S^1 (i.e., any subset other than the whole of S^1), then its preimage under e is a disjoint union of infinitely many spaces, each homeomorphic to U. Now suppose we have a small interval $[\delta_1, \delta_2] \subset [0, 1]$ whose image, under g, is contained in U. And suppose that $\tilde{g}(\delta_1)$ is already defined in such a way that $e\tilde{g}(\delta_1) = g(\delta_1)$. Then $\tilde{g}(\delta_1)$ lies in one of these spaces homeomorphic to U. We can then compose that homeomorphism with g to define $\tilde{g}$ on the interval $[\delta_1, \delta_2]$ so as to agree with the value on δ_1.



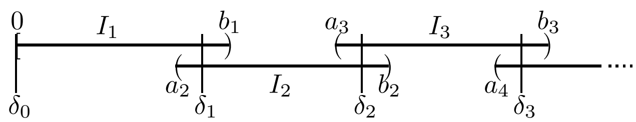
If we can split the interval $[0, 1]$ into a number of sections $[\delta_i, \delta_{i+1}]$, with $1 \leq i \leq n$, such that each section is mapped into some proper subset of S^1, then we can define $\tilde{g}$ inductively over the whole of $[0, 1]$, working with one of these sections at a time.

It is enough just to use two subsets of S^1, and we will use $U = S^1 - \{(1, 0)\}$ and $V = S^1 - \{(-1, 0)\}$. So both U and V are proper subsets and, between them, they contain every point of S^1. Note also that U and V are open, so the preimages $g^{-1}(U)$, $g^{-1}(V)$ will be open sets whose union is $[0,1]$. We can write $g^{-1}(U)$ and $g^{-1}(V)$ as a union of basic open sets, i.e., intervals (a,b), $[0,b)$ or $(a,1]$. This gives an open cover of $[0,1]$ where each set in the cover is a basic open set, and maps into either U or V. Since $[0,1]$ is compact, we can take a finite refinement of this cover, to get a list $I_1,\dots,I_n$.

Let us agree to order these open sets in the following way. First, 0 is contained in one of these sets; let that set be I_1. Then $I_1 = [0,b_1)$ for some $0 < b_1 \le 1$, and b_1 must be contained in another of these sets; let that set be I_2. Then $I_2 = (a_2,b_2)$ where $a_2 < b_1 < b_2$, or $I_2 = (a_2,1]$. In the first case, b_2 must be contained in another set; let that set be I_3. In the second case, I_1 and I_2 cover $[0,1]$, so we can take $n = 2$. And so forth.

In other words, we put the sets $I_1,\dots,I_n$ in the order in which we meet them as we travel from 0 to 1.

Let $\delta_0 = 0$, $\delta_n = 1$ and, for $1 \le i \le n-1$, $\delta_i = (a_{i+1}+b_i)/2$. Then $\delta_i \in I_i \cap I_{i+1}$ for $1 \le i \le n-1$ and so $[\delta_i,\delta_{i+1}] \subset I_{i+1}$ for $0 \le i \le n-1$.



We must have $\tilde g(0) = x$, so $\tilde g(\delta_0) = x$. Now $[\delta_0,\delta_1] \subset I_1$, and $g(I_1)$ is either contained in U or contained in V. In either case, there is a unique open interval of $\mathbf{R}$, containing x, and homeomorphic with U or V, whichever contains $g(I_1)$. We compose such a homeomorphism with g to get a continuous map $\tilde g : I_1 \to \mathbf{R}$ which sends δ_0 to x and is such that $e \circ \tilde g = g|I_1$.

In particular, we have defined $\tilde g(\delta_1)$. We can use the same argument again to define $\tilde g$ on the interval $[\delta_1,\delta_2]$, agreeing with the definition of $\tilde g(\delta_1)$. By the gluing lemma, 5.73, the extension of $\tilde g$ over $[0,\delta_2]$ is continuous.

Carrying on in the same way, we can define $\tilde g$ on the whole of $[0,1]$ and we have a continuous map $\tilde g : [0,1] \to \mathbf{R}$ such that $\tilde g(0) = x$ and $e \circ \tilde g = g$.

