MATH 8200 LECTURE NOTES (SPRING 2014)

MIKE USHER

1. Topological invariants and homotopy

A basic goal of topology is to determine under what circumstances two spaces X and Y are homeomorphic to each other–here a homeomorphism $f : X \to Y$ is a continuous bijection having a continuous inverse; since this means that a subset $V \subset X$ is open if and only if $f(V)$ is open this is evidently the appropriate notion of isomorphism for topological spaces.

If you think that two spaces X and Y are homeomorphic, the most obvious way to try to show this is to try to construct a specific homeomorphism $f : X \to Y$. For instance if you wanted to show that the open interval $(-1, 1)$ is homeomorphic to the real numbers $\mathbb{R}$ it would suffice to write down a function like $x \mapsto \frac{x}{1-x^2}$ and check that it is a continuous bijection between $(-1, 1)$ and $\mathbb{R}$ with continuous inverse. Likewise, introductory explanations of what topology is often note that a donut is homeomorphic to a coffee mug, as can be convincingly illustrated by an animation[1] in which the former continuously deforms to the latter.

However, if you think that X and Y are *not* homeomorphic, how would you prove this? For instance, it probably seems unlikely that a donut would be homeomorphic to a muffin, on the basis that transforming the donut to the muffin would presumably involve tearing it and not just deforming it and tearing is not a continuous transformation, but this bears no resemblance to an honest proof—perhaps there's some very clever continuous transformation that we just didn't think of.

The standard way of proving that spaces are not homeomorphic is by using *topological invariants*. The idea is to develop a rule which assigns to each topological space Z something that I'll denote $I(Z)$, and prove a general theorem that says that if two spaces X and Y are homeomorphic then $I(X)$ and $I(Y)$ are equivalent[2] (*i.e.*, I is "invariant" under homeomorphisms). Contrapositively, if we go compute the invariants $I(X)$ and $I(Y)$ and find them to be inequivalent then we can deduce that our two spaces X and Y are indeed not homeomorphic.

Whether or not you've seen this phrased in this specific way, you're probably familiar with this on some level. For a very easy example, presumably the way that you would prove that the topological spaces (with the discrete topology, for definiteness) $X = \{5\}$ and $Y = \{\pi, 63\}$ are not homeomorphic is by noting that X and Y have different numbers of elements. In the language of the previous paragraph, for any space Z let $I(Z)$ denote the number of elements of (the underlying set of) Z (or the symbol ∞ if Z is an infinite set). Since a homeomorphism is a bijection, if X were homeomorphic to Y then we would have $I(X) = I(Y)$, but we observe that in fact $I(X) \neq I(Y)$ and so X and Y are not homeomorphic.

For a more serious example, consider the statement that $X = [0, 1]$ and $Y = (0, 1)$ are not homeomorphic. If you're taking this class then you should know how to prove this—the observation is that X is compact (a non-trivial and important theorem from point set topology) and

[1] such as this one

[2] what "equivalent" means will depend on the context, as we'll see

Y isn't. We can phrase this in the language above: if Z is any space let

$$I(Z) = \begin{cases} \text{yes} & \text{if } Z \text{ is compact} \\ \text{no} & \text{otherwise} \end{cases}$$

Since the image of a compact space under a continuous function is compact, $I(Z)$ is evidently a topological invariant. So the fact that $I(X) = \text{yes}$ while $I(Y) = \text{no}$ implies that X and Y aren't homeomorphic.

Exercise 1.1. Prove that $\mathbb{R}$ is not homeomorphic to $\mathbb{R}^2$. (Hint: If $f : \mathbb{R} \to \mathbb{R}^2$ were a homeomorphism then it would restrict to a homeomorphism between $\mathbb{R} \setminus \{0\}$ and $\mathbb{R}^2 \setminus \{f(0)\}$.)

As the above examples illustrate, a topological invariant $I(Z)$ might usefully be a number, or the answer to a yes-or-no question. *Algebraic* topology is founded on the realization that it is often useful to have $I(Z)$ be an algebraic object—in this course usually it will be a group (the fundamental group or a homology group), and so if X and Y are homeomorphic then $I(X)$ and $I(Y)$ will be isomorphic as groups. In later courses you may learn about other topological invariants (notably cohomology) which are instead rings or more elaborate algebraic structures.

I'm going to take a somewhat unconventional approach of postponing the introduction of topologically invariant groups to the subject until we've learned a little more topology, but as a transition I will concentrate on invariants $I(Z)$ which are defined as sets.

Example 1.2. *To give another hopefully easy example, let $X = \mathbb{R} \setminus \{0\}$ and $Y = \mathbb{R} \setminus \{0, 1\}$. What invariant can we use to distinguish these spaces? I propose to define, for any space Z, $I(Z)$ to be the* set of all path components *of Z. (Recall that a path component of a topological space Z is an equivalence class under the equivalence relation on Z given by saying that two points $z_0, z_1 \in Z$ are equivalent if and only if there is a continuous map $\gamma\colon [0,1] \to Z$ such that $\gamma(0) = z_0$ and $\gamma(1) = z_1$.) So evidently $I(X) = \{(-\infty, 0), (0, \infty)\}$ and $I(Y) = \{(-\infty, 0), (0, 1), (1, \infty)\}$.*

In what sense is I an invariant? It's supposed to be true that the values of I on homeomorphic spaces should be "equivalent"—in this context $I(Z)$ is a set and the appropriate notion for sets to be equivalent is that there should be a bijection between them. In fact, if $f : X \to Y$ is a homeomorphism, then it is easy to see that $C \subset X$ is a path component of X if and only if $f(C) \subset Y$ is a path component of Y. ($x_0, x_1 \in X$ are joined by a path γ if and only if the path $f \circ \gamma$ joins $f(x_0)$ and $f(x_1)$). Consequently a homeomorphism $f : X \to Y$ would induce a bijection $f_*\colon I(X) \to I(Y)$. *But $I(X)$ has two elements and $I(Y)$ has three elements so no such bijection can exist and hence neither can the homeomorphism f.*

This example illustrates a property that most of the invariants of algebraic topology share: they can be seen as *functors* from the category of topological spaces (with morphisms given by continuous maps) to another category (here the category of sets with morphisms given by arbitrary functions). Without delving too much into categorical formalism (of course you can look up the precise definitions, see *e.g.* [H, p. 162] or Wikipedia) this means that in addition to associating a set $I(Z)$ to every space Z, the functor associates to *any* continuous map (not just a homeomorphism) $f : X \to Y$ between two topological spaces an induced map $f_*\colon I(X) \to I(Y)$ between the associated sets.[3] Moreover these induced maps are nicely-behaved in that the identity $1_X\colon X \to X$ induces the identity map on $I(X)$, and given composable continuous maps $f\colon X \to Y$ and $g\colon Y \to Z$ the map $(g \circ f)_*\colon I(X) \to I(Z)$ is the same as $g_* \circ f_*$. From these latter properties it is immediate that if f happens to be a homeomorphism then f_* is a bijection with inverse $(f^{-1})_*$. Thus functors automatically yield invariants, but they also yield quite

[3] In the case of Example 1.2, f_* sends a path component $C \subset X$ to the path component of $f(x)$ where x is an arbitrary element of C—it is easy to check that f_*C is independent of the choice of x.

a bit more, among other things since they can be used to study continuous maps that aren't necessarily homeomorphisms.

For one last introductory example, which will motivate several new ideas, consider the spaces $X = \mathbb{R}^2$ and $Y = \mathbb{R}^2 \setminus \{(0,0)\}$. These are both connected, noncompact, Hausdorff, second-countable, etc., and tricks like deleting a point in Exercise 1.1 don't seem to work, so introductory point-set topology seems powerless to tell the difference between X and Y. On the other hand if you try to construct a homeomorphism from X to Y you probably won't succeed.

We will see that X and Y are indeed not homeomorphic. To get a sense of why this should be, consider the unit circle $T \subset \mathbb{R}^2$, which is contained in both X and Y. In X, you can imagine "continuously shrinking T down to a point" without ever leaving the space. In fact the same would be true with T replaced by any circle embedded inside X. However it seems impossible to do this in Y, since shrinking down the circle T would require one to pass over $(0,0)$ at some time, and $(0,0)$ is not contained in Y.

It might be instructive at this point for you to try to write down a rigorous version of the above argument—if you haven't seen this material before it will probably be hard, as is often the case when you are trying to prove that something is impossible to do.

The way to prove the result involves the notion of *homotopy*, which is fundamental to the entire course. Here is the definition:

Definition 1.3. Let $h_0 \colon V \to X$ and $h_1 \colon V \to X$ be two continuous maps between the topological spaces V and X. We say that h_0 is *homotopic* to h_1 if and only if there is a continuous map $H \colon [0,1] \times V \to X$ such that $H(0,v) = h_0(v)$ and $H(1,v) = h_1(v)$ for all $v \in V$. In this case H is said to be a *homotopy* from h_0 to h_1.

Of course $[0,1] \times V$ has the standard product topology in the above definition. So a homotopy between two maps is essentially a continuous family of maps interpolating between them.

Exercise 1.4. Let $C(V,X)$ be the set of continuous maps from the topological space V to the topological space X and define a relation $\sim$ on $C(V,X)$ by saying that $h_0 \sim h_1$ if and only if h_0 is homotopic to h_1.

(a) Prove that $\sim$ is an equivalence relation.

(b) If Y is another topological space, $f \in C(X,Y)$, and $g \in C(Y,V)$, and if $h_0 \sim h_1$ where $h_0, h_1 \in C(V,X)$, prove that $f \circ h_0 \sim f \circ h_1$ and $h_0 \circ g \sim h_1 \circ g$.

Exercise 1.5. If V and X are two topological spaces let $[V,X]$ denote the set of equivalence classes of the relation $\sim$ on $C(V,X)$. If Y is another topological space and $f : X \to Y$ is continuous, "define" $f_* \colon [V,X] \to [V,Y]$ by, for any $h \in C(V,X)$, $f_*[h] = [f \circ h]$.

(a) Prove that f_* is indeed well-defined (*i.e.* that the image of an equivalence class under f_* is independent of the choice of representative of the equivalence class).

(b) Prove that such maps f_* are "functorial" in the sense that the map induced by the identity $1_X \colon X \to X$ is the identity on $[V,X]$, and that given continuous maps $f : X \to Y$ and $g \colon Y \to Z$ we have $g_* \circ f_* = (g \circ f)_*$.

(c) A *homotopy equivalence* $f : X \to Y$ is a continuous map with the property that there exists a continuous map $g \colon Y \to X$ such that $f \circ g$ is homotopic to the identity on Y and $g \circ f$ is homotopic to the identity on X. (So in particular a homeomorphism is obviously a homotopy equivalence.) Prove that if $f : X \to Y$ is a homotopy equivalence then for any topological space V the function $f_* \colon [V,X] \to [V,Y]$ is a bijection.

Definition 1.6. A space X is called *contractible* if X is homotopy equivalent to a one-point topological space.

Exercise 1.7. Prove that $\mathbb{R}^2$ is contractible, and that $\mathbb{R}^2 \setminus \{(0,0)\}$ is homotopy equivalent to the unit circle.

Let S^1 denote the quotient space $\frac{[0,1]}{0\sim 1}$. There is an obvious homeomorphism between S^1 and the unit circle in $\mathbb{C} \cong \mathbb{R}^2$ defined by $t \mapsto e^{2\pi i t}$. A special case of Exercise (1.5) (iii) says that if X and Y are homeomorphic then there is a bijection between $[S^1, X]$ and $[S^1, Y]$.

Returning to our example $X = \mathbb{R}^2$ and $Y = \mathbb{R}^2 \setminus \{(0,0)\}$, we will show that $[S^1, \mathbb{R}^2]$ and $[S^1, \mathbb{R}^2 \setminus \{(0,0)\}]$ are not in bijection. It is fairly easy to find the first of these sets; indeed quite generally we have:

Proposition 1.8. *For any topological space V and any n, $[V, \mathbb{R}^n]$ consists of only one element.*

Proof. It suffices to show that $f \colon V \to \mathbb{R}^n$ is any continuous map then f is homotopic to the constant map $g \colon V \to \mathbb{R}^n$ defined by $g(v) = 0$. To see this, simply note that the map $F \colon [0,1] \times V \to \mathbb{R}^n$ defined by $F(t, v) = (1-t)f(v)$ defines a homotopy from f to g (verification of this is left to the reader). $\square$

So to prove that $\mathbb{R}^2 \setminus \{(0,0)\}$ is not homotopy equivalent to $\mathbb{R}^2$ it is enough to show that there exists a continuous map $S^1 \to \mathbb{R}^2 \setminus \{(0,0)\}$ which is not homotopic to a constant map.[4] This is harder, and will involve a detour into covering space theory, which will eventually prove a stronger statement:

Theorem 1.9. *There is a bijection*

$$\deg \colon [S^1, \mathbb{R}^2 \setminus \{(0,0)\}] \to \mathbb{Z}$$

Roughly speaking, if $f \colon S^1 \to \mathbb{R}^2 \setminus \{(0,0)\}$ is a map with homotopy class $[f]$, then $\deg([f])$ measures the "number of times f winds around the origin," with counterclockwise winding counting positively and clockwise winding counting negatively. Viewing $\mathbb{R}^2 \setminus \{(0,0)\}$ as $\mathbb{C} \setminus \{0\}$, for any $n \in \mathbb{Z}$ we will see that the map $f_n(t) = e^{2\pi i n t}$ has $\deg([f_n]) = n$. Thus Theorem 1.9 asserts that no two of the maps f_n are homotopic, and that every continuous map $S^1 \to \mathbb{C} \setminus \{0\}$ is homotopic to one (and only one) of the f_n.

We will prove Theorem 1.9 using the theory of covering spaces, a notion which I will now introduce and which we will revisit in greater detail in a few weeks.

2. Introduction to Covering Spaces

Definition 2.1. If X is a topological space, a *covering space of X* is a continuous map $p \colon \tilde{X} \to X$ where $\tilde{X}$ is some other topological space, with the following property. There is an open cover $\{U_\alpha\}_{\alpha \in A}$ of X such that for each α the preimage $p^{-1}(U_\alpha) \subset \tilde{X}$ is a disjoint union of open subsets $V_{\alpha\beta}$ such that for each β the restriction $p|_{V_{\alpha\beta}}$ is a homeomorphism from $V_{\alpha\beta}$ to U_α.

One usually imagines $\tilde{X}$ hovering above the "base space" X. Then if one is standing at a point $x_0 \in X$ and looking up at $\tilde{X}$, one will see a discrete collection of copies of a small neighborhood of x_0. If one moves around in X, continuing to look up at $\tilde{X}$, something like these copies will continue to exist, though the definition allows for them to move around each other in various ways.

An obvious example of a covering space of X is given by choosing an arbitrary set S, endowed with the discrete topology, and putting $\tilde{X} = X \times S$ and p equal to the projection. This is a rather trivial example, and some references rule it out (except in the case where S is a one-point set)

[4]In view of Exercise 1.7 we could instead just look for a map $S^1 \to S^1$ which is not homotopic to a constant map, though we won't use this. By the way, a space X is called **simply connected** if $[S^1, X]$ has only one element, so we are going to show that $\mathbb{R}^2 \setminus \{(0,0)\}$ is not simply connected.

by requiring $\tilde{X}$ (and hence also X) to be connected. Connected covering spaces (when they are not homeomorphisms) are in some sense nontrivial, and our first example of one will be instrumental in the proof of Theorem 1.9

Example 2.2. *Let $X = \mathbb{C}\setminus\{0\}$. As you know, any element $z \in X$ can be written in polar coordinates as $z = re^{i\theta}$ for real numbers r and θ with $r > 0$. Of course r is uniquely determined by z (it is equal to $|z|$), but θ is not uniquely determined since adding an integer multiple of 2π to θ doesn't change z. (One way of ruling this issue out is by insisting that θ belong to a half-open interval like $[0, 2\pi)$, but from a topological perspective this is a bad idea since it means that θ doesn't depend continuously on z as one crosses the positive real axis.)*

The way of understanding this topologically is to say that polar coordinates don't give a homeomorphism but rather a nontrivial covering space. Specifically, define $\tilde{X} = (0,\infty)\times\mathbb{R}$ and $p\colon \tilde{X} \to X$ by

$$p(r, \theta) = re^{i\theta}$$

To see that this is a covering space, we cover X by the four open sets

$$U_1 = \{x + iy | x > 0\}$$
$$U_2 = \{x + iy | y > 0\}$$
$$U_3 = \{x + iy | x < 0\}$$
$$U_4 = \{x + iy | y < 0\}$$

We see that

$$p^{-1}(U_1) = \coprod_{n\in\mathbb{Z}} (0,\infty) \times (2n\pi - \pi/2, 2n\pi + \pi/2)$$
$$p^{-1}(U_2) = \coprod_{n\in\mathbb{Z}} (0,\infty) \times (2n\pi, 2n\pi + \pi)$$
$$p^{-1}(U_3) = \coprod_{n\in\mathbb{Z}} (0,\infty) \times (2n\pi + \pi/2, 2n\pi + 3\pi/2)$$
$$p^{-1}(U_4) = \coprod_{n\in\mathbb{Z}} (0,\infty) \times (2n\pi + \pi, 2n\pi + 2\pi)$$

and it is a routine matter to check that p restricts to each of the sets in the above disjoint unions representing $p^{-1}(U_i)$ as a homeomorphism to U_i. For instance the inverse of the restriction of p to $(0,\infty) \times (2n\pi - \pi/2, 2n\pi + \pi/2)$ is given by

$$x + iy \mapsto \left(\sqrt{x^2 + y^2}, 2n\pi + \arctan(y/x)\right)$$

(recalling that by convention arctan *takes values between $-\pi/2$ and $\pi/2$).*

The fact that $p\colon \tilde{X} \to X$ is a covering space can be seen as an expression of the fact that, although one cannot uniquely and continuously assign the polar coordinate θ to all nonzero complex numbers at the same time, one do so if one restricts to a sufficiently small part of $\mathbb{C} \setminus \{0\}$.

