The Circle Has a Hole In It
An Invitation to (Algebraic) Topology

Brendan Murphy

Abstract

Topology is the study of shape independent of geometry. Abstracting away from the concrete realization of a space as a subset of $\mathbb{R}^n$ gives us the tools to argue at a higher level, without getting into too many awkward details. But once we've forgotten about distance, angle, and orientation, what intrinsic properties are left? In this paper we develop machinery from algebraic topology in order to show that the circle and line are topologically distinct, and in doing so formalize the notion of what it means for a space, like the circle, to have a hole.

Contents

1 Introduction

In an introductory analysis course, a student learns to think about space in a more abstract way than they may have before. It dawns upon the student that the real line and the open interval $(0, 1)$ are in many ways similar, especially in a "local" sense. When learning vector calculus, this student learns the basics of differential forms, and discovers that on an annulus closed differential forms may fail to be exact. Later, when this student studies complex analysis, they find a similar phenomenon. Harmonic functions have conjugates on some domains, like disks, but not always on more complicated domains, like annuli. The contour integral of a function along two curves is the same when one can be continuously deformed into the other.

Each of these examples has at its core algebraic topology. The Poincare Lemma (not proven here) says that closed forms are the same as exact forms on a space S if the identity map on S is "homotopic" to a constant function. Global harmonic conjugates on a domain D always exist if and only if all paths in D are path-homotpic, and two contours give the same integral if they are homotopic.

This paper introduces concepts from algebraic topology and category theory, focusing on homotopy. Our motivating problem is showing that the circle and the interval are different on a very fundamental level, in that they are topological spaces which are not "homeomorphic". The first five sections of this paper set the stage for the meat of the paper in sections 6 and 7. A reader with prior exposure to topology and category theory should most likely skip ahead to these sections and look back at lemmas as they are used. In section 6, we develop the idea of the fundamental groupoid, which detects the number of holes in a space, and in section 7 we prove the Seifert van-Kampen Theorem (Theorem 7.3 and Theorem 7.7) that suffices to compute the fundamental groupoid in a variety of situations. In particular, it allows us to show that the circle and the interval have different fundamental groupoids, and from this we conclude in Corollary 7.9 that the circle and the interval are topologically distinct.

2 Topology

The fundamental object of topology is that of a space.

Definition 2.1. A **space** is a set X such that for any $x \in X$, there is a family of subsets of X called "neighborhoods" of x, such that

N1. If A is a neighborhood of x, then $x \in A$.

N2. The whole space X is a neighborhood of x.

N3. If A is a neighborhood of x and $A \subseteq B$, then B is a neighborhood of x.

N4. If A and B are neighborhoods of x, then so is $A \cap B$.

N5. For any neighborhood A of X, there is a neighborhood B of x contained in A such that B is a neighborhood of each of its points.

We refer to this collection of neighborhoods as a "topology" on X. We can make the familiar real line $\mathbb{R}$ into a space by saying that a set N is a neighborhood of some point x when it contains an interval (a, b) which also contains x. Axioms N1 and N2 are immediately satisfied, as is N3, and since the intersection of two open intervals is once again an open interval we get N4 immediately. Finally N5 is satisfied because an interval (a, b) is automatically a neighborhood of any of its points. In analysis, a set is open if it contains a ball around each of its points, which we can now restate as that set being a neighborhood of each of its points.

Definition 2.2. A subset U of a space X is **open** if it is a neighborhood of each of its points.

This definition allows us to state axiom N5 more succinctly: any neighborhood of x must contain an open set containing x. Some basic results about open sets are as follows.

Lemma 2.1. Suppose X is a space. Then the following properties holds:

(a) The whole space X and the empty set are open.

(b) If U and V are open sets of X, so is $U \cap V$.

(c) If Λ is some possibly infinite set and $\{U_\lambda\}_{\lambda \in \Lambda}$ is a family of open sets of X, then so is the union $\bigcup_{\lambda \in \Lambda} U_\lambda$.