Finally, we must prove that the lifting $\tilde g$ is unique. So suppose that $\bar g : [0,1] \to \mathbf{R}$ is another lift of g with $\bar g(0) = x = \tilde g(0)$. Since $e \circ \bar g = e \circ g$ we see that $\bar g(y) - \tilde g(y) \in \mathbf{Z}$ for all y. Thus we get a continuous map $\bar g - \tilde g : [0,1] \to \mathbf{Z}$. By Lemma 4.18, this map must be constant. Since $\bar g(0) = \tilde g(0) = x$, we conclude that $\bar g(y) - \tilde g(y) = 0$ for all y, i.e., $\bar g = \tilde g$. Hence the lift $\tilde g$ is unique. $\square$

If we take a continuous map $f : S^1 \to S^1$ and form the composite $g = f \circ \pi : [0,1] \to S^1$, then $g(0) = g(1)$. So if we apply this proposition to g, the resulting lift $\tilde g$ satisfies $e\tilde g(0) = e\tilde g(1)$. Now $e(t) = e(s)$ if, and only if, $t - s$ is an integer. So $\tilde g(1) - \tilde g(0) \in \mathbf{Z}$. Thus, for each map $f : S^1 \to S^1$, we obtain an integer $\tilde{g}(1) - \tilde{g}(0)$, which we call the **degree**, or **winding number**. of f, written $\deg(f)$. Of course, we need to verify that this only depends on f and not on the choice of lifting $\tilde{g}$. But by the uniqueness condition, we know that $\tilde{g}$ is determined by its start point $\tilde{g}(0)$. And this must be such that $e\tilde{g}(0) = g(0)$. Hence, if $\tilde{g}$, $\bar{g}$ are two lifts such that $e\tilde{g} = e\bar{g}$, then $e\tilde{g}(0) = e\bar{g}(0)$ and, so, $\tilde{g}(0) = \bar{g}(0) + c$ for some integer c. Then $x \mapsto \bar{g}(x) + c$ gives another lift of g which agrees with $\tilde{g}$ at 0. Hence, by the uniqueness, $\tilde{g} = \bar{g} + c$. In particular, $\tilde{g}(1) - \tilde{g}(0) = \bar{g}(1) + c - (\bar{g}(0) + c) = \bar{g}(1) - \bar{g}(0)$. In other words, $\tilde{g}$ and $\bar{g}$ give the same answer for the degree of g. Hence this degree does not depend on the choice of lifting.

Example 6.24

Any constant function $S^1 \to S^1$ has degree 0, for the composite $g = f \circ \pi$ will be constant, and the lift $\tilde{g}$ can be taken to be constant: If $x \in \mathbf{R}$ is such that $e(x) = g(0)$, and we define $\tilde{g}$ by $\tilde{g}(t) = x$ for all t, then $e\tilde{g}(t) = e(x) = g(0)$. Hence $\deg(f) = 0$.

Example 6.25

The identity map $S^1 \to S^1$ has degree 1. For $g : [0,1] \to S^1$ is the map $g(t) = (\cos(2\pi t), \sin(2\pi t))$, and a lift is given by $\tilde{g}(t) = t$.

Example 6.26

If f is the map

$$f(\cos(\theta), \sin(\theta)) = (\cos(2\theta), \sin(2\theta)),$$

so that $g : [0,1] \to S^1$ is the map $t \mapsto (\cos(4\pi t), \sin(4\pi t))$, then a lift $\tilde{g}$ is given by $\tilde{g}(t) = 2t$, so $\deg(f) = 2$.

Example 6.27

If n is an integer and f is the map

$$f(\cos(\theta), \sin(\theta)) = (\cos(n\theta), \sin(n\theta)),$$

so that $g : [0,1] \to S^1$ is the map $t \mapsto (\cos(2n\pi t), \sin(2n\pi t))$, then a lift $\tilde{g}$ is given by $\tilde{f}(t) = nt$, so $\deg(f) = n$.

So, we have constructed an integer, the degree, for any map $S^1 \to S^1$. This has not yet told us anything about homotopy classes of maps $S^1 \to S^1$.

However, it turns out that homotopic maps have equal degrees. This is proved using the following variant of Proposition 6.23, which shows that just as we can lift paths, so we can lift homotopies.