Definition 2.3. Let $p\colon \tilde{X} \to X$ be a covering space and let $f : Y \to X$ be any continuous map. A **lift** of f via p is a continuous map $\tilde{f} : Y \to \tilde{X}$ such that $p \circ \tilde{f} = f$.

The question of when a lift of $f : Y \to X$ exists will be fundamental both to our proof of Theorem 1.9 and to our detailed study of covering spaces later on. For instance the fact that $\mathbb{C} \setminus \{0\}$ is not simply connected can be understood in terms of the fact that there are maps $S^1 \to \mathbb{C} \setminus \{0\}$ (such as the f_n mentioned earlier for $n \neq 0$—it might be instructive for you to think about why this is plausible) which do *not* have lifts via the cover $(r, \theta) \mapsto re^{i\theta}$. The question of the existence of lifts is in some sense a global question as opposed to a local one: locally in X lifts are easy to understand as the following lemma shows.

Lemma 2.4. *Let Z be a connected space, let $p\colon \tilde{X} \to X$ be a covering space, and let $f : Z \to X$ be continuous. Let $\{U_\alpha\}_{\alpha\in A}$ be an open cover of X as in Definition 2.1 with corresponding open subsets $V_{\alpha\beta} \subset p^{-1}(U_\alpha)$. Suppose that $f(Z) \subset U_\alpha$ for some α. Then:*

(i) *If $\tilde{f} : Z \to \tilde{X}$ is any lift of f, we must have $\tilde{f}(Z) \subset V_{\alpha\beta}$ for some β.*
(ii) *For each β there is a unique lift $\tilde{f}$ of f with the property that $\tilde{f}(Y) \subset V_{\alpha\beta}$.*

Proof. For (i), note that the various subsets $\tilde{f}^{-1}(V_{\alpha\beta}) \subset Z$ are open and mutually disjoint, and moreover their union is all of Z since $f = p \circ \tilde{f}$ has image contained in U_α and any point of $\tilde{X}$ that is mapped by p to U_α is contained in one of the $V_{\alpha\beta}$. But since Z is connected, for any collection of disjoint open subsets of Z whose union is all of Z, one of these sets must equal Z and the others must be empty. (i) immediately follows.

As for (ii), note that since $p|_{V_{\alpha\beta}} : V_{\alpha\beta} \to U_\alpha$ is a homeomorphism it has an inverse $p|_{V_{\alpha\beta}}^{-1}$. If $\tilde{f} : Z \to V_{\alpha\beta}$ lifts f, this means that $p|_{V_{\alpha\beta}} \circ \tilde{f} = f$, and so applying $p|_{V_{\alpha\beta}}^{-1}$ to both sides shows that $\tilde{f} = p|_{V_{\alpha\beta}}^{-1} \circ f$, proving the uniqueness of $\tilde{f}$. Conversely $\tilde{f} = p|_{V_{\alpha\beta}}^{-1} \circ f$ is obviously a lift of f, proving existence. □

The most important tool in our study of lifts will be the following:

Theorem 2.5 (Unique homotopy lifting property)**.** *Let Y be a locally connected space*[5] *and let $p\colon \tilde{X} \to X$ be a covering space. Let $F\colon [0,1] \times Y \to X$ be continuous and suppose that there is a continuous map $\tilde{f}_0\colon Y \to \tilde{X}$ such that $p \circ \tilde{f}_0(y) = F(0, y)$ for all $y \in Y$. Then there is a unique continuous $\tilde{F} : [0,1] \times Y \to \tilde{X}$ such that $p \circ \tilde{F} = F$ and such that $\tilde{F}(0, y) = \tilde{f}_0(y)$ for all $y \in Y$.*

In other words, this says that, for any homotopy F between maps $f_0, f_1 \colon Y \to X$ such that f_0 has a lift to $\tilde{X}$, the entire homotopy F has a lift to $\tilde{X}$ (in particular, f_1 has a lift). Moreover this lift of F is unique once the lift of f_0 is specified. The special case where Y is a one-point space is already nontrivial and important—it says that paths in X can be uniquely lifted to paths in $\tilde{X}$ upon specifying a starting point for the lift.

Proof of Theorem 2.5. Most of the work will be done by the following lemma (a special case of which already gives the case of Theorem 2.5 where Y is a one-point space):

Lemma 2.6. *Under the hypotheses of Theorem 2.5, any point $y \in Y$ has a connected neighborhood N_y with the following property. There exists a lift $\tilde{F}^y : [0,1] \times N_y \to \tilde{X}$ of $F|_{[0,1]\times N_y}$ such that $\tilde{F}^y(0,z) = \tilde{f}_0(z)$ for all $z \in N_y$. Moreover, for any $z \in N_y$, the* only *lift $\tilde{\gamma}\colon [0,1] \to \tilde{X}$ of the map $t \mapsto F(t,z)$ such that $\tilde{\gamma}(0) = \tilde{f}_0(z)$ is the map $t \mapsto \tilde{F}^y(t,z)$.*

Proof. Let $y \in Y$ and let U_α and $V_{\alpha\beta}$ be sets as in the definition of Definition 2.1. By the continuity of the map $t \mapsto F(t, y)$, the sets $\{t \in [0,1]||F(t,y) \in U_\alpha\}$ form an open cover of $[0,1]$. Using the compactness of $[0,1]$, one can then find[6] numbers $0 = t_0 < t_1 < \cdots < t_k = 1$ and indices $\alpha_0, \ldots, \alpha_{k-1}$ such that, for each $i \in \{0, \ldots, k-1\}$, $F([t_i, t_{i+1}] \times \{y\}) \subset U_{\alpha_i}$. By standard properties of the product topology (sometimes this is called the "tube lemma"), since $[t_i, t_{i+1}]$ is compact there is then, for each i, an open neighborhood $N_{y,i}$ of y such that

[5] recall that this means that, for any point $y \in Y$, any open neighborhood of y has an open subneighborhood which is connected

[6] Check this yourself if it's not clear.

$F([t_i,t_{i+1}]\times N_{y,i})\subset U_{\alpha_i}$. Then $\cap_{i=1}^{k}N_{y,i}$ is also an open neighborhood of y, and so since Y is assumed locally connected we can find a connected open subneighborhood of $\cap_{i=0}^{k-1}N_{y,i}$; denote such a subneighborhood by N_y. In particular we have

$$F([t_i,t_{i+1}]\times N_y)\subset U_{\alpha_i} \text{ for all } i$$

We will prove by induction that, for all $i\in\{0,\ldots,k\}$, there is a lift $\tilde{F}^{y,i}$ of $F|_{[0,t_i]\times N_y}$ which coincides at $t=0$ with $\tilde{f}_0|_{N_y}$, and moreover that for all $z\in N_y$ the only lift of $t\mapsto F(t,z)$ on the interval $[0,t_i]$ which coincides at $t=0$ with $\tilde{f}_0(z)$ is $t\mapsto \tilde{F}^{y,i}(t,z)$. Note that since $t_0=0$, the base step of this induction is a tautology. Meanwhile since $t_k=1$, the $i=k$ version of the statement is precisely the conclusion of the lemma. So to prove the lemma we just need to prove the inductive step; assume that, for some $i\in\{0,\ldots,k-1\}$, we have a lift $\tilde{F}^{y,i}\colon [0,t_i]\times N_y\to\tilde{X}$ satisfying the indicated properties; we need to extend this to a map with domain $[0,t_{i+1}]\times N_y$.

Now we have arranged that $F([t_i,t_{i+1}]\times N_y)\subset U_{\alpha_i}$. By Lemma 2.4(i) there is β such that our inductively-existing map $\tilde{F}^{y,i}$ has $\tilde{F}^{y,i}(\{t_i\}\times N_y)\subset V_{\alpha_i\beta}$. By Lemma 2.4(ii), there is a lift of $F|_{[t_i,t_{i+1}]\times N_y}$ to a map $\tilde{G}\colon [t_i,t_{i+1}]\times N_y\to V_{\alpha\beta}$, and moreover by the uniqueness part of Lemma 2.4(ii) the restriction $\tilde{G}|_{\{t_i\}\times N_y}$ must coincide with $\tilde{F}^{y,i}|_{\{t_i\}\times N_y}$. Consequently we may define a map $\tilde{F}^{y,i+1}\colon [0,t_{i+1}]\times N_y\to\tilde{X}$ by setting $\tilde{F}^{y,i+1}$ equal to $\tilde{F}^{y,i}$ on $[0,t_i]\times N_y$ and to $\tilde{G}$ on $[t_i,t_{i+1}]\times N_y$. The "pasting lemma" from point-set topology shows that $\tilde{F}^{y,i+1}$ is continuous. The fact that $p\circ\tilde{F}^{y,i+1}(t,z)=F(t,z)$ for all (t,z) in the domain of $\tilde{F}^{y,i+1}$ follows directly from the corresponding facts for $\tilde{F}^{y,i}$ and $\tilde{G}$, and moreover $\tilde{F}^{y,i+1}$ restricts appropriately to $\{0\}\times N_y$ since it is equal to $\tilde{F}^{y,i}$ there.

It remains to prove the uniqueness statement in the inductive hypothesis. Let $\tilde{\gamma}\colon [0,t_{i+1}]\to\tilde{X}$ be a continuous map such that $p\circ\tilde{\gamma}(t)=F(t,z)$ for all t, where z is any given point of N_y. We must show that $\tilde{\gamma}(t)=\tilde{F}^{y,i+1}(t,z)$ for all $t\in[0,t_{i+1}]$. If $t\le t_i$ then this follows from the inductive hypothesis, as $\tilde{F}^{y,i}$ is assumed to satisfy the corresponding property and $\tilde{F}^{y,i+1}|_{[0,t_i]\times N_y}=\tilde{F}^{y,i}$ by construction. In particular $\tilde{\gamma}(t_i)\in V_{\alpha_i\beta}$ where α_i and β are as in the previous paragraph. Consequently since $F(t,z)\in U_{\alpha_i}$ for all $t\in[t_i,t_{i+1}]$, we have $\tilde{\gamma}([t_i,t_{i+1}])\subset V_{\alpha_i\beta}$ by Lemma 2.4(i). So $\tilde{\gamma}$ and $t\mapsto\tilde{F}^{y,i+1}(t,z)$ both restrict to $[t_i,t_{i+1}]$ as lifts of $t\mapsto F(t,z)$ having image contained in $V_{\alpha_i\beta}$; hence by Lemma 2.4(ii) they are equal. So indeed $\tilde{\gamma}(t)=\tilde{F}^{y,i+1}(t,z)$ for all $t\in[0,t_{i+1}]$, completing the induction. □

Having proven Lemma 2.6 we now complete the proof of Theorem 2.5. The various open sets N_y as y varies through Y obviously cover Y (since $y\in N_y$), and for each of these we have a map $\tilde{F}^y\colon [0,1]\times N_y\to\tilde{X}$. If $z\in Y$ belongs to two different sets N_{y_1} and N_{y_2}, then the maps $t\mapsto\tilde{F}^{y_1}(t,z)$ and $t\mapsto\tilde{F}^{y_2}(t,z)$ are both lifts of $t\mapsto F(t,z)$ which send 0 to $\tilde{f}_0(z)$, so by the uniqueness part of Lemma 2.6 they are equal. So we can define a map $\tilde{F}\colon [0,1]\times Y\to\tilde{X}$ by setting $\tilde{F}(t,z)=\tilde{F}^y(t,z)$ for any y such that $z\in N_y$; as just explained this definition is independent of the particular choice of y with $z\in N_y$. The fact that $\tilde{F}$ is a continuous lift of F restricting at $t=0$ to $\tilde{f}_0$ is inherited directly from the corresponding facts for the $\tilde{F}^y$, completing the proof of existence in Theorem 2.5. Moreover if $\tilde{H}\colon [0,1]\times Y\to\tilde{X}$ were another map satisfying the same properties then for any $z\in Y$ the map $t\mapsto\tilde{H}(t,z)$ would, by Lemma 2.6, coincide with $t\mapsto\tilde{F}^y(t,z)$ for any y with $z\in N_y$, and hence would coincide with $t\mapsto\tilde{F}(t,z)$. This proves uniqueness in Theorem 2.5. □

We now use Theorem 2.5 to construct the function $\deg\colon [S^1,\mathbb{C}\setminus\{0\}]\to\mathbb{Z}$ in Theorem 1.9.

Proposition 2.7. *If $f: [0,1] \to \mathbb{C} \setminus \{0\}$ is any continuous map there are continuous functions $r: [0,1] \to (0,\infty)$ and $\theta: [0,1] \to \mathbb{R}$ such that $f(t) = r(t)e^{i\theta(t)}$ for all $t \in [0,1]$ Moreover any two such pairs of functions (r_1,θ_1) and (r_2,θ_2) have $r_1(t) = r_2(t)$ and $\theta_1(t) - \theta_2(t) = 2\pi n$ for all $t \in [0,1]$ and some $n \in \mathbb{Z}$ which is independent of t.*

Proof. The first statement follows directly from the existence part of Theorem 2.5 applied with Y equal to a point and with p equal to the cover $(0,\infty)\times\mathbb{R} \to \mathbb{C}\setminus\{0\}$ defined by $(r,\theta) \mapsto re^{i\theta}$. (One just needs to choose arbitrarily a pair $(r(0),\theta(0)) \in (0,\infty)\times\mathbb{R}$ such that $r(0)e^{i\theta(0)} = f(0)$ and then apply Theorem 2.5.) For the second part, if $r_1, r_2: [0,1] \to (0,\infty)$ and $\theta_1,\theta_2: [0,1] \to \mathbb{R}$ are continuous and both obey $r_1(t)e^{i\theta_1(t)} = r_2(t)e^{i\theta_2(t)} = f(t)$ for all t, setting $t = 0$ we see that $r_1(0) = r_2(0)$ and that there is $n \in \mathbb{Z}$ such that $\theta_1(0) = 2\pi n + \theta_2(0)$. Then the maps $t \mapsto (r_1(t), \theta_1(t) - 2n\pi)$ and $t \mapsto (r_2(t), \theta_2(t))$ are both lifts via p of the map $f: [0,1] \to \mathbb{C}\setminus\{0\}$, and they coincide at $t = 0$, so by the uniqueness part of Theorem 2.5 they coincide for all t. Thus $r_1(t) = r_2(t)$ and $\theta_1(t) - 2n\pi = \theta_2(t)$ $\square$

Recall that, by definition, $S^1 = \frac{[0,1]}{0\sim 1}$. Where $\pi: [0,1] \to S^1$ is the quotient projection, any continuous map $f: S^1 \to \mathbb{C}\setminus\{0\}$ gives rise to a continuous map $f\circ\pi: [0,1] \to \mathbb{C}\setminus\{0\}$. By Proposition 2.7 we can write $f\circ\pi(t) = r(t)e^{i\theta(t)}$ where $r: [0,1] \to (0,\infty)$ and $\theta: [0,1] \to \mathbb{R}$ are continuous. Since $\pi(0) = \pi(1)$, we have $r(0)e^{i\theta(0)} = r(1)e^{i\theta(1)}$, forcing $r(0) = r(1)$ and

$$\frac{\theta(1) - \theta(0)}{2\pi} \in \mathbb{Z}$$

Moreover if (r_2,θ_2) is a different choice of continuous functions with $f\circ\pi(t) = r_2(t)e^{i\theta_2(t)}$, by Proposition 2.7 we have $\theta(t) = \theta_2(t) + 2\pi n$ for some $n \in \mathbb{Z}$ which is independent of t. In particular $\theta_2(0) - \theta(0) = \theta_2(1) - \theta(1)$, from which it follows by rearranging terms that

$$\frac{\theta_2(1) - \theta_2(0)}{2\pi} = \frac{\theta(1) - \theta(0)}{2\pi}$$

Thus we may **define the degree of a continuous map** $f: S^1 \to \mathbb{C}\setminus\{0\}$ as

$$d(f) = \frac{\theta(1) - \theta(0)}{2\pi}$$

for any continuous maps $r: [0,1] \to (0,\infty), \theta: [0,1] \to \mathbb{R}$ such that $f(\pi(t)) = r(t)e^{i\theta(t)}$ for all $t \in [0,1]$. The above discussion shows that this definition is independent of the particular choice of r,θ obeying the required property, and that $d(f) \in \mathbb{Z}$.