Proof. For (a), note that X is open since it is a neighborhood of all its points by N2, and the empty set is vacuously a neighborhood of all its points. To show (b), suppose you have some $x \in A \cap B$. Then $x \in A$ and $x \in B$, so A and B are neighborhoods of x. But then by N4, so is $A \cap B$. Finally for (c), suppose $x \in \bigcup_{\lambda \in \Lambda} U_\lambda$. Then for some fixed λ_0, we must have $x \in U_{\lambda_0}$. Since U_{λ_0} is open, this shows it is a neighborhood of x. But then $\bigcup_{\lambda \in \Lambda} U_\lambda$ contains a neighborhood of x, so by N3 it is a neighborhood of x. $\square$

Definition 2.3. The **interior** of a subset S of a space X, written $\operatorname{Int} S$, is the set of all points of which S is a neighborhood.

Immediately we see that $\operatorname{Int} S \subseteq S$, since any point of which S is a neighborhood is contained in S. Also taking the interior is monotone, i.e. if $A \subseteq B \subseteq X$ then $\operatorname{Int} A \subseteq \operatorname{Int} B$ (this is immediate from N3). We have the following characterization of the interior:

Lemma 2.2. The interior of some set S is the largest open set contained in S. Formally,

$$\operatorname{Int} S = \bigcup_{U \subseteq S, U \text{ open}} U.$$

Proof. We first show $\operatorname{Int} S$ contains all open sets contained in S. Suppose U is an open subset of S. Then for any $x \in U$, U is a neighborhood of x, so by axiom N3 so is S, which means x is in the interior of S. We now show $\operatorname{Int} S$ is open. If x is in the interior of S, then S is a neighborhood of x, so by axiom N5 there is an open set U containing x contained in S. But as we just noted, $U \subseteq \operatorname{Int} S$, so by axiom N3 $\operatorname{Int} S$ is also a neighborhood of x. Since $\operatorname{Int} S$ is a neighborhood of all of its points, it is open. $\square$

As a corollary, we see $\operatorname{Int} U = U$ for any open U, and in particular $\operatorname{Int}\operatorname{Int} S = \operatorname{Int} S$ for any S. Dual to the open sets are the closed sets. The way we think of closed sets in analysis is in terms of limits and sequences, but this are insufficient without a way to measure distance. We do have an equivalent characterization of closedness which generalizes better to arbitrary topological spaces, given below.

Definition 2.4. A subset C of a space X is **closed** if $X \setminus C$ is open.

We also have a dual to Lemma 2.1. Each of the following properties can be proven from Lemma 2.1 by applying DeMorgan's Laws.

Lemma 2.3. Suppose X is a space. Then the following properties holds:

(a) The empty set and the whole space X are closed.

(b) If C and K are closed sets of X, so is $C \cup K$.

(c) If Λ is some possibly infinite set and $\{C_\lambda\}_{\lambda \in \Lambda}$ is a family of closed sets of X, then so is the intersection $\bigcap_{\lambda \in \Lambda} U_\lambda$.

Following this pattern, there is a dual to the interior of a subset. The interior is formed by taking away all points which are not "interior points", i.e. points which the set is a neighborhood of. The closure is found by adding in all "closure points". Inuitively, a closure point of S is a point which is arbitrarily close to the points in S, in that no neighborhood, no matter how small, will seperate S from the point.

Definition 2.5. The **closure** of a subset S of a space X, written $\overline{S}$, is the set of all points which have no neighborhoods disjoint from S.

More positively, we have that for any $x \in \overline{S}$ and any neighborhood N of x, there is some point $y \in S \cap N$. The precise way that this is dual to the interior is given below

Lemma 2.4. For any subset S of a space X,

$$X \setminus \overline{S} = \operatorname{Int}(X \setminus S).$$

Proof. A point x is in $X \setminus \overline{S}$ if and only if there is some neighborhood N of x which is disjoint from S. But N being disjoint from S is equivalent to being contained in $X \setminus S$, and there being a neighborhood of x contained in $X \setminus S$ is the same as saying $X \setminus S$ is itself a neighborhood of x, or equivalently $x \in \operatorname{Int}(X \setminus S)$. □

This gives us immediately that the closure of any set is closed (as the name suggests), because its complement is open. And by writing $\overline{A} = X \setminus \operatorname{Int}(X \setminus A)$ and using the monotonicity of the interior, we get that $\overline{A} \subseteq \overline{B}$ if $A \subseteq B$. Also, $S \subseteq \overline{S}$ since

$$X \setminus \overline{S} = \operatorname{Int}(X \setminus S) \subseteq X \setminus S.$$