Proposition 6.28 (Homotopy Lifting)

If $F : [0,1] \times [0,1] \to S^1$ is a continuous function, and $x \in \mathbf{R}$ is any point such that $e(x) = F(0,0)$, then there is a unique continuous function $\tilde{F} : [0,1] \times [0,1] \to \mathbf{R}$ such that $e\tilde{F}(s,t) = F(s,t)$ for all $s,t \in [0,1]$ and $\tilde{F}(0,0) = x$. So the following triangle commutes:

$$\begin{tikzcd} & \mathbf{R} \arrow[d, "e"] \\ {[0,1] \times [0,1]} \arrow[ur, "\tilde{F}"] \arrow[r, "F"'] & S^1 \end{tikzcd}$$

The basic idea of the proof is, as you would expect, the same as for Proposition 6.23. But splitting the square $[0,1] \times [0,1]$ into smaller chunks requires more care than splitting the interval $[0,1]$. Since the ideas we need to split the square will be used a few more times in the book, we present them in a slightly more general form here.

To simplify the statement of the next result, we say that a subset of $\mathbf{R}^n$ has **diameter** less than d if the distance between any pair $\mathbf{x}, \mathbf{y}$ of points in the subset is less than d.

Proposition 6.29 (Domain Splitting)

Suppose we have a map $f : X \to Y$, where X is a compact subset of $\mathbf{R}^n$, and an open cover $\mathcal{U}$ of Y. Then there is some number $\delta > 0$ such that whenever V is a subset of X of diameter less than δ, its image $f(V)$ is contained in one of the sets in $\mathcal{U}$.

Proof

As the map f is continuous, the preimages of the open sets in $\mathcal{U}$ will be open sets in X and these will give an open cover $\mathcal{W}$ of X. Any subset V of X which is contained in one of the sets in $\mathcal{W}$ will, then, have the property that its image, $f(V)$, is contained in one of the sets in $\mathcal{U}$.

The number δ then comes from the Lebesgue lemma, stated next. $\square$

Lemma 6.30 (Lebesgue Lemma)

Given a compact subspace X of $\mathbf{R}^n$ and an open cover $\mathcal{U}$ of X, there is some $\delta > 0$ such that any subset U of X of diameter less than δ is contained in one of the sets in $\mathcal{U}$.

Proof

Since X is compact, we can refine $\mathcal{U}$ to a finite list $U_1, \ldots, U_n$ of open subsets of X. Then, for $1 \leq i \leq n$, define $f_i : X \to \mathbf{R}$ by setting $f_i(x)$ to be the largest radius r such that $B_r(x)$ is contained in U_i. We take $f_i(x)$ to be 0 if $x \notin U_i$. This is continuous, as is more easily seen by considering f_i as the distance from x to the point in $X - U_i$ nearest to x. Thus the function $f : X \to \mathbf{R}$ defined by $f(x) = \max\{f_i(x) : 1 \leq i \leq n\}$ is also continuous. This function gives the largest radius r such that $B_r(x)$ is contained in one of the open sets U_i.

If there is some $\delta > 0$ such that $f(x) \geq \delta$ for all x, then every open ball of radius less than δ is contained in some open set U_i. Every set of diameter less than δ is contained in an open ball of radius δ, and so the lemma follows.

To see that there is such a δ, note that $f(x) > 0$ for all x, so 0 is not in the image of f. Since X is compact, Proposition 4.27 shows that the image of f will be a compact subset of $\mathbf{R}$. By the Heine–Borel Theorem 4.29, it is thus a closed subset of $\mathbf{R}$, so its complement is open. As this complement contains 0, it also contains some interval $(-\delta, \delta)$ around 0. Hence $f(x) \geq \delta$ for all x. $\square$

Proof (of Proposition 6.28)

Suppose, then, that we have a homotopy $F : [0,1] \times [0,1] \to S^1$. By covering S^1 with the open sets $U = S^1 - \{(1,0)\}$, $V = S^1 - \{(-1,0)\}$ as before, we can obtain a number $\delta > 0$ such that any subset of $[0,1] \times [0,1]$ of diameter less than δ is mapped into either U or V by F.

We split $[0,1] \times [0,1]$ into an $n \times n$ grid, where n is chosen so that $1/n < \delta/\sqrt{2}$, i.e., each square has diameter less than δ. Hence each square is mapped by F into either U or V.