Proposition 2.8. *If $f_0, f_1: S^1 \to \mathbb{C}\setminus\{0\}$ are two homotopic maps then $d(f_0) = d(f_1)$.*

Proof. If $\underline{F}: [0,1]\times S^1 \to \mathbb{C}\setminus\{0\}$ is a homotopy from f_0 to f_1, then evidently the map $F: [0,1]\times[0,1] \to \mathbb{C}\setminus\{0\}$ defined by $F(s,t) = \underline{F}(s,\pi(t))$ is a homotopy between the maps $f_0\circ\pi, f_1\circ\pi: [0,1] \to \mathbb{C}\setminus\{0\}$, where again $\pi: [0,1] \to S^1$ is the quotient projection.

If $(r_0,\theta_0): [0,1] \to (0,\infty)\times\mathbb{R}$ is a lift via p of $f_0\circ\pi$, the homotopy lifting property gives a lift via p of F restricting at $s = 0$ to (r_0,θ_0); we may write this lift as $(R,\Theta): [0,1]\times[0,1] \to (0,\infty)\times\mathbb{R}$ for some continuous functions (R,Θ). So $(R(0,\cdot),\Theta(0,\cdot)) = (r_0,\theta_0)$, while $(R(1,\cdot),\Theta(1,\cdot))$ is a lift of $f_1\circ\pi$. So

$$d(f_0) = \frac{\Theta(0,1) - \Theta(0,0)}{2\pi} \qquad d(f_1) = \frac{\Theta(1,1) - \Theta(1,0)}{2\pi}$$

Moreover for all $t \in [0,1]$ it holds that $R(s,1)e^{i\Theta(s,1)} = R(s,0)e^{i\Theta(s,0)} = \underline{F}(s,0)$, so for all $s \in [0,1]$ we have $\frac{\Theta(s,1)-\Theta(s,0)}{2\pi} \in \mathbb{Z}$. Thus $s \mapsto \frac{\Theta(s,1)-\Theta(s,0)}{2\pi}$ is a continuous, $\mathbb{Z}$-valued function which equals

$d(f_0)$ at $s = 0$ and $d(f_1)$ at $s = 1$. But the only continuous functions from $[0,1]$ to $\mathbb{Z}$ are constants, so $d(f_0) = d(f_1)$. $\square$

Proof of Theorem 1.9. By Proposition 2.8 we obtain a well-defined function $\deg\colon [S^1, \mathbb{C}\setminus\{0\}] \to \mathbb{Z}$ by setting $\deg(c) = d(f)$ where f is any continuous map representing a given homotopy class c. We will show that deg is a bijection. Now that we know that deg is well-defined, surjectivity is immediate: for $n \in \mathbb{Z}$ the function $f_n(t) = e^{2\pi i n t}$ lifts to a map $t \mapsto (1, 2\pi n t)$ and so has degree $\frac{2\pi n - 0}{2\pi} = n$. (Note that this already completes the proof that $\mathbb{C} \setminus \{0\}$ is not homotopy equivalent to the contractible space $\mathbb{C}$.)

To prove injectivity we need to prove that any two maps having the same degree are homotopic. Let $f_0, f_1\colon S^1 \to \mathbb{C}\setminus\{0\}$ be two maps with the same degree, say $n \in \mathbb{Z}$. We can then write $f_0 \circ \pi(t) = r_0(t)e^{i\theta_0(t)}$ and $f_1 \circ \pi(t) = r_1(t)e^{i\theta_1(t)}$ for continuous $r_0, r_1\colon [0,1] \to (0,\infty)$ and $\theta_0, \theta_1\colon [0,1] \to \mathbb{R}$ with $r_0(1) = r_0(0)$, $r_1(1) = r_1(0)$, and $\theta_1(1) - \theta_1(0) = \theta_0(1) - \theta_0(0) = 2\pi n$.

Define $R\colon [0,1]\times[0,1] \to (0,\infty)$ by $R(s,t) = (1-s)r_0(t) + s r_1(t)$ and $\Theta\colon [0,1]\times[0,1] \to \mathbb{R}$ by $\Theta(s,t) = (1-s)\theta_0(t) + s\theta_1(t)$. For any $s \in [0,1]$ we have

$$\begin{aligned}\Theta(s,1) - \Theta(s,0) &= (1-s)\theta_0(1) + s\theta_1(1) - (1-s)\theta_0(0) - s\theta_1(0) \\ &= (1-s)(\theta_0(1) - \theta_0(0)) + s(\theta_1(1) - \theta_1(0)) = (1-s)2\pi n + s 2\pi n = 2\pi n.\end{aligned}$$

Likewise $R(s,1) = R(s,0)$ for all $s \in [0,1]$, so we have $R(s,1)e^{i\Theta(s,1)} = R(s,0)e^{i\Theta(s,0)}$ for all $s \in [0,1]$. This allows us to define $\underline{F}\colon [0,1]\times S^1 \to \mathbb{C}\setminus\{0\}$ by $\underline{F}(s,t) = R(s,t)e^{i\Theta(s,t)}$, and it is straightforward to see that $\underline{F}$ is a homotopy from f_0 to f_1.

$\square$

Exercise 2.9. Let us define the degree of a continuous map $g\colon S^1 \to S^1$ to be the degree of the composition $g \circ f_1\colon S^1 \to \mathbb{C}\setminus\{0\}$ where $f_1(t) = e^{2\pi i t}$. (So if you think of S^1 as identified with the unit circle T via f_1 then the degree of g is $\frac{\theta(1)-\theta(0)}{2\pi}$ where $g(t) = r(t)e^{i\theta(t)}$ for continuous $r\colon [0,1] \to (0,\infty)$ and $\theta\colon [0,1] \to \mathbb{R}$, just like the definition of the degree of a map to $\mathbb{C}\setminus\{0\}$.)

(a) Prove that if a continuous map $g\colon S^1 \to S^1$ is not surjective then its degree is zero.

(b) Prove that if $g\colon S^1 \to S^1$ is a homeomorphism then its degree is *not* zero. (Actually its degree is ± 1 but we'll need the fundamental group to prove this.)

(Hint for both parts: A map $S^1 \to S^1$ has degree zero if and only if it is homotopic to a constant map.)

3. Some applications of Theorem 1.9

Our determination of the set of homotopy classes of maps from S^1 to $\mathbb{C}\setminus\{0\}$ has some classic consequences going beyond the original motivating question of distinguishing $\mathbb{C}\setminus\{0\}$ from $\mathbb{C}$.

First, let $T = \{z \in \mathbb{C} | |z| = 1\}$ denote the unit circle, so by Exercise 1.7 T is homotopy equivalent to $\mathbb{C}\setminus\{0\}$. So by Theorem 1.9 and Exercise 1.5(c), the set of homotopy classes $[S^1, T]$ is countably infinite. Meanwhile if $D = \{z \in \mathbb{C} | |z| = 1\}$ is the closed unit disk, a straightforward modification of the proof of Exercise 1.7 or Proposition 1.8 shows that $[S^1, D]$ consists of a single element.

Definition 3.1. Let X be a topological space and let $Y \subset X$. A *retraction* from X to Y is a continuous map $r\colon X \to Y$ such that $r(y) = y$ for all $y \in Y$.

Proposition 3.2. *There does **not** exist any retraction $r\colon D \to T$.*

Proof. Suppose that $r\colon D \to T$ were a retraction. Another way of expressing this is that, where $i\colon T \to D$ is the inclusion and 1_T is the identity on T, $r \circ i = 1_T$. Now we have induced maps $r_*\colon [S^1, D] \to [S^1, T]$ and $i_*\colon [S^1, T] \to [S^1, D]$, and by Exercise 1.5 (c) these obey $r_* \circ i_* =$

$(r\circ i)_* = 1_{T*}$. But i_* is a map to the one-element set $[S^1, D]$, so $r_*\circ i_*$ must have image consisting of only one element. On the other hand 1_{T*} is the identity on the infinite set $[S^1, T]$ so its image is infinite, a contradiction. □

This leads to one of the most famous results of early algebraic topology (though actually we've proven it without any algebra):

Theorem 3.3 (Brouwer Fixed Point Theorem, 1912). *If $f: D \to D$ is any continuous map then there is a point $x \in D$ such that $f(x) = x$.*

(Note that if D were replaced by $\mathbb{R}^2$, or $D \setminus \{0\}$, or the open unit disk, this result would be false, as you should be able to convince yourself via specific counterexamples.)

Proof. We will prove this by assuming for contradiction that $f: D \to D$ is a continuous map with no fixed points and then using this to construct a retraction $r: D \to T$, in violation of Proposition 3.2. Specifically, assuming that $f(x) \neq x$ for all $x \in D$, let $r: D \to T$ be the map defined by letting $r(x)$ be the point on T obtained as follows: $r(x)$ is the unique point on T lying on the open ray that starts at $f(x)$ and passes through x. [7] From this geometric description it's clear that $r(x) = x$ if $x \in T$. This contradicts Proposition 3.2. □

We also obtain a quick proof of a result with which you are likely familiar from other courses (the proof also perhaps justifies the use of the term "degree" in the proof of Theorem 1.9):

Theorem 3.4 (Fundamental theorem of algebra). *If $f(z) = \sum_{j=0}^{n} a_j z^j$ is a polynomial having no zeros in $\mathbb{C}$ then f is constant.*

Proof. Let $f(z) = \sum_{j=0}^{n} a_j z^j$ be as in the statement of the theorem; without loss of generality we can assume that $a_n \neq 0$ and then our goal is to show that $n = 0$.

Define a map $\gamma: S^1 \to \mathbb{C} \setminus \{0\}$ by $\gamma(t) = f(e^{2\pi i t})$. Let us compute the degree of γ in two ways.

First define $G: [0,1] \times S^1 \to \mathbb{C} \setminus \{0\}$ by $G(s, e^{it}) = f(se^{2\pi i t})$. Then G is a homotopy between the constant map to $a_0 = f(0)$ and the loop γ (it takes values in $\mathbb{C} \setminus \{0\}$ because f has no zeros in the closed unit disk). So by the homotopy-invariance of the degree (Proposition 2.8) γ evidently has degree 0.

On the other hand let us define $H: [0,1] \times S^1 \to \mathbb{C} \setminus \{0\}$ by

$$H(s,t) = s^n f(s^{-1}e^{2\pi i t}) = \sum_{j=0}^{n} a_j s^{n-j} e^{2\pi i j t}$$

(of course the formula in the middle is ill-defined at s=0, so really we are defining this using the formula on the right, but the formula in the middle makes clear that this takes values in $\mathbb{C} \setminus \{0\}$ for $s \neq 0$ since f has no zeros outside the open unit disk). At $s = 0$ the only nonzero term on the right is where $j = n$; on the other hand we have $H(1, \cdot) = \gamma$. So γ is homotopic to the map $t \mapsto a_n e^{2\pi i n t}$ which has degree n.

Since γ both has degree n and has degree 0 we have shown that $n = 0$.

□

[7] If you want a formula, it's $r(x) = f(x) + t(x)(x - f(x))$ where

$$t(x) = \frac{x \cdot (f(x) - x) + \sqrt{(x \cdot (f(x) - x))^2 + 4|x - f(x)|^2(1 - |f(x)|^2)}}{2|x - f(x)|^2}$$

Recall that the n-dimensional sphere is by definition $\{(x_1,\dots,x_n) \in \mathbb{R}^{n+1} | \sum_1^{n+1} x_i^2 = 1\}$, endowed with the subspace topology from $\mathbb{R}^{n+1}$.[8] There is an important homeomorphism from S^n to itself, the *antipodal map* $A\colon S^n \to S^n$, defined by

$$A(x_1,\dots,x_{n+1}) = (-x_1,\dots,-x_{n+1})$$

Theorem 3.5 (Borsuk–Ulam theorem (two dimensions))**.** *If $f : S^2 \to \mathbb{R}^2$ is any continuous map then there is $x \in S^2$ such that $f(x) = f(Ax)$.*

Proof. As with the proof of the Brouwer fixed point theorem we will assume that no such x exists and use this to construct a different map having contradictory properties. Namely, identifying $\mathbb{R}^2$ with $\mathbb{C}$ as usual, if $f(x) \neq f(Ax)$ for all x then we obtain a map $F\colon S^2 \to \mathbb{C} \setminus \{0\}$ by setting

$$F(x) = f(x) - f(Ax)$$

This new map evidently has the property that, for all $x \in S^2$,

$$F(Ax) = -F(x) \tag{1}$$

Consider the loop $\gamma\colon S^1 \to S^2$ given by

$$\gamma(t) = (\cos(2\pi t), \sin(2\pi t), 0)$$

so that γ goes once around the equator. We will consider the degree of the composition $F \circ \gamma\colon S^1 \to \mathbb{C} \setminus \{0\}$.

Observe that γ is homotopic to a constant map: $\Gamma\colon [0,1] \times S^1 \to S^2$ defined by $\Gamma(s,t) = (s\cos(2\pi t), s\sin(2\pi t), \sqrt{1-s^2})$ defines a homotopy from the constant map to the north pole $(1,0,0)$ to γ. So by Exercise 1.4 $F \circ \gamma$ is homotopic to the constant map to $F(1,0,0)$, and so by Proposition 2.8 the degree of $F \circ \gamma$ is the same as the degree of this constant map, namely zero.

Since $d(F \circ \gamma) = 0$ we may write $F(\gamma(t)) = r(t)e^{i\theta(t)}$ where $r\colon [0,1] \to (0,\infty)$ and $\theta\colon [0,1] \to \mathbb{R}$ are continuous functions such that $\theta(1) = \theta(0)$. Now γ has the property that $A(\gamma(t)) = \gamma(t+1/2)$ for $t \in [0,1/2]$, so by (1) we have, for all $t \in [0,1/2]$,

$$r(t+1/2)e^{i\theta(t+1/2)} = -r(t)e^{i\theta(t)}$$

So taking magnitudes shows that $r(t+1/2) = r(t)$, and so $e^{i\theta(t+1/2)} = -e^{i\theta(t)}$. This implies that, for some $n(t) \in \mathbb{Z}$, $\theta(t+1/2) - \theta(t) = (2n(t)+1)\pi$. But since θ is continuous, so is the map $t \mapsto n(t)$, and so since $n(t) \in \mathbb{Z}$ it must be that $n(t) = n$ for some (constant) $n \in \mathbb{Z}$ and we have $\theta(t+1/2) - \theta(t) = (2n+1)\pi$ for all $t \in [0,1/2]$.

Thus we obtain

$$\theta(1) - \theta(0) = (\theta(1) - \theta(1/2)) + (\theta(1/2) - \theta(0)) = (4n+2)\pi$$

Since $n \in \mathbb{Z}$ this contradicts the fact that (since $d(F \circ \gamma) = 0$) $\theta(1) = \theta(0)$.

□

4. The fundamental group(oid)

What we have done up to now should be enough to convince you that studying sets constructed out of the paths and loops in a space can yield significant information about the space. We can go further than this by performing algebraic operations on paths and loops.

For example, we have shown that $\mathbb{C} \setminus \{0\}$ can be topologically distinguished from $\mathbb{C}$ by the fact that the set $[S^1, \mathbb{C} \setminus \{0\}]$ of homotopy classes of loops in $\mathbb{C} \setminus \{0\}$ is countably infinite, while the corresponding set for $\mathbb{C}$ has just one element. But what happens if we delete another point,

[8] Apologies for the slight abuse of notation that I previously defined S^1 to be not exactly this but rather something homeomorphic to it.

and consider $\mathbb{C}\setminus\{0,1\}$? It turns out to be possible to show that $[S^1,\mathbb{C}\setminus\{0,1\}]$ is also countably infinite; this successfully distinguishes it from $\mathbb{C}$, but gives no information about whether or not it is homeomorphic to $\mathbb{C}\setminus\{0\}$. However, after we construct the fundamental group we will find that the fundamental groups of $\mathbb{C}\setminus\{0\}$ and $\mathbb{C}\setminus\{0,1\}$, while again both countably infinite, are not isomorphic, with the latter much larger than the former from the standpoint of group theory (even though there exists a non-homomorphic bijection between them).

To inject algebra into this story we need to have a natural binary operation which one can perform on paths and/or loops. There does not seem to be such an operation on arbitrary loops, but there is a "partially defined" binary operation on paths arising from the observation that if one has a way of getting from point A to point B and also a way of getting from point B to point C then one has a way of getting from point A to point C. This gives rise to something called the *fundamental groupoid*, and then specializing to the case that $A=B=C$ (so one is dealing with paths from A to A, *i.e.* loops "based at" A) gives the fundamental group.

If X is a topological space and $x,y\in X$ let us denote

$$\mathcal{P}_X(x,y)=\{\gamma\colon [0,1]\to X|\gamma \text{ is continuous, } \gamma(0)=x,\ \gamma(1)=y\},$$

i.e. $\mathcal{P}_X(x,y)$ is the set of paths from x to y. The remark in the previous paragraph reflects the fact that there is a *concatenation operation*

$$\begin{aligned}\mathcal{P}_X(x,y)\times\mathcal{P}_X(y,z)&\to\mathcal{P}_X(x,z)\\ (\alpha,\beta)&\mapsto\alpha*\beta\end{aligned}$$

where for $\alpha\in\mathcal{P}_X(x,y),\beta\in\mathcal{P}_X(y,z)$ the path $\alpha*\beta$ from x to z is defined by

$$\alpha*\beta(t)=\begin{cases}\alpha(2t) & 0\le t\le 1/2\\ \beta(2t-1) & 1/2\le t\le 1\end{cases}$$

Thus $\alpha*\beta$ goes along α (but twice as fast) to get from x to y, then goes along β (but twice as fast) to get from y to z; since $\alpha(1)=\beta(0)=y$ it is clear from the pasting lemma that $\alpha*\beta$ is continuous. There is also a *reversal operation*

$$\begin{aligned}\mathcal{P}_X(x,y)&\to\mathcal{P}_X(y,x)\\ \gamma&\mapsto\bar{\gamma}\end{aligned}$$

defined by

$$\bar{\gamma}(t)=\gamma(1-t)$$

The sets $\mathcal{P}_X(x,y)$ are generally too large to readily give usable information by themselves, and don't particularly satisfy nice algebraic properties (like forming a groupoid). We will instead consider homotopy classes of paths between points of X. In general, given two continuous maps $f_0,f_1\colon Y\to Z$ and a subset $A\subset Y$, a **homotopy rel** A from f_0 to f_1 is a continuous map $F\colon [0,1]\times Y\to Z$ which is a homotopy from f_0 to f_1 (so $F(0,y)=f_0(y)$, $F(1,y)=f_1(y)$) with the additional property that $F(t,a)$ is independent of t for every $a\in A$. (So in particular this obviously forces $f_1|_A=f_0|_A$.) It is straightforward to see that saying that $f_0\sim f_1$ if and only if there is a homotopy rel A from f_0 to f_1 defines an equivalence relation on the set of maps $Y\to Z$ (or for that matter on the set of maps $Y\to Z$ whose restriction to A coincides with some given map $A\to Z$).

Now for $x,y\in X$ we define

$$\Pi_X(x,y)=\frac{\mathcal{P}_X(x,y)}{\sim}$$

where for $\gamma_0,\gamma_1\in\mathcal{P}_X(x,y)$ we say $\gamma_0\sim\gamma_1$ if and only if there is a homotopy rel $\{0,1\}$ from $\gamma_0\colon [0,1]\to X$ to $\gamma_1\colon [0,1]\to X$. In other words $\Pi_X(x,y)$ is the set of equivalence classes of paths from x to y, where paths γ_0,γ_1 are considered equivalent provided that there is a continuous $\Gamma\colon [0,1]\times[0,1]\to X$ such that for all t $\Gamma(0,t)=\gamma_0(t)$ and $\Gamma(1,t)=\gamma_1(t)$, and for all s $\Gamma(s,0)=x$ and $\Gamma(s,1)=y$.

For $\gamma\in\mathscr{P}_X(x,y)$ we denote its equivalence class in $\Pi_X(x,y)$ by $[\gamma]$. Generic elements of $\Pi_X(x,y)$ which are not given as equivalence classes of specific loops will be given names like a or c.

Exercise 4.1. Prove that if we instead had said that $\gamma_0,\gamma_1\in\mathscr{P}_X(x,y)$ were equivalent if there were *any* homotopy from γ_0 to γ_1 (not necessarily a homotopy rel $\{0,1\}$) then all elements of $\mathscr{P}_X(x,y)$ would have been equivalent.

Exercise 4.2. Let $r\colon [0,1]\to[0,1]$ be any continuous map with $r(0)=0$ and $r(1)=1$. Prove that for any $\gamma\in\mathscr{P}_X(x,y)$ we have $[\gamma\circ r]=[\gamma]$. (Hint: Exploit the fact that $[0,1]$ is a convex set.)

Exercise 4.3. Assume we have $\alpha_0,\alpha_1\in\mathscr{P}_X(x,y)$ and $\beta_0,\beta_1\in\mathscr{P}_X(y,z)$ such that $[\alpha_0]=[\alpha_1]$ and $[\beta_0]=[\beta_1]$. Prove that, where $*$ is the concatenation operation defined earlier,

$$[\alpha_0*\beta_0]=[\alpha_1*\beta_1]$$

In view of Exercise 4.3 we may now define a concatenation operation on (relative, though in future I'll leave this word out) homotopy classes of paths:

$$\begin{aligned}\Pi_X(x,y)\times\Pi_X(y,z)&\to\Pi_X(x,z)\\ (a,b)&\mapsto a*b:=[\alpha*\beta]\quad\text{for any }\alpha,\beta\text{ with }[\alpha]=a,\ [\beta]=b\end{aligned}$$

Moreover this operation is associative in the following sense:

Proposition 4.4. *For $w,x,y,z\in X$ and $a\in\Pi_X(w,x), b\in\Pi_X(x,y), c\in\Pi_X(y,z)$ we have*

$$a*(b*c)=(a*b)*c$$

Proof. Let $\alpha\in\mathscr{P}_X(w,x)$, $\beta\in\mathscr{P}_X(x,y)$, $\gamma\in\mathscr{P}_X(y,z)$ be representatives of the classes a,b,c respectively. By definition $a*(b*c)$ is represented by the path $\alpha*(\beta*\gamma)$ from w to z where

$$\beta*\gamma(t)=\begin{cases}\beta(2t) & 0\le t\le 1/2\\ \gamma(2t-1) & 1/2\le t\le 1\end{cases}$$

and so

$$\alpha*(\beta*\gamma)(t)=\beta*\gamma(t)=\begin{cases}\alpha(2t) & 0\le t\le 1/2\\ \beta*\gamma(2t-1) & 1/2\le t\le 1\end{cases}=\begin{cases}\alpha(2t) & 0\le t\le 1/2\\ \beta(4t-2) & 1/2\le t\le 3/4\\ \gamma(4t-3) & 3/4\le t\le 1\end{cases}$$

Similarly

$$(\alpha*\beta)*\gamma(t)=\begin{cases}(\alpha*\beta)(2t) & 0\le t\le 1/2\\ \gamma(2t-1) & 1/2\le t\le 1\end{cases}=\begin{cases}\alpha(4t) & 0\le t\le 1/4\\ \beta(4t-1) & 1/4\le t\le 1/2\\ \gamma(2t-1) & 1/2\le t\le 1\end{cases}$$

(In particular $\alpha*(\beta*\gamma)$ is *not* the same path as $(\alpha*\beta)*\gamma$; we are just proving that they are homotopic.) Defining $r\colon [0,1]\to[0,1]$ by

$$r(t)=\begin{cases}t/2 & 0\le t\le 1/2\\ t-1/4 & 1/2\le t\le 3/4\\ 2t-1 & 3/4\le t\le 1\end{cases}$$

we see from the above formulas that, for all $t \in [0,1]$, $((\alpha * \beta) * \gamma)(r(t)) = (\alpha * (\beta * \gamma))(t)$. So since the "reparametrization function" $r\colon [0,1] \to [0,1]$ is continuous it follows from Exercise 4.2 that

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

□

If one specializes to the case that $x = y$ we get the *fundamental group* at x, $\Pi_X(x,x)$ (or more commonly written as $\pi_1(X,x)$) consisting of homotopy classes of paths from x to x, *i.e.* of loops based at x. The name fundamental group will be justified shortly; evidently Proposition 4.4 already shows (something stronger than) associativity, and the following will imply the existence of an identity:

Proposition 4.5. *For each $x \in X$, where $e_X \in \Pi_X(x,x)$ denotes the homotopy class of the constant path $\gamma_x\colon [0,1] \to X$ defined by $\gamma_x(t) = x$ for all $t \in [0,1]$, the following holds:*

(i) *If $w \in X$ and $a \in \Pi_X(w,x)$ then $a * e_x = a$*
(ii) *If $y \in X$ and $b \in \Pi_X(x,y)$ then $e_X * b = b$.*

Proof. For (i), if $a = [\alpha]$ where $\alpha \in \mathscr{P}_X(w,x)$, we have $a * e_x = [\alpha * \gamma_x]$ where

$$\alpha * \gamma(t) = \begin{cases} \alpha(2t) & 0 \le t \le 1/2 \\ x & 1/2 \le t \le 1 \end{cases}$$

Thus where $r\colon [0,1] \to [0,1]$ is defined by $r(t) = 2t$ for $0 \le t \le 1/2$ and $r(t) = 1$ for $1/2 \le t \le 1$ we have $\alpha * \gamma_x = \alpha \circ r$. Thus by Proposition 4.2, $[\alpha * \gamma_x] = [\alpha]$, proving (i). Essentially the same argument (but with $r(t) = 0$ for $0 \le t \le 1/2$ and $r(t) = 2t - 1$ for $1/2 \le t \le 1$) proves (ii). □

Likewise we have a generalization of inverses (in the sense that the case $x = y$ in the proposition below gives the existence of inverses in the group $\Pi_X(x,x)$); recall that for $\gamma \in \mathscr{P}_X(x,y)$ we defined $\bar{\gamma} \in \mathscr{P}_X(y,x)$ by $\bar{\gamma}(t) = \gamma(1-t)$.

Proposition 4.6. *For $x, y \in X$ and $\gamma \in \mathscr{P}_X(x,y)$ we have*

$$[\gamma] * [\bar{\gamma}] = e_x \in \Pi_X(x,x) \qquad \textit{and} \qquad [\bar{\gamma}] * [\gamma] = e_y \in \Pi_X(y,y)$$

Proof. We will prove the second equality; the proof of the first is very similar and is left to the reader (or for that matter can be proven by formally manipulating the second equality). $[\bar{\gamma}] * [\gamma] \in \Pi_X(y,y)$ is the homotopy class of the loop $\bar{\gamma} * \gamma$ defined by

$$\bar{\gamma} * \gamma(t) = \begin{cases} \gamma(1-2t) & 0 \le t \le 1/2 \\ \gamma(2t-1) & 1/2 \le t \le 1 \end{cases}$$

Define $\Gamma\colon [0,1] \times [0,1] \to X$ by

$$\Gamma(s,t) = \begin{cases} \gamma(1-2t) & 0 \le t \le s/2 \\ \gamma(1-s) & s/2 \le t \le 1 - s/2 \\ \gamma(2t-1) & 1 - s/2 \le t \le 1 \end{cases}$$

In particular $\Gamma(0,t)\Gamma(s,0) = \Gamma(s,1) = \gamma(1) = y$ for all s and t, while $\Gamma(1,t) = (\bar{\gamma} * \gamma)(t)$ for all t. Thus Γ gives a homotopy rel $\{0,1\}$ from the constant path γ_y to $\bar{\gamma} * \gamma$, proving that $[\bar{\gamma}] * [\gamma] = e_y$. □

Propositions 4.4, 4.5, and 4.6 comprise a proof that the fundamental groupoid Π_X, which comprises the data of the sets $\Pi_X(x,y)$ for all $x, y \in X$ with the concatenation operation $*$ and the distinguished elements $e_x \in \Pi_X(x,x)$, is indeed a groupoid—in (concise) categorical terms this means that Π_X is a small category in which every morphism is an isomorphism, namely one has:

- A set of "objects" of the groupoid, which in this case is just the underlying set of the space X.
- For each $x, y \in X$ an associated "set of morphisms from x to y," in this case $\Pi_X(x, y)$, together with maps $*\colon \Pi_X(x, y)\times\Pi_X(y, z) \to \Pi_X(x, z)$ obeying the associativity relation $a * (b * c) = (a * b) * c$.
- For each $x \in X$ a distinguished "identity morphism" $e_x \in \Pi_X(x, x)$, obeying the property in Proposition 4.5.[9]
- Each element $[\gamma] \in \Pi_X(x, y)$ has an "inverse element" in $\Pi_X(y, x)$ as in Proposition 4.6.

In this groupoid as in any other, for any object x the "endomorphisms" $\Pi_X(x, x)$ form a group, as one sees by setting $w = x = y = z$ in Propositions 4.4, 4.5, and 4.6. Groups are simpler to study than groupoids, so this will tend to be our primary object of study:

Definition 4.7. If X is a topological space and $x_0 \in X$, the *fundamental group of X at x_0* is the set $\Pi_X(x_0, x_0)$ of homotopy classes of paths from x_0 to x_0, endowed with the concatenation operation $*$.

On the other hand, algebra that we can do on more general homotopy classes of paths in $\Pi_X(x, y)$ will sometimes be useful—this makes sense, since in exploring a space one would presumably like to follow paths wherever they go rather than always having to return to a single basepoint x_0.

Proposition 4.8. *If X and Y are spaces with $x_0 \in X$, and if $f : X \to Y$ is any continuous map, there is a well-defined induced map $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ defined by, for $\gamma \in \mathscr{P}_X(x_0, x_0)$,*

$$f_*([\gamma]) = [f \circ \gamma]$$

Moreover f_* is a homomorphism of groups. Also, if $g\colon Y \to Z$ is another continuous map of topological spaces then the induced map $(g \circ f)_*\colon \pi_1(X, x_0) \to \pi_1(Z, g(f(x_0)))$ obeys

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

Proof. Exercise (this is very similar to Exercise 1.5—the main novelty is the statement that f_* is a homomorphism, but if you think about what this means you shouldn't find it hard). □

In categorical language Proposition 4.8 says that the fundamental group is a functor from the category of *pointed* topological spaces to the category of groups, where a pointed topological space is a topological space together with a specifically-chosen "basepoint" in that space, and the only allowed maps from one pointed topological space (X, x_0) another one (Y, y_0) are those continuous maps $X \to Y$ which map x_0 to y_0. In particular Proposition 4.8 shows that if there is a homeomorphism $f : X \to Y$ taking x_0 to y_0 then $\pi_1(X, x_0)$ must be isomorphic to $\pi_1(Y, y_0)$, since the induced map $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, y_0)$ will have inverse given by $(f^{-1})_*\colon \pi_1(Y, f(x_0)) \to \pi_1(X, x_0)$. So π_1 is an invariant of *pointed* topological spaces. This need to arbitrarily specify and then keep track of basepoints is annoying, but inevitable unless you want to analyze the entire groupoid structure. However we at least have the following:

Proposition 4.9. *Suppose that $x_0, x_1 \in X$ belong to the same path component. Then there is an isomorphism $\phi\colon \pi_1(X, x_0) \to \pi_1(X, x_1)$. In particular, if X is path-connected, then the isomorphism type of $\pi_1(X, x)$ is independent of the choice of $x \in X$.*

---

[9] These three items define a "small category," with the term "small" referring to the fact that the collection of objects is a set rather than something too big to be a set, such as the class of all sets or the class of all topological spaces.