If $\tilde{F}(0,0) = x$, then that determines a component of $e^{-1}(U)$ or $e^{-1}(V)$ and, hence, a homeomorphism between that component and U or V. By this homeomorphism, we define $\tilde{F}$ on the square $[0, \frac{1}{n}] \times [0, \frac{1}{n}]$. In particular, this defines $\tilde{F}(0, \frac{1}{n})$ and, by the same process, we can define $\tilde{F}$ on the square $[0, \frac{1}{n}] \times [\frac{1}{n}, \frac{2}{n}]$. However, this means defining $\tilde{F}$ on the path $[0, \frac{1}{n}] \times \frac{1}{n}$, based on its value at $(0, \frac{1}{n})$. The problem is that we have already defined $\tilde{F}$ on $[0, \frac{1}{n}] \times \frac{1}{n}$ when we defined it on the square $[0, \frac{1}{n}] \times [0, \frac{1}{n}]$, so we have two definitions which may contradict each other. Fortunately, the uniqueness of path lifting ensures that this cannot happen – if these two paths agree on $(0, \frac{1}{n})$ then they agree everywhere. Hence we can $\tilde{F}$ define on $[0, \frac{1}{n}] \times [0, \frac{2}{n}]$ without problem. Similarly, we can define $\tilde{F}$ on $[0, \frac{1}{n}] \times [0, \frac{3}{n}]$, and so on, until we have $\tilde{F}$ defined on the entire strip $[0, \frac{1}{n}] \times [0, 1]$. Then we use the definition of $\tilde{F}(\frac{1}{n}, 0)$ to define $\tilde{F}$ on the square $[\frac{1}{n}, \frac{2}{n}] \times [0, \frac{1}{n}]$. This entails redefining $\tilde{F}$ on the edge $\frac{1}{n} \times [0, \frac{1}{n}]$ but, again, the uniqueness of path lifting ensures that this definition agrees with the previous one. Next we define $\tilde{F}$ on the square $[\frac{1}{n}, \frac{2}{n}] \times [\frac{1}{n}, \frac{2}{n}]$ based on its value at $(\frac{1}{n}, \frac{1}{n})$. This entails redefining $\tilde{F}$ on two edges $\frac{1}{n} \times [\frac{1}{n}, \frac{2}{n}]$ and $[\frac{1}{n}, \frac{2}{n}] \times \frac{1}{n}$. However, the uniquess of path lifting can be used in both cases to show that the new definition agrees with the old. Then, in a similar way, we can define $\tilde{F}$ on the rest of the strip $[\frac{1}{n}, \frac{2}{n}] \times [0, 1]$ and, continuing similarly, on the whole of the square $[0, 1] \times [0, 1]$.

As before, this lift is unique as, if $\bar{F}$ is a different lift, then $\bar{F}(x) - \tilde{F}(x)$ is an integer for all $x \in [0, 1] \times [0, 1]$. As both $\bar{F}$ and $\tilde{F}$ are continuous, this integer must be constant, i.e., independent of x. If $\bar{F}(0, 0) = \tilde{F}(0, 0)$, then this integer must be 0, i.e., $\bar{F} = \tilde{F}$. □

Having now established that homotopies can be lifted, we can deduce that homotopic maps have the same degree.

Corollary 6.31

If $f, g : S^1 \to S^1$ are homotopic, then $\deg(f) = \deg(g)$.

Proof

Let $H : S^1 \times [0, 1] \to S^1$ be a homotopy between f and g. Considered as a map defined on $[0, 1] \times [0, 1]$, we can lift this to a map $\tilde{H} : [0, 1] \times [0, 1] \to \mathbf{R}$. Then $\tilde{H}$ restricted to $[0, 1] \times \{0\}$ will give a lift for f, so $\deg(f) = \tilde{H}(1, 0) - \tilde{H}(0, 0)$. And $\tilde{H}$ restricted to $[0, 1] \times \{1\}$ will give a lift for g, so that $\deg(g) = \tilde{H}(1, 1) - \tilde{H}(0, 1)$. In fact, we can use $\tilde{H}$ to define a continuous map $D : [0, 1] \to \mathbf{Z}$ by $D(t) = \tilde{H}(1, t) - \tilde{H}(0, t)$. Then $\deg(f) = D(0)$ and $\deg(g) = D(1)$. However, by Lemma 4.18, such a function D must be constant, since $[0, 1]$ is connected. Hence $\deg(f) = \deg(g)$. □

Corollary 6.32

The circle is not contractible.