Proof. Let $\gamma \in \mathscr{P}_X(x_0,x_1)$, which then determines a path $\bar{\gamma} \in \mathscr{P}_X(x_1,x_0)$ such that $[\gamma] \in \Pi_X(x_0,x_1)$ and $[\bar{\gamma}] \in \mathscr{P}_X(x_1,x_0)$ obey the "inverse" property of Proposition 4.6. Using the concatenation operation in the whole fundamental groupoid, and recalling that by definition $\pi_1(X,x_j) = \Pi_X(x_j,x_j)$ for $j = 0,1$, we define $\phi\colon \pi_1(X,x_0) \to \pi_1(X,x_1)$ by

$$\phi(a) = [\bar{\gamma}] * a * [\gamma]$$

(of course the very act of writing the above formula without parentheses requires Proposition 4.4). To see that ϕ is a homomorphism we simply note that since $[\gamma] * [\bar{\gamma}] = e_{x_0}$ is the identity in $\pi_1(X,x_0)$,

$$\phi(a) * \phi(b) = [\bar{\gamma}] * a * [\gamma] * [\bar{\gamma}] * b * [\gamma] = [\bar{\gamma}] * a * e_{x_0} * b * [\gamma] = \phi(a * b).$$

If we define $\psi\colon \pi_1(X,x_1) \to \pi_1(X,x_0)$ by

$$\psi(c) = [\gamma] * c * [\bar{\gamma}],$$

repeatedly applying Propositions 4.4,4.5, and 4.6 shows that

$$\phi(\psi(c)) = [\bar{\gamma}] * [\gamma] * c * [\bar{\gamma}] * [\gamma] = e_{x_1} * c * e_{x_1} = c$$

for all $c \in \pi_1(X,x_1)$ and likewise

$$\phi(\psi(a)) = [\gamma] * [\bar{\gamma}] * a * [\gamma] * [\bar{\gamma}] = e_{x_0} * a * e_{x_0} = a$$

for all $a \in \pi_1(X,x_0)$. This ϕ is bijective, with inverse ψ. So since ϕ was already shown to be a homomorphism it is an isomorphism. □

Note however that the isomorphism ϕ depended on the path γ; if we had chosen a path representing a different homotopy class we might have gotten a different isomorphism. One accordingly says that the fundamental groups at different basepoints of a path-connected space are isomorphic, but not canonically isomorphic. One could apply the construction in the proof to the case where $x_0 = x_1$, giving an isomorphism $\pi_1(X,x_0) \to \pi_1(X,x_0)$—in this case the homotopy class $[\gamma]$ itself belongs to $\pi_1(X,x_0)$ and the isomorphism is given be conjugation by the element $[\gamma]$. It could very well be the case that this isomorphism is not the identity, since we will eventually see examples of spaces with nonabelian fundamental groups.

In any case, we now have a new example of an invariant of path-connected topological spaces, even if it's slightly less functorial than we might like: the isomorphism type of the fundamental group. To reiterate, we choose a point $x_0 \in X$ and define $\pi_1(X,x_0)$ to be the set of equivalence classes of paths $\gamma\colon [0,1] \to X$ with $\gamma(0) = \gamma(1) = x_0$, where two such γ_0, γ_1 are equivalent if they are homotopic rel $\{0,1\}$, *i.e.* if there is a continuous map $\Gamma\colon [0,1] \times [0,1]$ with $\Gamma(s,0) = \Gamma(s,1) = x$ for all $s \in [0,1]$, and $\Gamma(0,t) = \gamma_0(t)$ and $\Gamma(1,t) = \gamma_1(t)$ for all $t \in [0,1]$. If X isn't path connected one likewise has an invariant given by the function which assigns to any path component of X the isomorphism type of the fundamental group at a point in that path component.

Remark 4.10. Of course, a path $\gamma\colon [0,1] \to X$ having $\gamma(0) = \gamma(1) = x_0$ is essentially the same thing as continuous map $\eta\colon S^1 \to X$ having $\eta(0) = x_0$. (As before $S^1 = \frac{[0,1]}{0\sim 1}$; where $\pi\colon [0,1] \to S^1$ is the quotient projection γ and η are related by $\gamma = \eta \circ \pi$.) Moreover the paths γ_0, γ_1 from x_0 to x_0 are homotopic rel $\{0,1\}$ if and only if the corresponding loops $\eta_0, \eta_1\colon S^1 \to X$ are homotopic rel $\{0\}$. So it is entirely equivalent to define the fundamental group $\pi_1(X,x_0)$ as consisting of equivalence classes up to homotopy rel $\{0\}$ of continuous maps $\eta\colon S^1 \to X$ with $\eta(0) = x_0$. (The more common term for "homotopy rel $\{0\}$" is "based homotopy.") The group operation is of course given by the same concatenation operation as before, just regarding the domains to be S^1 rather than $[0,1]$: namely $[\eta_0]*[\eta_1]$ is the based homotopy class of the map $\eta_0*\eta_1\colon S^1\to X$ defined by

$$(\eta_0*\eta_1)(t)\begin{cases}\eta_0(2t) & 0\le t\le 1/2\\ \eta_1(2t-1) & 1/2\le t\le 1\end{cases}$$

We'll often use this alternative characterization.

Exercise 4.11. As mentioned before, a *pointed topological space* is a pair (X,x_0) where X is a topological space and $x_0\in X$, and a *based map* $f\colon (X,x_0)\to(Y,y_0)$ between two pointed topological spaces is by definition a continuous map $f\colon X\to Y$ such that $f(x_0)=y_0$. We say that two based maps $f_0,f_1\colon (X,x_0)\to(Y,y_0)$ are *based homotopic* if there is a homotopy $F\colon [0,1]\times X\to Y$ such that $F(t,x_0)=y_0$ for all $t\in[0,1]$.

(a) Prove that if $f_0,f_1\colon (X,x_0)\to(Y,y_0)$ are based homotopic then the induced maps $(f_0)_*$ and $(f_1)_*$ from $\pi_1(X,x_0)$ to $\pi_1(Y,y_0)$ are equal.

(b) A *based homotopy equivalence* $f\colon (X,x_0)\to(Y,y_0)$ is by definition a based map $f\colon (X,x_0)\to(Y,y_0)$ such that there exists a based map $g\colon (Y,y_0)\to(X,x_0)$ such that $g\circ f$ and $f\circ g$ are based homotopic to the identities $(X,x_0)\to(X,x_0)$ and $(Y,y_0)\to(Y,y_0)$, respectively. Prove that if f is a based homotopy equivalence then the induced map $f_*\colon \pi_1(X,x_0)\to\pi_1(Y,y_0)$ is an isomorphism.

(In fact, with a bit more effort one can show that any homotopy equivalence $f\colon X\to Y$, based or not, induces an isomorphism on fundamental groups; see [H, Proposition 1.18].)

We'll now compute a few examples.

Exercise 4.12. Let $S\subset\mathbb{R}^n$ be any subset which is *star-shaped* in the sense that for all $\vec{v}\in S$ and $t\in[0,1]$ it holds that $t\vec{v}\in S$. (So in particular S is path-connected and contains $\vec{0}$.) Prove that $\pi_1(S,\vec{0})$ is the trivial group.

Proposition 4.13. *There is a group isomorphism $\pi_1(\mathbb{C}\setminus\{0\},1)\cong\mathbb{Z}$.*

Proof. We already did most of the work here when we proved Theorem 1.9 classifying homotopy classes of maps from $S^1\to\mathbb{C}\setminus\{0\}$, but the question answered by the current proposition is just a little bit different. The proof of Theorem 1.9 shows that every continuous map $f\colon S^1\to\mathbb{C}\setminus\{0\}$ is homotopic to the map $f_n(t)=e^{2\pi int}$ for one and only one value of $n\in\mathbb{Z}$. Conveniently, the maps f_n all have $f_n(0)=1$, so they each represent classes $[f_n]\in\pi_1(\mathbb{C}\setminus\{0\},1)$. Moreover since Proposition 2.8 shows that if $m\neq n$ then f_m is not even homotopic to f_n, much less based homotopic, the classes $[f_n]$ are all distinct. So we have a well-defined injection

$$\begin{aligned}\phi\colon \mathbb{Z}&\to\pi_1(\mathbb{C}\setminus\{0\},1)\\ n&\mapsto[f_n]\end{aligned}$$

We will now show that ϕ is surjective. Suppose that $f\colon S^1\to\mathbb{C}\setminus\{0\}$ is any continuous map with $f(0)=1$. We know that f is homotopic to some f_n; we need to improve this to the statement that f is based homotopic to f_n. Let $F\colon [0,1]\times S^1\to\mathbb{C}\setminus\{0\}$ be a homotopy from f to f_n. We have $F(0,0)=F(1,0)=1$, but for $s\notin\{0,1\}$, $F(s,0)$ may not be 1. But if we define $\hat{F}\colon [0,1]\times S^1\to\mathbb{C}\setminus\{0\}$ by

$$\hat{F}(s,t)=\frac{F(s,t)}{F(s,0)}$$

then $\hat{F}$ will indeed give a based homotopy from f to f_n, proving that $[f]=[f_n]\in\pi_1(\mathbb{C}\setminus\{0\},1)$.

So ϕ is a bijection; it remains to show that ϕ is a group isomorphism, which is equivalent to the statement that $[f_m]*[f_n]=[f_{m+n}]$. Now what we have done up to this point shows that, quite generally, two loops $S^1\to\mathbb{C}\setminus\{0\}$ based at 1 will represent the same element of

$\pi_1(\mathbb{C}\setminus\{0\},1)$ if and only if they have the same degree. Now the loop $f_m * f_n\colon S^1 \to \mathbb{C}\setminus\{0\}$ is given by $f_m * f_n(t) = e^{i\theta(t)}$ for the continuous function $\theta\colon [0,1]\to\mathbb{R}$ defined by

$$\theta(t) = \begin{cases} 4\pi m t & 0 \le t \le 1/2 \\ 2\pi m + 2\pi n(2t-1) & 1/2 \le t \le 1 \end{cases}$$

So $\theta(1)-\theta(0) = 2\pi(m+n)$, and the degree of $f_m * f_n$ is $m+n$, which is the same as the degree of f_{m+n}. □

Now the unit circle $T \subset \mathbb{C}$ (which of course is homeomorphic to S^1) is easily seen to be based homotopy equivalent to $\mathbb{C}\setminus\{0\}$ (with basepoint 1 for both spaces): the inclusion $T \to \mathbb{C}\setminus\{0\}$ has based homotopy inverse $z \mapsto \frac{z}{|z|}$. So we have:

Corollary 4.14. $\pi_1(S^1,0) \cong \mathbb{Z}$

The story is different for the other spheres, however:

Proposition 4.15. *For $n \ge 2$, where $x_N \in S^n$ is the "north pole" $(0,\dots,0,1)$, $\pi_1(S^n,x_N)$ is the trivial group.*

Proof. Let $x_S \in S^1$ denote the south pole: $x_S = (0,\dots,0,-1)$. The main observation is that stereographic projection gives a homeomorphism $\pi\colon S^n\setminus\{x_S\} \to \mathbb{R}^n$ and $\mathbb{R}^n$ is based homotopy equivalent to the one-point space $\{\pi(x_N)\}$, so a loop contained in $S^n\setminus\{x_S\}$ can easily be basepoint-preservingly shrunk to x_N.[10] In formulas define $\pi\colon S^n\setminus\{x_S\}\to\mathbb{R}^n$ by

$$\pi(x_1,\dots,x_{n+1}) = \left(\frac{x_1}{1+x_{n+1}},\dots,\frac{x_n}{1+x_{n+1}}\right)$$

which has inverse

$$\pi^{-1}(y_1,\dots,y_n) = \left(\frac{2\|\vec{y}\|^2}{\|\vec{y}\|^2+1}y_1,\dots,\frac{2\|\vec{y}\|^2}{\|\vec{y}\|^2+1}y_n,\frac{\|\vec{y}\|^2-1}{\|\vec{y}\|^2+1}\right)$$

Then if $\gamma\colon (S^1,0)\to(S^n,x_N)$ has $\gamma(t)\in S^n\setminus\{x_S\}$ for all $t\in S^1$ we can define $\Gamma\colon [0,1]\times S^1\to S^n$ by $\Gamma(s,t) = \pi^{-1}((1-s)\pi(\gamma(t)))$ to obtain a homotopy from γ to the constant map to x_N; since $\Gamma(s,0) = \pi^{-1}((1-s)\pi(x_N)) = \pi^{-1}(\vec{0}) = x_N$ for all s, Γ is in fact a based homotopy and so $[\gamma]$ is the identity element of $\pi_1(S^n,x_N)$.

So to prove the triviality of $\pi_1(S^n,x_N)$ (for $n\ge 2$) it suffices to show that, if $n\ge 2$, any path $\gamma\colon [0,1]\to S^n$ with $\gamma(0)=\gamma(1)=x_N$ is homotopic rel $\{0,1\}$ to a map γ' with image in $S^n\setminus\{x_S\}$, since we will then have $[\gamma]=[\gamma']$ and the previous paragraph shows that $[\gamma']$ is the identity.

So let $\gamma\in\mathscr{P}_{S^n}(x_N,x_N)$ be arbitrary, and let U denote the southern hemisphere $\{x_1,\dots,x_{n+1}\in S^n|x_{n+1}<0\}$. Let $\partial U = \{(x_1,\dots,x_{n+1})\in S^n|x_{n+1}=0\}$. Then $\gamma^{-1}(U)$ is an open subset of $[0,1]$ containing neither 0 nor 1, and so can be written as a disjoint union of open intervals (a,b); necessarily the endpoints a and b of these intervals are not contained in $\gamma^{-1}(U)$ (otherwise (a,b) would not be disjoint from the other open intervals in the union), and so $\gamma(a),\gamma(b)\in\partial U$ by continuity. The closed, hence compact set $\gamma^{-1}(\{x_S\})$ is contained in $\gamma^{-1}(U)$, so we can find finitely many of these intervals that cover $\gamma^{-1}(\{x_S\})$; say the intervals are $(a_1,b_1),\cdots,(a_k,b_k)$ where $0<a_1<b_1<a_2<\cdots<b_k<1$. For notational convenience write $b_0=0$ and $a_{k+1}=1$; then we have

$$[0,1] = \left(\cup_{i=0}^k[b_i,a_{i+1}]\right)\cup\left(\cup_{i=1}^k[a_i,b_i]\right)$$

[10] In fact the same argument applies with any point other than x_N in place of x_S, allowing us to contract any loop $\gamma\colon (S^1,0)\to(S^n,x_N)$ which is not surjective, but the existence of space-filling curves shows that an additional argument is still needed.

where, for all i, $\gamma|_{[b_i,a_{i+1}]}$ has image contained in $S^n \setminus \{x_S\}$, while $\gamma|_{[a_i,b_i]}$ has image meeting x_S and contained in $\bar{U} = U \cup \partial U$, with $\gamma(a_i), \gamma(b_i) \in \partial U$.

Now the fact that $n \geq 2$ implies that ∂U (which is evidently homeomorphic to S^{n-1} is path-connected. Our loop γ' will be obtained from γ by, for each i, setting $\gamma'|_{[b_i,a_{i+1}]}$ equal to $\gamma|_{[b_i,a_{i+1}]}$ while setting $\gamma'_{[a_i,b_i]}$ equal to an arbitrary continuous map $\gamma' \colon [a_i,b_i] \to \partial U$ such that $\gamma'(a_i) = \gamma(a_i)$ and $\gamma'(b_i) = \gamma(b_i)$. Since $x_S \notin \partial U$, γ' is clearly a continuous map with image in $S^n \setminus \{x_S\}$, and so is based homotopic to the constant map to x_N as we have already seen. It only remains to check that γ is based homotopic to γ'. For this it clearly suffices to see that $\gamma|_{[a_i,b_i]}$ is homotopic rel $\{a_i, b_i\}$ to $\gamma'|_{[a_i,b_i]}$. But this is easy to check: for instance there is a homeomorphism between $\bar{U}$ to the closed unit disk in $\mathbb{R}^n$ (for instance by taking the projection to the first n factors) and in terms of this homeomorphism one obtains a homotopy F_i rel $\{a_i, b_i\}$ from γ to γ' by setting $F_i(s,t)$ equal to the position at time s along a line segment from $\gamma(t)$ to $\gamma'(t)$. (Alternately, to give a more formal argument in terms of the fundamental groupoid, since $\bar{U}$ is homeomorphic to the unit disk we know that $\pi_1(\bar{U}, \gamma(a_i)) = \Pi_{\bar{U}}(\gamma(a_i), \gamma(a_i))$ has just one element, in view of which the same is true of $\Pi_{\bar{U}}(\gamma(a_i), \gamma(b_i))$ since for an arbitrary path $\eta \in \mathscr{P}_U(\gamma(a_i), \gamma(b_i))$ the map $c \mapsto c * [\eta]$ gives a bijection from $\Pi_{\bar{U}}(\gamma(a_i), \gamma(a_i))$ to $\Pi_{\bar{U}}(\gamma(a_i), \gamma(b_i))$ with inverse $d \mapsto d * [\bar{\eta}]$.)

Summing up, we have shown first that any loop based at x_N in S^n which does not intersect x_S is based homotopic to the constant loop, and then that any loop based at x_N in S^n at all is based homotopic to one which does not intersect x_S, in view of which $\pi_1(S^n, x_N)$ consists only of the class of the constant loop. □

Corollary 4.16. *For $n \geq 3$, $\mathbb{R}^2$ is not homeomorphic to $\mathbb{R}^n$.*

(Of course, you already showed in Exercise 1.1 that $\mathbb{R}^2$ is not homeomorphic to $\mathbb{R}$, and the same argument should show that $\mathbb{R}$ is not homeomorphic to $\mathbb{R}^n$ for $n \geq 3$ either. Distinguishing the various $\mathbb{R}^n$ for $n \geq 3$ from each other will have to wait until we learn about homology later in the course.)