Proof

Suppose that $f : S^1 \to \{0\}$ and $g : \{0\} \to S^1$ were homotopy equivalences. So $g \circ f \simeq 1_{S^1}$. Now, $(g \circ f)(x,y) = g(0)$ for all $(x,y) \in S^1$, i.e., this composite is a constant function, and hence has degree 0. Conversely, the identity map has degree 1. As these are different, $g \circ f$ cannot be homotopic to the identity map, so S^1 is not contractible. $\square$

To compute $[S^1, S^1]$, we also need the following converse to Corollary 6.31:

Theorem 6.33

If $f, g : S^1 \to S^1$ are such that $\deg(f) = \deg(g)$, then f and g are homotopic.

Proof

The idea is to define a homotopy "upstairs". For simplicity we will assume that $(f \circ \pi)(0) = (g \circ \pi)(0)$, so that we can lift f and g to maps $\tilde{f}, \tilde{g} : [0,1] \to \mathbf{R}$ which satisfy $\tilde{f}(0) = \tilde{g}(0)$. Hence

$$\tilde{f}(1) = \deg(f) + \tilde{f}(0) = \deg(g) + \tilde{g}(0) = \tilde{g}(1).$$

Thus if we define $\tilde{H} : [0,1] \times [0,1] \to \mathbf{R}$ by

$$\tilde{H}(s,t) = t\tilde{f}(s) + (1-t)\tilde{g}(s),$$

then $\tilde{H}(0,t) = \tilde{f}(0) = \tilde{g}(0)$ does not depend on t, and $\tilde{H}(1,t) = \tilde{f}(1) = \tilde{g}(1)$ is, similarly, independent of t. In particular, $\tilde{H}(1,t) - \tilde{H}(0,t) = \deg(f)$ is an integer. Hence when we compose with the exponential map $e : R \to S^1$, we find that $(e \circ \tilde{H})(0,t) = (e \circ \tilde{H})(1,t)$, so we can consider this as a map $H : S^1 \times [0,1] \to S^1$, which is a homotopy between f and g.

If $(f \circ \pi)(0) \neq (g \circ \pi)(0)$, then we use the following lemma to replace g by a function which does agree with f on $\pi(0)$. $\square$

Lemma 6.34

If $g : S^1 \to S^1$ and $(x,y) \in S^1$, then there is a map $h : S^1 \to S^1$ which is homotopic to g and such that $h(\pi(0)) = (x,y)$.

Proof

Let θ be the angle from $g(\pi(0))$ to (x,y). Define $H : S^1 \times [0,1] \to S^1$ so that $H((x',y'),t)$ is the rotation of (x',y') through the angle $t\theta$. Hence

$H((x', y'), 0) = (x', y')$, and $H((x', y'), 1)$ is (x', y') rotated by θ. In particular, $H(g(\pi(0)), 1) = (x, y)$, while $H(g(x, y), 0) = g(x, y)$. As H is continuous, it gives a homotopy from g to the map $h : S^1 \to S^1$ defined by $h(x', y') = H(g(x', y'), 1)$ which satisfies $h(\pi(0)) = H(g(\pi(0)), 1) = (x, y)$. □

Having completed the proof of Theorem 6.33, we can now give the promised calculation of the set of homotopy classes of self-maps of S^1.

Corollary 6.35

The set of homotopy classes of maps $S^1 \to S^1$ is in one-to-one correspondence with the set of integers, i.e., $[S^1, S^1] = \mathbf{Z}$.

Proof

Every continuous map $S^1 \to S^1$ has a degree, which is an integer. Homotopic maps have the same degree, and non-homotopic maps have different degrees. Hence $[S^1, S^1] \subset \mathbf{Z}$. To complete the proof, we note that all integers occur as the degree of a map, since, for any $n \in \mathbf{Z}$, the map $z \mapsto z^n$ ($z \in \mathbf{C}$, $|z| = 1$) has degree n. □

6.4 Brouwer's Fixed-Point Theorem

We have already seen a theorem saying that any continuous map from $[0, 1]$ to itself must have a fixed point. Our study of continuous maps from S^1 to S^1 can be used to prove a two-dimensional version of this theorem, due to Brouwer.

Theorem 6.36 (Brouwer's Fixed-Point Theorem)

Let $f : D^2 \to D^2$ be a continuous map, where D^2 is the closed disc

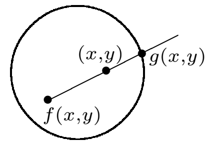
$$D^2 = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 \leq 1\}.$$

Then f has a fixed point, i.e., there is some point $(x, y) \in D^2$ with the property that $f(x, y) = (x, y)$.

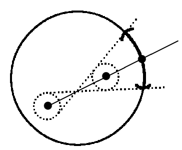
Proof

Suppose that $f : D^2 \to D^2$ does not have a fixed point, so that $f(x, y) \neq (x, y)$ for all $(x, y) \in D^2$. So, for each point $(x, y) \in D^2$ we get two points (x, y) and

$f(x,y)$, and we can draw a line through them both. Extend this line beyond (x,y) until it meets the boundary of D^2 (i.e., S^1), and let $g(x,y)$ be the point where this happens. So we get a function $g : D^2 \to S^1$ as in the picture.



This map g is continuous, essentially because if (x',y') is sufficiently close to (x,y), then $f(x',y')$ will be close to $f(x,y)$ (since f is continuous) and, hence, $g(x',y')$ will be reasonably close to $g(x,y)$. More rigorously, if A is an open arc around $g(x,y)$, then there is some radius r such that whenever (x',y') is in the open ball $B_r(x,y)$ and $f(x',y')$ is in the open ball $B_r(f(x,y))$, then $g(x',y')$ is in A, as depicted below, where A is indicated by a bold line, and the balls around (x,y) and $f(x,y)$ are indicated by the dotted circles of their perimeters. Any straight line which passes through both balls will hit the circle in the region A.



Since f is continuous, there is some radius δ such that $f(x',y') \in B_r(f(x,y))$ whenever $(x',y') \in B_\delta(x,y)$. Hence the preimage $g^{-1}(A)$ contains $B_\delta(x,y)$. The same argument can be applied to any point in the preimage, so $g^{-1}(A)$ is open, i.e., g is continuous.

If (x,y) is on the boundary of D^2, then $g(x,y) = (x,y)$ no matter what $f(x,y)$ is.

Now define a map

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

by $F((x,y),t) = g(tx,ty)$.

This map F is continuous, so we can think of it as a homotopy between the map $h : S^1 \to S^1$ defined by $h(x,y) = F((x,y),0)$ and $j : S^1 \to S^1$ defined by $j(x,y) = F((x,y),1)$. Now $h(x,y) = g(0,0)$ for all (x,y), so h is the constant map and thus $\deg(h) = 0$. On the other hand, however, $j(x,y) = g(x,y) = (x,y)$ for all (x,y), so j is the identity map and $\deg(j) = 1$. If F is a homotopy between h and j, then these degrees must be equal. Since they are not, the map F cannot exist. Hence nor can g, showing in turn that the map f must have had a fixed point in the first place. □

6.5 Vector Fields

One of the most celebrated theorems of topology is the “Hairy ball theorem”. In simple language this says that you cannot comb a hairy ball. To make this more precise, we need the notion of a “vector field”.

When combing a surface, such as the sphere, we move the comb in a certain direction, tangential to the surface. This gives a function which assigns, to each point on the surface, a direction, i.e., a vector which is tangential to the surface. For example, if we comb the sphere S^2, we will get a function $v : S^2 \to \mathbf{R}^3$ with the property that $v(s)$ is tangential to the surface of S^2 at s.

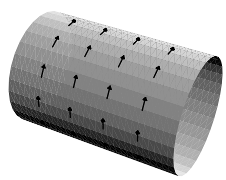
In general, combing a surface $S \subset \mathbf{R}^n$ will give rise to a function $v : S \to \mathbf{R}^n$. Of course, this function v should be continuous, as the comb is presumed to move in a continuous way. A continuous tangential vector-valued function such as this is called a **vector field**.

Example 6.37

At any given moment in time there is a vector field which assigns to each point on the surface of the earth, a vector representing the wind felt at that point.

Example 6.38

Another example of a vector field is given by combing a hairy cylinder.



If we constantly comb round the cylinder, then we get a nowhere-zero vector field, i.e., $v(s) \neq 0$ for all s in the cylinder.