Proof. If $\underline{f} \colon \mathbb{R}^2 \to \mathbb{R}^n$ were a homeomorphism then by composing $\underline{f}$ with the homeomorphism $\vec{x} \mapsto \vec{x} - \underline{f}((0,0))$ of $\mathbb{R}^n$ we would obtain a homeomorphism $f \colon \mathbb{R}^2 \to \mathbb{R}^n$ sending $\vec{0}$ to $\vec{0}$. So f would restrict as a homeomorphism $f \colon \mathbb{R}^2 \setminus \{\vec{0}\} \to \mathbb{R}^n \setminus \{\vec{0}\}$ Hence $\mathbb{R}^2 \setminus \{\vec{0}\}$ and $\mathbb{R}^n \setminus \{\vec{0}\}$ would have isomorphic fundamental groups. But the inclusion $S^{n-1} \to \mathbb{R}^n \setminus \{\vec{0}\}$ is a homotopy equivalence with homotopy inverse $\vec{x} \mapsto \frac{\vec{x}}{\|\vec{x}\|}$. So $\mathbb{R}^n \setminus \{\vec{0}\}$ has trivial fundamental group for $n \geq 3$ by Proposition 4.15, whereas the fundamental group of $\mathbb{R}^2 \setminus \{\vec{0}\}$ is nontrivial by Proposition 4.13. □

Exercise 4.17. Recall that earlier we defined a space X to be simply connected if and only if $[S^1, X]$ has only one element. Prove that X is simply connected if and only if X has only one path component and $\pi_1(X, x_0)$ is the trivial group for (one and hence any) $x_0 \in X$.

Here is a simple general observation:

Proposition 4.18. *Let X and Y be two spaces with $x_0 \in X$, $y_0 \in Y$. Then $\pi_1(X \times Y, (x_0, y_0))$ is the direct product of the groups $\pi_1(X, x_0)$ and $\pi_1(Y, y_0)$.*

Proof. Let $p_X \colon X \times Y \to X$ and $p_Y \colon X \times Y \to Y$ denote the projections. We then have induced homomorphisms $p_{X*} \colon \pi_1(X \times Y, (x_0, y_0)) \to \pi_1(X, x_0)$ and $p_{Y*} \colon \pi_1(X \times Y, (x_0, y_0)) \to \pi_1(Y, y_0)$. So we can form the homomorphism $\phi \colon \pi_1(X \times Y, (x_0, y_0)) \to \pi_1(X, x_0) \times \pi_1(Y, y_0)$ by setting $\phi(c) = (p_{X*}(c), p_{Y*}(c))$.

We will show that ϕ is an isomorphism. To see that it is surjective, note that any element of $\pi_1(X,x_0)\times\pi_1(Y,y_0)$ can be represented as $([\gamma_X],[\gamma_Y])$ for based loops $\gamma_X\colon(S^1,0)\to(X,x_0)$ and $\gamma_Y\colon(S^1,0)\to(Y,y_0)$, and the definition obviously implies that, where $\gamma\colon(S^1,0)\to(X\times Y,(x_0,y_0))$ is defined by $\gamma(t)=(\gamma_X(t),\gamma_Y(t))$, we have $\phi([\gamma])=([\gamma_X],[\gamma_Y])$.

As for injectivity (equivalently, triviality of the kernel), if $\gamma\colon(S^1,0)\to(X\times Y,(x_0,y_0))$ has the property that $\phi([\gamma])$ is the identity, this means that $p_X\circ\gamma$ is based homotopic to the constant map to x_0 in X and that $p_Y\circ\gamma$ is based homotopic to the constant map to y_0 in Y. Letting $F_X\colon[0,1]\times S^1\to X$ and $F_Y\colon[0,1]\times S^1\to Y$ be the based homotopies whose existences are asserted in the previous sentence, the map $F\colon[0,1]\times S^1\to X\times Y$ defined by $F(s,t)=(F_X(s,t),F_Y(s,t))$ is a based homotopy from γ to the constant map to (x_0,y_0) in $X\times Y$, proving that $[\gamma]$ is the identity in $\pi_1(X\times Y,(x_0,y_0))$. □

For instance, it follows from Propositions 4.18 and 4.13 that $\pi_1(S^1\times S^1,(0,0))\cong\mathbb{Z}\times\mathbb{Z}$, with the two integers representing the number of times a loop "wraps around" the two S^1 factors of $S^1\times S^1$. Together with Proposition 4.15 this answers a question raised at the very start of these notes (in the third paragraph): you should be able to convince yourself that (the surface of) a donut is homeomorphic to $S^1\times S^1$, while a muffin is homeomorphic to S^2, so they have distinct fundamental groups and so are not homeomorphic.

By induction one can see that $(S^1)^n$ has fundamental group $\mathbb{Z}^n$; one might ask what other groups can arise. As a matter of fact *every* group can; a proof of this based on van Kampen's theorem (our next topic) appears in [H, Proposition 1.28]. Here is an important example of a space whose fundamental group is nontrivial but finite.

Exercise 4.19. One of several equivalent ways of defining *real projective space* $\mathbb{R}P^n$ is as the quotient of the sphere S^n by the equivalence relation that identifies $\vec{x}$ with $-\vec{x}$ for all $\vec{x}\in S^n\subset\mathbb{R}^{n+1}$. In particular $\mathbb{R}P^0$ is a one-point space and $\mathbb{R}P^1$ can be shown to be homeomorphic to S^1.

(a) Prove that the quotient map $p\colon S^n\to\mathbb{R}P^n$ is a covering space.

(b) From now on assume $n\geq 2$. Let $x_0\in S^n$ and let $\gamma\colon[0,1]\to\mathbb{R}P^n$ be continuous with $\gamma(0)=\gamma(1)=p(x_0)$. So according to the $Y=\{*\}$ case of Theorem 2.5 γ has a unique lift via p, denoted $\tilde{\gamma}\colon[0,1]\to S^n$, such that $\tilde{\gamma}(0)=x_0$. Prove (using Theorem 2.5 and Proposition 4.15) that, γ represents the identity element of $\pi_1(\mathbb{R}P^n,p(x_0))$ *if and only if* $\tilde{\gamma}(1)=x_0$.

(c) Prove that $\pi_1(\mathbb{R}P^n,p(x_0))$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ for $n\geq 2$.

5. Pushouts and the van Kampen theorem

The van Kampen theorem will allow us to compute the fundamental group of a space X by covering X by two suitable open sets U_0 and U_1, based on the fundamental groups of U_0,U_1, and $U_0\cap U_1$. (In [H, Section 1.2] there is a more general version that applies to certain covers by arbitrarily many open sets, but the case of a cover by two sets is the one that is most often used and I would say that it has a more elegant statement and proof.) The theorem as I will formulate it will be a sort of group-theoretic version of the following simple fact from point set topology.

Proposition 5.1. *Let X be a space and let $U_0,U_1\subset X$ be open sets such that $X=U_0\cup U_1$. For $k\in\{0,1\}$ let $j_k\colon U_0\cap U_1\to U_k$ and $i_k\colon U_k\to X$ be the inclusions. Then given any topological space Y and continuous maps $\phi_0\colon U_0\to Y$ and $\phi_1\colon U_1\to Y$ such that $\phi_0\circ j_0=\phi_1\circ j_1$, there is a unique continuous map $\phi\colon X\to Y$ such that $\phi\circ i_0=\phi_0$ and $\phi\circ i_1=\phi_1$.*

The proposition is more succinctly summarized by the following diagram (where all triangles and squares should be understood to be commutative):

$$\begin{tikzcd}
U_0 \cap U_1 \arrow[r, "j_0"] \arrow[d, "j_1"] & U_0 \arrow[d, "i_0"] \arrow[ddr, "\phi_0"] & \\
U_1 \arrow[r, "i_1"] \arrow[drr, "\phi_1"'] & X \arrow[dr, dashed, "\exists!\phi"] & \\
 & & Y
\end{tikzcd}$$

Proof. If $\phi: X \to Y$ is any function (continuous or not) satisfying $\phi \circ i_0 = \phi_0$, we must have $\phi(x) = \phi_0(x)$ whenever $x \in U_0$, and likewise the condition that $\phi \circ i_1 = \phi_1$ implies that $\phi(x) = \phi_1(x)$ whenever $x \in U_1$. Moreover the assumption that $\phi_0 \circ j_0 = \phi_1 \circ j_1$ shows that if $x \in U_0 \cap U_1$ then $\phi_0(x) = \phi_1(x)$. So there is a well-defined function $\phi: X \to Y$ given by

$$\phi(x) = \begin{cases} \phi_0(x) & x \in U_0 \\ \phi_1(x) & x \in U_1 \end{cases}$$

and this is the only function of any kind obeying $\phi \circ i_0 = \phi_0$ and $\phi \circ i_1 = \phi_1$. So the proposition will follow as soon as we show that ϕ is continuous. To see this, if $V \subset Y$ is open we have $\phi^{-1}(V) = \phi_0^{-1}(V) \cup \phi_1^{-1}(V)$. Now for $k \in \{0, 1\}$, ϕ_k was assumed continuous, so $\phi_k^{-1}(V)$ is relatively open in U_k. Since U_k is in turn open in X this implies that $\phi_k^{-1}(V)$ is open in X (for the fact that $\phi_k^{-1}(V)$ is relatively open in U_k amounts to the statement that there is an open $G_k \subset X$ so that $\phi_k^{-1}(V) = G_k \cap U_k$). Thus $\phi^{-1}(V)$ is a union of two open sets and so is open. □

We now abstract the behavior in the previous proposition into a general definition, which we give simultaneously in the category of topological spaces and in the category of groups.

Definition 5.2. Let A be a topological space (respectively, a group), and let $j_0: A \to B$ and $j_1: A \to C$ be continuous maps to other topological spaces B and C (respectively, homomorphisms to other groups B and C). A **pushout** of the diagram

$$\begin{tikzcd}
A \arrow[r, "j_0"] \arrow[d, "j_1"] & B \\
C &
\end{tikzcd}$$

is a space (respectively, group) D together with continuous maps (respectively, homomorphisms) $i_0: B \to D$ and $i_1: C \to D$ such that $i_0 \circ j_0 = i_1 \circ j_1$ and satisfying the following "universal property": for any other space (resp. group) Y with continuous maps (resp. homomorphisms) $\phi_0: B \to Y$ and $\phi_1: C \to Y$ obeying $\phi_0 \circ j_0 = \phi_1 \circ j_1$ there is a unique continuous map (resp. homomorphism) $\phi: D \to Y$ such that $\phi \circ i_0 = \phi_0$ and $\phi \circ i_1 = \phi_1$. In this case the commutative diagram

$$\begin{tikzcd}
A \arrow[r, "j_0"] \arrow[d, "j_1"] & B \arrow[d, "i_0"] \\
C \arrow[r, "i_1"] & D
\end{tikzcd}$$

is called a **pushout square**.

Again, this definition is summarized by the diagram

$$\begin{tikzcd} A \arrow[r, "j_0"] \arrow[d, "j_1"] & B \arrow[d, "i_0"] \arrow[ddr, "\phi_0"] & \\ C \arrow[r, "i_1"] \arrow[drr, "\phi_1"] & D \arrow[dr, dashed, "\exists!\phi"] & \\ & & Y \end{tikzcd}$$

(where either A, B, C, D, and Y are topological spaces and the arrows are continuous maps or A, B, C, D, and Y are groups and the arrows are homomorphisms).

Example 5.3. *According to Proposition 5.1, if X is a space with $X = U_0 \cup U_1$ where U_0 and U_1 are open the diagram*

$$\begin{tikzcd} U_0 \cap U_1 \arrow[r] \arrow[d] & U_0 \arrow[d] \\ U_1 \arrow[r] & X \end{tikzcd}$$

where all arrows are inclusion maps is a pushout square.

Example 5.4. *For a rather simple-minded group-theoretic example, where 1 denotes the trivial group, note that for any other group G there is a unique homomorphism $G \to 1$* ***and*** *a unique homomorphism $1 \to G$, with the image of the latter consisting only of the identity in G. In view of this it is easy (and left to the reader) to check that*

$$\begin{tikzcd} G \arrow[r] \arrow[d] & 1 \arrow[d] \\ 1 \arrow[r] & 1 \end{tikzcd}$$

is a pushout square.

Here is a more complicated example:

Exercise 5.5. Let U and V be spaces, $A \subset U$ a subspace, and $g : A \to V$ a continuous map, and form the "adjunction space"

$$U \cup_g V = \frac{U \coprod V}{a \sim g(a) \text{ for } a \in A}$$

(*i.e.* $U \cup_g V$ is the quotient space formed from the disjoint union $U \coprod V$ by the equivalence relation given by saying that $a \in U$ is equivalent to $b \in V$ iff $a \in A$ and $b = g(a)$ (and by requiring the relation to be reflexive, symmetric, and transitive, so in particular if $a_1, a_2 \in A$ with $g(a_1) = g(a_2)$ then $a_1 \sim a_2$)). Let $j_U : U \to U \cup_g V$ and $j_V : V \to U \cup_g V$ be the maps obtained by composing the inclusion of U or V into $U \coprod V$ with the quotient projection to $U \cup_g V$.

(a) Prove that j_V is a homeomorphism onto its image in $U \cup_g V$. (Said differently, V appears in an obvious way as a subset of $U \cup_g V$, and you need to show that the resulting subspace topology is the same as the original topology on V.)

(b) Where $i\colon A \to U$ is the inclusion, prove that the diagram

$$\begin{tikzcd} A \arrow[r, "i"] \arrow[d, "g"'] & U \arrow[d, "j_U"] \\ V \arrow[r, "j_V"] & U \cup_g V \end{tikzcd}$$

is a pushout square.

(c) Given any map $g_0 : A_0 \to V$ between topological spaces, apply the above construction with $U = A_0 \times [0,1]$, $A = A_0 \times \{0\}$, and $g(a,0) = g_0(a)$. The resulting adjunction space

$$M_{g_0} := (A_0 \times [0,1]) \cup_g V$$

is called the "mapping cylinder" of g_0. Prove that in this case the inclusion $j_V : V \to M_{g_0}$ is a homotopy equivalence.

An important fact about pushouts is that they are unique:

Proposition 5.6. *If A is a topological space, $j_0\colon A \to B$ and $j_1\colon A \to C$ are continuous maps, and*

$$\begin{tikzcd} A \arrow[r, "j_0"] \arrow[d, "j_1"'] & B \arrow[d, "i_0"] \\ C \arrow[r, "i_1"] & D \end{tikzcd}$$

and

$$\begin{tikzcd} A \arrow[r, "j_0"] \arrow[d, "j_1"'] & B \arrow[d, "h_0"] \\ C \arrow[r, "h_1"] & E \end{tikzcd}$$

are both pushout squares then there is a homeomorphism $\phi\colon D \to E$. The same statement also holds with "topological space," "continuous maps," and "homeomorphism" replaced respectively by "group," "homomorphism," and "group isomorphism."

Proof. Applying the universal property for D with $Y = D'$, and then for D' with $Y = D$, gives rise to a commutative diagram

$$\begin{tikzcd} A \arrow[r, "j_0"] \arrow[d, "j_1"'] & B \arrow[d, "i_0"'] \arrow[ddr, "h_0"] \arrow[dddrr, "i_0"] & & \\ C \arrow[r, "i_1"] \arrow[drr, "h_1"'] \arrow[ddrrr, "i_1"'] & D \arrow[dr, dashed, "\phi"] & & \\ & & E \arrow[dr, dashed, "\phi'"] & \\ & & & D \end{tikzcd}$$

In particular the composition $\phi' \circ \phi\colon D \to D$ obeys $(\phi' \circ \phi) \circ i_0 = i_0$ and $(\phi' \circ \phi) \circ i_1 = i_1$. But where 1_D is the identity, we of course also have $1_D \circ i_0 = i_0$ and $1_D \circ i_1 = i_1$. Applying the uniqueness part of the universal property to the case $Y = D$, $\phi_0 = i_0$, $\phi_1 = i_1$ then implies that $\phi' \circ \phi = 1_D$.

The same argument with D and E interchanged shows that $\phi \circ \phi' = 1_E$. Thus $\phi\colon D \to E$ is a homeomorphism (or group isomorphism) with inverse ϕ'.

□

Thus, going back to the earlier examples, *any* pushout of a diagram

$$\begin{array}{ccc} G & \longrightarrow & 1 \\ \downarrow & & \\ 1 & & \end{array}$$

must be given by the trivial group, and any topological space obtained as a pushout of a diagram

$$\begin{array}{ccc} A & \xrightarrow{i} & U \\ \downarrow{\scriptstyle g} & & \\ V & & \end{array}$$

is homeomorphic to $U \cup_g V$ (in particular this applies to X as in Proposition 5.1). More broadly, stating that a group appears as the bottom right corner of a pushout square suffices to uniquely specify that group up to isomorphism (and likewise for topological spaces up to homeomorphism), provided that the other groups and the upper and left homomorphisms in the diagram are known. With that in mind, here is the van Kampen theorem:

Theorem 5.7 (Van Kampen). *Let X be a topological space and $U_0, U_1 \subset X$ open with $X = U_0 \cup U_1$ and U_0, U_1, and $U_0 \cap U_1$ all path connected. For $k \in \{0,1\}$ let $j_k\colon U_0 \cap U_1 \to U_k$ and $i_k\colon U_k \to X$ be the inclusions. Then for $x_0 \in U_0 \cap U_1$ the diagram of induced maps*

$$\begin{array}{ccc} \pi_1(U_0 \cap U_1, x_0) & \xrightarrow{j_{0*}} & \pi_1(U_0, x_0) \\ \downarrow{\scriptstyle j_{1*}} & & \downarrow{\scriptstyle i_{0*}} \\ \pi_1(U_1, x_0) & \xrightarrow{i_{1*}} & \pi_1(X, x_0) \end{array}$$

is a pushout square.

Proof. To start we choose, for every $x \in X$, an arbitrary path $\eta_x\colon [0,1] \to X$ with $\eta(0) = x_0$ and $\eta(1) = x$, such that if $x \in U_0 \cap U_1$ then $\eta([0,1]) \subset U_0 \cap U_1$, if $x \in U_0$ then $\eta([0,1]) \subset U_0$, and if $x \in U_1$ then $\eta([0,1]) \subset U_1$. (Of course this is possible since we assumed $U_0, U_1, U_0 \cap U_1$ where all path connected.) Also we take η_{x_0} to be the constant path at x_0.

If $[a,b] \subset [0,1]$ and $\gamma\colon [0,1] \to X$ we will abuse notation slightly by denoting by $\gamma|_{[a,b]}$ the map $[0,1] \to X$ given by $\gamma|_{[a,b]}(t) = \gamma(a + (b-a)t)$. Thus $\gamma|_{[a,b]} \in \mathscr{P}_X([\gamma(a), \gamma(b)])$.

If the map $\phi\colon \pi_1(X, x_0) \to G$ is to satisfy the required properties we will now determine how it must evaluate on the homotopy class of an arbitrary loop $\gamma\colon [0,1] \to X$ based at x_0.

We have an open cover $[0,1] = \gamma^{-1}(U_0) \cap \gamma^{-1}(U_1)$ where $\gamma^{-1}(U_0)$ and $\gamma^{-1}(U_1)$ can each be expressed as a disjoint union of open intervals; by passing to a finite subcover by these intervals and then shrinking the intervals slightly we obtain

$$0 = t_0 < t_1 < t_2 < \cdots < t_m = 1$$

where for each i there is $\epsilon_i \in \{0,1\}$ such that $\gamma([t_{i-1}, t_i]) \subset U_{\epsilon_i}$. Evidently we have (in the fundamental groupoid Π_X) an identity

$$[\gamma] = [\gamma|_{[t_0,t_1]}] * [\gamma|_{[t_1,t_2]}] * \cdots * [\gamma|_{[t_{m-1},t_m]}]$$

Here $[\gamma|_{[t_{i-1},t_i]}] \in \Pi_X(\gamma(t_{i-1}),\gamma(t_i))$. Now by Proposition 4.6 the concatenation $[\bar{\eta}_{\gamma(t_i)}]*[\eta_{\gamma(t_i)}]$ acts as the identity in $\Pi_X(\gamma(t_i),\gamma(t_i))$ for each i, as for that matter do $[\eta_{\gamma(t_0)}]$ and $[\bar{\eta}_{\gamma(t_m)}]$ since these were chosen to be the homotopy class of the constant path at $x_0 = \gamma(t_0) = \gamma(t_m)$. As a consequence of this (and of the associativity of concatenation in the fundamental groupoid) we have

$$[\gamma] = \bigstar_{i=1}^m \left([\eta_{\gamma(t_{i-1})}] * [\gamma|_{[t_{i-1},t_i]}] * [\bar{\eta}_{\gamma(t_i)}]\right)$$

Note that each factor $[\eta_{\gamma(t_{i-1})}] * [\gamma|_{[t_{i-1},t_i]}] * [\bar{\eta}_{\gamma(t_i)}]$ belongs to $\Pi_X(x_0,x_0) = \pi_1(X,x_0)$. So if $\phi\colon \pi_1(X,x_0) \to G$ is to be a homomorphism we must have $\phi([\gamma]) = \prod_{i=1}^m \phi([\eta_{\gamma(t_{i-1})}] * [\gamma|_{[t_{i-1},t_i]}] * [\bar{\eta}_{\gamma(t_i)}])$.

In fact, by the way that the intervals $[t_{i-1},t_i]$ and the paths $\eta_{\gamma(t_i)}$ were chosen, each $[\eta_{\gamma(t_{i-1})}]*[\gamma|_{[t_{i-1},t_i]}]*[\bar{\eta}_{\gamma(t_i)}]$ is the homotopy class of a loop based at x_0 which remains entirely within the open set U_{ϵ_i}, and so lies in the domain of the map ϕ_{ϵ_i}. Since ϕ was assumed to obey $\phi \circ i_{\epsilon_i *} = \phi_i$ where $i_{\epsilon_i}\colon U_{\epsilon_i} \to X$ is the inclusion we necessarily have

$$\phi([\eta_{\gamma(t_{i-1})}] * [\gamma|_{[t_{i-1},t_i]}] * [\bar{\eta}_{\gamma(t_i)}]) = \phi_{\epsilon_i}([\eta_{\gamma(t_{i-1})}] * [\gamma|_{[t_{i-1},t_i]}] * [\bar{\eta}_{\gamma(t_i)}])$$

for each i. Hence we have shown that, if $\phi\colon \pi_1(X,x_0) \to G$ is any homomorphism satisfying the required properties, its value on the homotopy class $[\gamma]$ of any loop $\gamma \in \mathscr{P}_X(x_0,x_0)$ may be determined as follows: choose a subdivision $[0,1] = \cup_{i=1}^m [t_{i-1},t_i]$ such that for each i there is $\epsilon_i \in \{0,1\}$ such that $\gamma([t_{i-1},t_i]) \subset U_{\epsilon_i}$, and then $\phi([\gamma])$ is given by

$$\phi([\gamma]) = \prod_{i=1}^m \phi_{\epsilon_i}([\eta_{\gamma(t_{i-1})}] * [\gamma|_{[t_{i-1},t_i]}] * [\bar{\eta}_{\gamma(t_i)}]) \tag{2}$$

This is more than sufficient to show that ϕ is unique if it exists. Our main remaining task is to show that (2) gives a well-defined function $\phi\colon \pi_1(X) \to G$. In this direction we will show first that the right hand side of (2) depends only on the path γ, and not on arbitrary choices that were made after we chose γ; and then, having completed this first step, we will show that $\phi([\gamma])$ in (2) depends only on the equivalence class of γ in $\pi_1(X,x_0)$ (*i.e.*, we will show that the right hand side of (2) is unchanged under homotopy rel endpoints).

To complete the first of these two steps, let us be more specific about what the arbitrary choices that were made after the choice of γ:

(a) We chose a partition of $[0,1]$ into subintervals $[t_{i-1},t_i]$ each of which is mapped either to U_0 or to U_1.
(b) Having chosen this partition into subintervals, for any i with the property that $\gamma([t_{i-1},t_i]) \subset U_0 \cap U_1$ we chose a specific $\epsilon_i \in \{0,1\}$.

First we address (b): In the case that $\gamma([t_{i-1},t_i]) \subset U_0 \cap U_1$ the term $[\eta_{\gamma(t_{i-1})}] * [\gamma|_{[t_{i-1},t_i]}] * [\bar{\eta}_{\gamma(t_i)}]$ represents an element of $\Pi_{U_0\cap U_1}(x_0,x_0) = \pi_1(U_0\cap U_1,x_0)$. Consequently since it was assumed that $\phi_0 \circ j_{0*} = \phi_1 \circ j_{1*}$ we will have

$$\phi_0([\eta_{\gamma(t_{i-1})}] * [\gamma|_{[t_{i-1},t_i]}] * [\bar{\eta}_{\gamma(t_i)}]) = \phi_1([\eta_{\gamma(t_{i-1})}] * [\gamma|_{[t_{i-1},t_i]}] * [\bar{\eta}_{\gamma(t_i)}]).$$

Thus the right hand side of (2) is independent of the choices of ϵ_i in those cases where there was more than one choice that we could have made.

As for (a), note that if $\cup_{i=1}^m [t_{i-1},t_i]$ and $\cup_{j=1}^n [s_{j-1},s_j]$ are two decompositions of $[0,1]$ with the relevant property, then by interspersing the s_j between the t_i we get a new subdivision refining each of the previous ones. So it suffices to show that the right hand side of (2) is unchanged under replacing one subdivision of $[0,1]$ by a finer subdivision (as any two subdivisions have a common refinement). Meanwhile the process of refining a subdivision $[0,1] = \cup_{i=1}^m [t_{i-1},t_i]$ consists of a finite sequence of operations involving choosing $i \in \{1,\dots,m\}$ and replacing the interval $[t_{i-1},t_i]$ by two consecutive intervals $[t_{i-1},t^*],[t^*,t_i]$ where $t_{i-1}<t^*<t_i$. This operation affects the right hand side of (2) by replacing $\phi_{\epsilon_i}([\eta_{\gamma(t_{i-1})}]*[\gamma|_{[t_{i-1},t_i]}]*[\bar{\eta}_{\gamma(t_i)}])$ with

$$\phi_{\epsilon_i}([\eta_{\gamma(t_{i-1})}]*[\gamma|_{[t_{i-1},t^*]}]*[\bar{\eta}_{\gamma(t^*)}])*\phi_{\epsilon_i}([\eta_{\gamma(t^*)}]*[\gamma|_{[t^*,t_i]}]*[\bar{\eta}_{\gamma(t_i)}]) \tag{3}$$

(Note that since we have already addressed (b) we can certainly use the same ϵ_i for both factors.) But since ϕ_{ϵ_i} is a homomorphism and since in $\Pi_X(\gamma(t_{i-1}),\gamma(t_i))$ we have an identity

$$[\gamma|_{[t_{i-1},t_i]}]=[\gamma|_{[t_{i-1},t^*]}]*[\bar{\eta}_{\gamma(t*)}]*[\eta_{\gamma(t^*)}]*[\gamma|_{[t^*,t_i]}]$$

it follows that (3) is equal to $\phi_{\epsilon_i}([\eta_{\gamma(t_{i-1})}]*[\gamma|_{[t_{i-1},t_i]}]*[\bar{\eta}_{\gamma(t_i)}])$. Thus splitting $[t_{i-1},t_i]$ into $[t_{i-1},t^*]$, $[t^*,t_i]$ does not affect the right hand side of (2).

Thus at this point (2) gives a formula for $\phi([\gamma])$ which does not depend on the particular (allowable) choices of t_i and ϵ_i, but still might depend on the particular path γ within its homotopy class. We will now show that, on the contrary, ϕ is unchanged under based homotopies of γ.

Accordingly let $\Gamma\colon [0,1]\times[0,1]\to X$ be a continuous map such that $\Gamma(s,0)=\Gamma(s,1)=x_0$ for all s, and write $\gamma_0(t)=\Gamma(0,t)$ and $\gamma_1(t)=\Gamma(1,t)$. We must show that the right-hand side of (2) defined using γ_0 is the same as the one defined using γ_1.

Just as with the reasoning that led to (2), since $[0,1]\times[0,1]$ is compact and is covered by the open sets $\Gamma^{-1}(U_0),\Gamma^{-1}(U_1)$, we may cover $[0,1]\times[0,1]$ by rectangles $[a,b]\times[c,d]$ each of which is mapped by Γ either to U_0 or to U_1. After subdividing these rectangles we may assume that they are of the form $[t_{i-1},t_i]\times[u_{j-1},u_j]$ where $0=t_0<t_1<\cdots<t_m=1$ and $0=u_0<u_1<\cdots<u_n=1$, and we choose $\epsilon_{ij}\in\{0,1\}$ with $\Gamma([t_{i-1},t_i]\times[u_{j-1},u_j])\subset U_{\epsilon_{ij}}$.

We have thus divided the square $[0,1]\times[0,1]$ into a rectangular grid. Let us encode the various ways of traveling from $(0,0)$ to $(1,1)$ northward and eastward along the edges of the grid by a tuple $\vec{v}\in\{0,1\}^{m+n}$ consisting of m zeros and n ones: starting at $(0,0)$, one successively reads through the entries of $\vec{v}$ and moves along an edge $[t_{i-1},t_i]\times\{u\}$ when the entry of $\vec{v}$ is 0 and along an edge $\{t\}\times[u_{j-1},u_j]$ when the entry of $\vec{v}$ is 1. For any such $\vec{v}$, and for $k\in\{1,\dots,m\}$, let $e_k(\vec{v})$ be the kth edge of $\vec{v}$ as just described, which will have the form $[t_{i_k(\vec{v})-1},t_{i_k(\vec{v})}]\times\{u_{j_k(\vec{v})}\}$ (if $v_k=0$ or $\{t_{i_k(\vec{v})}\}\times[u_{j_k(\vec{v})-1},u_{j_k(\vec{v})}]$ (if $v_k=1$) for appropriate integers $i_k(v),j_k(v)$. Also let $p_k(\vec{v}),q_k(\vec{v})$ be the initial and terminal points, respectively, of $e_k(\vec{v})$.

Let $\gamma_{k,\vec{v}}$ be the restriction of Γ to the edge $e_k(\vec{v})$, linearly reparametrized to have domain $[0,1]$. Thus in the notation just defined $\gamma_{k,\vec{v}}\in\mathscr{P}_X(\gamma(p_k(\vec{v})),\gamma(q_k(\vec{v})))$, and the image of $\gamma_{k,\vec{v}}$ is contained in $U_{\epsilon_{i_k(\vec{v})j_k(\vec{v})}}$.

Now define an element $\alpha(\vec{v})$ of G by

$$\alpha(\vec{v})=\prod_{k=1}^{m+n}\phi_{\epsilon_{i_k(\vec{v})j_k(\vec{v})}}\left([\eta_{p_k(\vec{v})}]*[\gamma_{k,\vec{v}}]*[\bar{\eta}_{q_k(\vec{v})}]\right)$$

Note incidentally that, just as earlier, if k and $\vec{v}$ are such that $\gamma_{k,\vec{v}}$ has image both in U_0 and in U_1, the fact that $\phi_0\circ j_{0*}=\phi_1\circ j_{1*}$ implies that the kth term in the product defining $\alpha_k(\vec{v})$ is independent of whether $\epsilon_{i_k(\vec{v})j_k(\vec{v})}$ is 0 or 1.

Observe that if $\vec{v}=(0,\dots,0,1,\dots,1)$ (m zeros followed by n ones) then the first m terms in the product defining $\alpha_k(\vec{v})$ are each the identity (since for $k\le m$ $e_k(\vec{v})$ is part of the lower edge $[0,1]\times\{0\}$ which is mapped by Γ to x_0), while in each of the last n terms of the product $\gamma_{k,\vec{v}}$ is the segment $\gamma_1|_{[u_{k-m-1},u_{k-m}]}$. Thus for this choice of $\vec{v}$, $\alpha(\vec{v})$ is equal to the γ_1-version of (2) (for the particular subdivision given by the u_j, but we showed earlier that (2) does not depend on the choice of a subdivision).

Similarly if $\vec{v} = (1,\dots,1,0,\dots,0)$ (so the path associated to $\vec{v}$ travels first along $\{0\} \times [0,1]$ and then along $[0,1] \times \{1\}$), the first n terms in the product defining $\alpha_k(\vec{v})$ reproduce the γ_0-version of (2) and the last m terms are the identity.