We can use our knowledge about homotopy classes of maps from S^1 to S^1 to tell us about vector fields, as illustrated by the next two theorems.

Theorem 6.39

If you stir a cup of coffee, then, at any given moment in time, some particle on the surface is stationary.

Proof

Let $v : D^2 \to \mathbf{R}^2$ be the vector field indicating how the surface of the coffee is moving, so that $v(x, y)$ is the velocity of a particle of coffee at the point (x, y). If v is nowhere zero, then we can define a continuous map $g : S^1 \to S^1$ by

$$g(x, y) = \frac{-v(x, y)}{|v(x, y)|},$$

thinking of $(x, y) \in S^1$ as a point on the boundary of D^2. Then there is a homotopy $G : g \simeq id$ defined by

$$G((x, y), t) = \frac{t(x, y) - (1 - t)v(x, y)}{|t(x, y) - (1 - t)v(x, y)|}.$$

It takes some thought to see that this is continuous, as we must verify that the denominator $|t(x, y) - (1 - t)v(x, y)|$ cannot be zero. If it were zero, then this would say that $t(x, y) = (1 - t)v(x, y)$. If $t = 0$, then that would mean $v(x, y) = 0$, which cannot happen by assumption. If $t = 1$, then that would mean that $(x, y) = 0$, which cannot happen as $(x, y) \in S^1$. If $0 < t < 1$, then $t(x, y) - (1 - t)v(x, y) = 0$ implies that $v(x, y) = \frac{t}{1-t}(x, y)$, i.e., $v(x, y)$ is a positive multiple of (x, y). Since (x, y) is on the perimeter of the cup, this would be saying that the coffee is moving out of the cup, which cannot happen. Hence $|t(x, y) - (1 - t)v(x, y)|$ is never zero, so G is a continuous map.

On the other hand, if v is nowhere zero, then we can also define a homotopy $F : S^1 \times [0, 1] \to S^1$ by

$$F((x, y), t) = \frac{-v(tx, ty)}{|v(tx, ty)|}.$$

If $t = 1$, then $F((x, y), 1) = g(x, y)$ and if $t = 0$, $F((x, y), 0) = -v(0, 0)/|v(0, 0)|$ is constant. So F is a homotopy between g and a constant map. Putting these homotopies together, we get

$$id \simeq g \simeq \text{constant}.$$

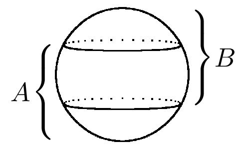
A constant map has degree 0, and the identity has degree 1, hence these two cannot be homotopic. So v must be zero somewhere, i.e., some point is stationary. □

Theorem 6.40 (Hairy Ball Theorem)

Let $v : S^2 \to \mathbf{R}^3$ be a vector field on the sphere. Then there is some point $\mathbf{x} \in S^2$ such that $v(\mathbf{x}) = 0$.

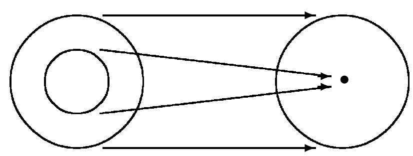
Proof

To prove this, we will split the sphere up into three sections, by latitude:



We will first consider the region A, i.e., everything below (and including) the upper line. This region is homeomorphic to the closed disc D^2, by stereographic projection. More importantly, if we think of v as placing an arrow at each point on S^2 tangential to S^2, then under this stereographic projection, v corresponds to a continuous map $\tilde{v}$ from D^2 to $\mathbf{R}^2$, placing an arrow at every point of D^2 tangential to D^2. (This correspondence sounds plausible for points near the South Pole. To prove that it works for the whole region A requires methods from multivariate calculus whose details we omit. See Section 7c of [4] for more information, or Section 2.2 of [5] for a different approach.)

Now we modify $\tilde{v}$ as follows. Let $h : D^2 \to D^2$ be the continuous map which shrinks the disc of radius $1/2$ within D^2 down to a point and stretches out the remainder of D^2 accordingly.