Thus to prove that (2) returns the same element of G regardless of the choice of loop within a homotopy class, it suffices to show that $\alpha(0,\dots,0,1,\dots,1) = \alpha(1,\dots,1,0,\dots,0)$. To see this, for any $\vec{v} \in \{0,1\}^{m+n}$ with m entries equal to 0 and n entries equal to 1, consider the effect on $\alpha_k(\vec{v})$ of swapping two adjacent entries in $\vec{v}$—for convenience suppose $v_{k-1} = 0$ and $v_k = 1$ and define $\vec{w} \in \{0,1\}^{m+n}$ by $w_{k-1} = 1$, $w_k = 0$, and $w_l = v_l$ for all other l. All of the factors in $\alpha(\vec{v})$ and $\alpha(\vec{w})$ except the $(k-1)$th and kth are then identical. Meanwhile if we combine the $(k-1)$th and kth factors (using that $\phi_{\epsilon_{i_{k-1}(\vec{v})j_{k-1}(v)}}$ is a homomorphism and that the images of all relevant segments are contained in $U_{\epsilon_{i_{k-1}(\vec{v})j_{k-1}(v)}}$) we see that the results for $\vec{v}$ and $\vec{w}$ are, respectively, images under $\phi_{\epsilon_{i_{k-1}(\vec{v})j_{k-1}(v)}}$ of loops that are based homotopic to each other within $U_{\epsilon_{i_{k-1}(\vec{v})j_{k-1}(v)}}$ (essentially the homotopy is formed by deforming a path consisting of the bottom and right edges of a rectangle to the path consisting of the left and top edges, and then applying Γ). Thus since $\phi_{\epsilon_{i_{k-1}(\vec{v})j_{k-1}(v)}}$ was assumed to be defined on π_1 (not just on the space of loops), it follows that the product of the $(k-1)$th and kth terms in $\alpha(\vec{v})$ and $\alpha(\vec{w})$ will be equal. Hence if $\vec{w}$ is formed from $\vec{v}$ by swapping an adjacent 0 and 1 then $\alpha(\vec{v}) = \alpha(\vec{w})$.

Of course, $(1,\dots,1,0,\dots,0)$ may be obtained from $(0,\dots,0,1,\dots,1)$ by repeatedly swapping adjacent 0's and 1's. So iteratively applying the previous paragraph shows that $\alpha(0,\dots,0,1,\dots,1) = \alpha(1,\dots,1,0,\dots,0)$. So by our earlier remarks we have finally shown that the γ_0 and γ_1 versions of (2) are equal whenever γ_0 and γ_1 represent the same class in $\pi_1(X,x_0)$. Thus we have a well-defined function $\phi\colon \pi_1(X,x_0) \to G$.

Given that ϕ is well-defined it is almost immediate that it satisfies the required properties. If $\gamma \in \mathscr{P}_{U_k}(x_0,x_0)$ for $k \in \{0,1\}$ then in computing $\phi([i_k \circ \gamma])$ via (2) we are free to set $\epsilon_i = k$ for each i, and then the fact that $\phi_k\colon \pi_1(U,x_0) \to G$ is a homomorphism quickly shows that $\phi([i_k \circ \gamma]) = \phi_k([\gamma])$. Thus $\phi \circ i_{k*} = \phi_k$ for $k \in \{0,1\}$. Also ϕ is easily seen to be a homomorphism: the outcome of applying ϕ in (2) to the concatenation of two loops is, after obvious adjustments for parametrization, the product in G of ϕ applied to the first with ϕ applied to the second. So since we long ago established that ϕ was the only possible map $\pi_1(X,x_0)$ that could satisfy these properties (as long as it was well-defined) this establishes the required universal property and so completes the proof.

□

Example 5.8. *For $n \geq 2$ let $X = S^n$, $U_0 = \{\vec{x} \in S^n | x_{n+1} < 1/2\}$, and $U_1 = \{\vec{x} \in S^n | x_{n+1} > -1/2\}$. As one sees from sterographic projection, U_0 and U_1 are both contractible, and in particular have $\pi_1 = 1$. So by Example 5.4 and Theorem 5.7 we recover the fact that $\pi_1(S^n,(0,\dots,1)) = 1$. Note here that the assumption that $n \geq 2$ was necessary to imply that $U_0 \cap U_1$ is path connected; in particular the same example with $n = 1$ illustrates that the path-connectedness assumption on $U_0 \cap U_1$ in Theorem 5.7 is necessary—otherwise we would incorrectly conclude that $\pi_1(S^1,0) = 1$.*

Before our next example we give a general definition. Recall first that a *pointed topological space* is simply a pair (X,x_0) where X is a topological space and $x_0 \in X$.

Definition 5.9. Let (Y,y_0) and (Z,z_0) be two pointed topological spaces. The *wedge sum* of Y and Z is the pointed topological space $(Y \vee Z, x_0)$ where

$$Y \vee Z = \frac{Y \coprod Z}{y_0 \sim z_0}$$

and x_0 is the image of y_0 (or, equivalently, of z_0) under the quotient projection $Y \coprod Z \to Y \vee Z$.

Example 5.10. *Let $(Y, y_0), (Z, z_0)$ be two copies of the pointed topological space $(S^1, 0)$, yielding a new topological space $S^1 \vee S^1$ with basepoint which we will still denote 0. You should be able to convince yourself that $S^1 \vee S^1$ is homeomorphic to the symbol '8', and also that $S^1 \vee S^1$ is homotopy equivalent to $\mathbb{C} \setminus \{0, 1\}$, so that computing the fundamental group of $S^1 \vee S^1$ will achieve our earlier goal of computing the fundamental group of $\mathbb{C} \setminus \{0, 1\}$. Naively I would like to apply van Kampen's theorem with $U_0 = Y$ and $U_1 = Z$ (identifying Y and Z with their images under the quotient map $Y \coprod Z \to Y \vee Z = S^1 \vee S^1$), but this does not quite work because Y and Z are not open in $S^1 \vee S^1$. But we can easily fix this: let $V_0 \subset Y$, $V_1 \subset Z$ be small open neighborhoods of y_0 and z_0 respectively which are each homeomorphic to open intervals, and use $U_0 = Y \vee V_1$ and $U_1 = V_0 \vee Z$. So U_0 and U_1 are each homeomorphic to the symbol 'α' and hence are homotopy equivalent to S^1 and so have fundamental group $\mathbb{Z}$. Meanwhile $U_0 \cap U_1$ is homeomorphic to the symbol 'x' and so is contractible. Thus van Kampen's theorem gives a pushout square*

$$\begin{array}{ccc} 1 & \longrightarrow & \mathbb{Z} \\ \downarrow & & \downarrow{\scriptstyle i_{0*}} \\ \mathbb{Z} & \xrightarrow{i_{1*}} & \pi_1(S^1 \vee S^1, 0) \end{array} \tag{4}$$

(Of course the upper and left maps send the unique element of the trivial group 1 to the identity 0 in $\mathbb{Z}$.)

According to Proposition 5.6, the pushout square (4) completely determines $\pi_1(S^1 \vee S^1, 0)$ up to group isomorphism, so in some sense we can say that we have computed $\pi_1(S^1 \vee S^1, 0)$. All the same it may not quite feel like we have really done so—you are probably more accustomed to defining a group by saying what its elements are and how the binary operation is defined rather than by saying that it behaves in a particular way with respect to certain commutative diagrams. I will eventually more explicitly exhibit a group that fits into a pushout square like that in (4), and hence is isomorphic to $\pi_1(S^1 \vee S^1, 0)$ by Proposition 5.6, but first let us see how one can get concrete information out of the universal property satisfied by $\pi_1(S^1 \vee S^1, 0)$ without identifying another group to which it is isomorphic.

Proposition 5.11. *There are elements $a, b \in \pi_1(S^1 \vee S^1, 0)$ which both have infinite order such that $ab \neq ba$ (and, more generally, whenever either $m, n \in \mathbb{Z}$ are both nonzero or $k, l \in \mathbb{Z}$ are both nonzero, $a^m b^n \neq b^k a^l$). Consequently $\mathbb{C} \setminus \{0, 1\}$ is not homeomorphic to $\mathbb{C} \setminus \{0\}$.*

Proof. Where $GL(2, \mathbb{R})$ is the group of invertible 2×2 real matrices, define homomorphisms $\phi_0, \phi_1 \colon \mathbb{Z} \to GL(2, \mathbb{R})$ by

$$\phi_0(m) = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \qquad \phi_1(n) = \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix}$$

(It is easy to check that these are indeed homomorphisms.) Where i_{0*}, i_{1*} are as in (4), let $a = i_{0*}(1)$ and $b = i_{1*}(1)$. So since i_{0*} and i_{1*} are homomorphisms, we have $a^m = i_{0*}(m)$ and $b^n = i_{1*}(n)$ for all $m, n \in \mathbb{Z}$.

Now the universal property of pushouts that is satisfied by $\pi_1(S^1 \vee S^1, x_0)$ yields a homomorphism $\phi \colon \pi_1(S^1 \vee S^1, 0) \to GL(2, \mathbb{R})$ such that $\phi \circ i_{0*} = \phi_0$ and $\phi \circ i_{1*} = \phi_1$. (Since the group in the upper left of the pushout square is in this case trivial the commutativity condition is satisfied vacuously.)

We then have, for $m, n \in \mathbb{Z}$,

$$\phi(a^m) = \phi_0(m) = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \qquad \phi(b^n) = \phi_1(n) = \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix}$$

So since ϕ is a homomorphism,

$$\phi(a^m b^n) = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix} = \begin{pmatrix} 1+mn & m \\ n & 1 \end{pmatrix}$$

while

$$\phi(b^k a^l) = \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}\begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & l \\ k & 1+kl \end{pmatrix}$$

So if $a^m b^n = b^k a^l$ we would have $\phi(a^m b^n) = \phi(b^k a^l)$ and hence in particular $mn = 0$ and $kl = 0$ (by considering upper left and lower right entries). Thus indeed $a^m b^n \neq b^k a^l$ whenever either m and n are both nonzero or k and l are both nonzero.

It remains to prove the last sentence. We have shown that $\pi_1(S^1 \vee S^1, 0)$ is nonabelian, so it is certainly not isomorphic to the cyclic group $\pi_1(S^1, 0)$. Thus $S^1 \vee S^1$ and S^1 are not based homotopy equivalent. But $\mathbb{C} \setminus \{0, 1\}$ is based homotopy equivalent to $S^1 \vee S^1$ while $\mathbb{C} \setminus \{0\}$ is based homotopy equivalent to S^1, so if $\mathbb{C} \setminus \{0, 1\}$ were homeomorphic to $\mathbb{C} \setminus \{0\}$ we would obtain a based homotopy equivalence between $S^1 \vee S^1$ and S^1.

□

5.1. **Free groups and free products.** We now develop some of the theory of the sorts of groups that arise from the pushout squares appearing in van Kampen's theorem.

Definition 5.12. Let $\{G_\alpha\}_{\alpha\in A}$ be any collection of groups. A *free product* of $\{G_\alpha\}_{\alpha\in A}$ is a group $*_\alpha G_\alpha$ together with homomorphisms $i_\beta \colon G_\beta \to *_\alpha G_\alpha$ for each $\beta \in A$, obeying the following universal property: for any other group G together with homomorphisms $\phi_\beta \colon G_\beta \to G$ for all β, there is a unique homomorphism $\phi \colon *_\alpha G_\alpha \to G$ such that $\phi \circ i_\beta = \phi_\beta$ for all β. In the case that $\{G_\alpha\}_{\alpha\in A}$ is a finite collection $\{G_1, \ldots, G_m\}$ we will often write $G_1 * \cdots * G_m$ instead of $*_\alpha G_\alpha$.

An easy modification of the proof of Theorem 5.6 shows that $*_\alpha G_\alpha$ is unique up to isomorphism (if this isn't clear to you then you should think about the proof of Theorem 5.6 until it is. Note though that we haven't yet shown that $*_\alpha G_\alpha$ exists.). Similarly to Example 5.10 one sees from van Kampen's theorem that, under suitable hypotheses on Y and Z, the wedge sum $(Y \vee Z, x_0)$ obeys

$$\pi_1(Y \wedge Z, x_0) \cong \pi_1(Y, y_0) * \pi_1(Z, z_0)$$

(for these "suitable hypotheses" one could use the assumption that there are open neighborhoods U of y_0 in Y and V of z_0 in Z such that the inclusions $\{y_0\} \to U$ and $\{z_0\} \to V$ are based homotopy equivalences—just as in Example 5.10 the need for such an assumption arises from the fact that van Kampen's theorem does not directly apply to the cover $Y \vee Z = Y \cup Z$ since Y and Z are not open in $Y \vee Z$). Thus for instance $S^1 \vee S^1$ has fundamental group $\mathbb{Z} * \mathbb{Z}$, and $S^1 \vee \mathbb{R}P^2$ has fundamental group $\mathbb{Z} * \mathbb{Z}_2$. By induction we likewise have $\pi_1((S^1)^{\vee n}, x_0) \cong \mathbb{Z}^{*n}$ for all $n \in \mathbb{N}$ (here of course $(S^1)^{\vee n}$ means the wedge sum of n copies of S^1—perhaps most easily visualized as a "flower with n petals"—and $\mathbb{Z}^{*n}$ means the free product of n copies of $\mathbb{Z}$. The version of van Kampen's theorem in [H] allows one to generalize this to infinite wedge sums and free products.

Definition 5.13. If S is a set, a *free group on S* is a group $F(S)$ together with a function $\iota \colon S \to F(S)$ obeying the following universal property: if G is any group and $f \colon S \to G$ is any function then there is a unique homomorphism $h \colon F(S) \to G$ such that $h \circ \iota = f$.

Again a straightforward modification of the proof of Theorem 5.6 shows that $F(S)$ is uniquely characterized up to isomorphism by this universal property.

Proposition 5.14. *If S is any set, let $\{G_\alpha | \alpha \in S\}$ be a family of groups parametrized by S, each of which is isomorphic to $\mathbb{Z}$. Then $*_{\alpha\in S} G_\alpha$ is a free group on S.*

Proof. For each $\beta \in S$ and let e_β denote a generator of G_β (which we have assumed to be isomorphic to $\mathbb{Z}$). As part of the definition of $*_\alpha G_\alpha$, we have homomorphisms $i_\beta \colon G_\beta \to *_\alpha G_\alpha$; write $1_\beta = i_\beta(e_\beta)$.

Define $\iota \colon S \to *_\alpha G_\alpha$ by $\iota(\beta) = 1_\beta$

If $f : S \to G$ is any function where G is another group, for each β there is a unique homomorphism $\phi_\beta \colon G_\beta \to G$ obeying $\phi_\beta(e_\beta) = f(\beta)$. (Namely, each element of G_β is given by e_β^n for some $n \in \mathbb{Z}$, and so the unique such ϕ is defined by $\phi_\beta(e_\beta^n) = f(\beta)^n$.) The universal property obeyed by $*_\alpha G_\alpha$ then gives a unique homomorphism $\phi \colon *_\alpha G_\alpha \to G$ obeying $\phi \circ i_\beta = \phi_\beta$ for all β. So for $\beta \in S$ we have

$$(\phi \circ \iota)(\beta) = \phi(1_\beta) = \phi(i_\beta(e_\beta)) = \phi_\beta(e_\beta) = f(\beta)$$

and so $\phi \circ \iota = f$. This proves the existence part of the universal property for $F(S)$; for the uniqueness part just observe that any homomorphism $\phi' \colon *_\alpha G_\alpha \to G$ obeying $\phi' \circ \iota = f$ would need to obey $\phi'(i_\beta(e_\beta)) = \phi_\beta(e_\beta)$, and so since e_β generates G_β would obey $\phi' \circ i_\beta = \phi_\beta$ for each β, so by the uniqueness part of the universal property for $*_\alpha G_\alpha$ would be equal to ϕ. □

In particular it follows that $\pi_1((S^1)^{\vee n}, 0)$ is a free group $F(\{1, \ldots, n\})$ on $\{1, \ldots, n\}$. Note that we have not yet given any *group-theoretic* argument for why $F(\{1, \ldots, n\})$ should even exist, but what we have done consitutes a topological proof of the existence of $F(\{1, \ldots, n\})$: there is indeed a well-defined group $\pi_1((S^1)^{\vee n}, 0)$, and it satisfies the universal property characterizing any free group $F(\{1, \ldots, n\})$.

You should be able to convince yourself that $\mathbb{C} \setminus \{1, \ldots, n\}$ is based homotopy equivalent to $(S^1)^{\vee n}$, so the following shows that all of the spaces $\mathbb{C} \setminus \{1, \ldots, n\}$ are mutually nonhomeomorphic as n varies through $\mathbb{N}$.

Exercise 5.15. If G is a group let $[G, G]$ denote the subgroup of G generated by all elements of form $aba^{-1}b^{-1}$ where $a, b \in G$.

(a) Prove that $[G, G]$ is a normal subgroup of G, and that the quotient $\frac{G}{[G,G]}$ is an abelian group which, together with the quotient map $\pi : G \to \frac{G}{[G,G]}$, satisfies the following universal property: If H is any abelian group and if $\phi_0 : G \to H$ is a homomorphism then there is a unique homomorphism $\phi : \frac{G}{[G,G]} \to H$ such that $\phi \circ \pi = \phi_0$. ($\frac{G}{[G,G]}$ is called the *abelianization* of G)

(b) Prove that if $m \neq n$, the free groups $F(\{1, \ldots, m\})$ and $F(\{1, \ldots, n\})$ are not isomorphic. (Perhaps the easiest way to do this is by considering their abelianizations.)

References

[H] A. Hatcher. *Algebraic Topology*. Cambridge U. Press, 2002. `http://www.math.cornell.edu/~hatcher/AT/ATpage.html`