So, using polar coordinates, h can be written as

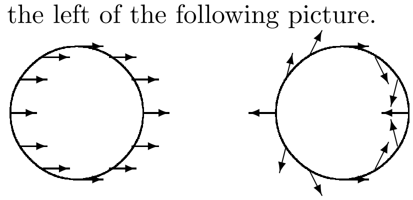
$$h(r\cos(\theta), r\sin(\theta)) = \begin{cases} ((2r-1)\cos(\theta), (2r-1)\sin(\theta)) & \text{if} \quad r \geq \frac{1}{2}, \\ (0,0) & \text{if} \quad r \leq \frac{1}{2}. \end{cases}$$

We define a new function $\tilde{w} : D^2 \to \mathbf{R}^2$ by $\tilde{w} = \tilde{v} \circ h : D^2 \to \mathbf{R}^2$, and we can think of this as a vector field on D^2. Note that $\tilde{w}(x,y) = \tilde{v}(0,0)$ if $|(x,y)| \leq 1/2$, i.e., $\tilde{w}$ is constant throughout the disc of radius $1/2$ inside D^2. But when $|(x,y)| = 1$, $\tilde{w}(x,y) = \tilde{v}(x,y)$, so $\tilde{v}$ and $\tilde{w}$ agree with each other on the perimeter of D^2.

The fact that $\tilde{v}$ and $\tilde{w}$ agree on the perimeter of D^2 means that we can patch $\tilde{w}$ in, in place of $\tilde{v}$, in our original vector field on S^2, to get a new continuous tangential vector field w. Because $\tilde{w}$ is constant in the middle of D^2, so w is "constant" south of the lower tropic.

Now we look at the region B, north of, and including this lower tropic. Again, by stereographic projection, this region is homeomorphic to D^2, but look what happens to the perimeter of this region under this homeomorphism.

On the sphere, this perimeter corresponds to the lower tropic, on which w is constant, depicted in the left of the following picture.



However, when we apply stereographic projection, it has the effect of turning this circle inside out. The arrows on the perimeter then point in different directions. In fact, as you can see from the right-hand picture above, the arrows rotate through 720° as you pass around the perimeter circle.

We can use this to define a continuous map $f : S^1 \to S^1$ by

$$f(x, y) = \frac{w'(x, y)}{|w'(x, y)|},$$

where $w' : D^2 \to \mathbf{R}^2$ is the function corresponding to w under stereographic projection. Since the arrows rotate through 720° as we go around the circle, so $\deg(f) = 2$. However, if v has no zeros, then w will have no zeros and so, in particular, w' will be nowhere zero. We could then define a homotopy $H : S^1 \times [0, 1] \to S^1$ by

$$H((x, y), t) = \frac{w'(t(x, y))}{|w'(t(x, y))|}.$$

If w' is nowhere zero, then this is continuous. When $t = 0$, $H((x, y), 0) = w'(0, 0)/|w'(0, 0)|$ is constant. So H would be a homotopy from a degree 2 map to a constant map. Since this cannot happen, w' must have a zero somewhere. Consequently, so must w and, in turn, v. $\square$

Since, as we have seen, wind can be considered as a vector field on the surface of the earth, which is homeomorphic to S^2, we get the following meteorological consequence.

Corollary 6.41

At any moment in time, there is some point on the earth where there is no wind.

Remark: Notice how we have applied our knowledge of $[S^1, S^1]$ to problems about vector fields, which don't directly involve S^1 or homotopy. A little knowledge can, indeed, go a long way, and this shows how useful the "homotopy" concept is that it can solve problems that have no apparent connection to it.

EXERCISES

6.1. Write down a homotopy equivalence between $(0,1)$ and $[0,1]$.

6.2. List all homotopy classes of maps $(0,1) \to (0,1)$.

6.3. Prove that a discrete space consisting of m points is homotopy equivalent to a discrete space consisting of n points if, and only if, $m = n$.

6.4. Let X be any space and $f : X \to S^n$ a continuous map. Using Proposition 6.5, show that if f is not surjective, then f is homotopic to a constant map.

6.5. Show that the map $f : S^1 \to S^1$ given by $f(x,y) = (x,-y)$ is homotopic to the identity map.

6.6. If $f, g : S^1 \to S^1$ are two continuous maps, express $\deg(f \circ g)$ in terms of $\deg(f)$ and $\deg(g)$. Use this to show that $f \circ g$ is homotopic to $g \circ f$.

6.7. Which of the following surfaces do you think can be combed (i.e., which admit a nowhere-zero tangential vector field): (1) a Möbius band, (2) a surface of genus two, (3) a torus, and (4) a Klein bottle?