Lemma 2.5. The closure of set S is the smallest closed set containing in S. Formally,

$$\overline{S} = \bigcap_{C \supseteq S, C \text{ closed}} C.$$

Proof. ad We've already shown that the closure of S is a closed set containing S, so it suffices to show it is contained within any other closed set containing S. Suppose C is closed and $S \subseteq C$. Then $X \setminus C$ is open and $X \setminus C \subseteq X \setminus S$, so by Lemma 2.2, $X \setminus C \subseteq \operatorname{Int}(X \setminus S) = X \setminus \overline{S}$, which means $\overline{S} \subseteq C$ as desired. □

We immediately get from this that $\overline{C} = C$ if C is closed, and $\overline{\overline{S}} = \overline{S}$ always. These five notions (neighborhood, open, closed, interior, closure) are almost all of the primitives from which the rest of topology is built. There is still one important concept left, though. As we will see in section 4, a mathematical theory comes not just with objects but also maps.

Definition 2.6. If X and Y are spaces, and $f : X \to Y$ is a function between them as sets, then f is **continuous** if for any open set U of Y, the preimage $f^{-1}(U)$ is open in X.

This is the same as saying preimages of closed sets are closed, since preimages commute with the complement. We also have an equivalent definition in terms of neighborhoods.

Lemma 2.6. A function $f : X \to Y$ is continuous if and only if for any point $x \in X$ and neighborhood N of $f(x)$ in Y, the preimage $f^{-1}(N)$ is a neighborhood of x in X.

Proof. First suppose f is continuous. Suppose we have some $x \in X$ and a neighborhood N of $f(y)$ in Y. Then there is an open subneighborhood U of $f(x)$, and by continuity $f^{-1}(U)$ is also open. But also $x \in f^{-1}(U)$ since U is a neighborhood of $f(x)$, and so $f(x) \in U$. Since U is an open set containing x, it is a neighborhood of x, and since $f^{-1}(U) \subseteq f^{-1}(N)$ so is $f^{-1}(N)$.

Now suppose that for any point $x \in X$ and neighborhood N of $f(x)$ in Y, the preimage $f^{-1}(N)$ is a neighborhood of x in X, and let U be an open set in Y. Let $f^{-1}(U)$ and suppose $x \in V$. Then $f(x) \in U$, so U is a neighborhood of $f(x)$, and thus $f^{-1}(U)$ is a neighborhood of x by assumption. Thus $f^{-1}(U)$ is open, and so f is continuous. □

The ability to speak about continuity is essentially why we care about topology. As we see below, the basic kinds of functions that can defined for arbitrary sets and combinations of continuous functions are always continuous, which will be important in section 4.

Lemma 2.7. The identity function $id_X : X \to X$, defined by $id_X(x) = x$, is continuous (for any space X).

Proof. This is immediate from the fact that $id_X^{-1}(U) = U$. □

Lemma 2.8. Let X, Y, Z be spaces. Suppose $f : X \to Y$ and $g : Y \to Z$ are continuous functions. Then $g \circ f : X \to Z$ is continuous.

Proof. Note that $(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U))$. So if U is open in Z, then $g^{-1}(U)$ is open in Y the continuity of g, and thus by the continuity of f, $f^{-1}(g^{-1}(U))$ is open in X. Thus $g \circ f$ is continuous. □

Lemma 2.9. Let X, Y be spaces and suppose $b \in Y$. Then the constant function $f(x) = b$ is continuous.

Proof. Suppose U is an open set of Y. Then if $b \in U$, $f^{-1}(U) = X$, and otherwise $f^{-1}(U) = \emptyset$. But both of these sets are open, f is continuous. □

Now we have some basic topological definitions, but how do we build up new spaces? The topological structure of the real line should show us how to turn subsets of that line like the interval $[0, 1]$ into a space. Similarly, we should be able to get the structure of the Cartesian plane $\mathbb{R} \times \mathbb{R}$ from that of $\mathbb{R}$.

Definition 2.7. Suppose X is a space and $S \subseteq X$ is a subset. Then S can be made into a space using the **subspace topology** as follows. The neighborhoods of a point $x \in S$ are of the form $S \cap N$ where N is a neighborhood of x in X.

Lemma 2.10. The subspace topology is a well defined topology.

Proof. If N is a neighborhood of x in the subspace topology, then by definition $N = N' \cap S$ for some N' a neighborhood of x in X. Then N' contains x by N1, so $x \in N'$, and $x \in S$ by assumption, so $x \in N' \cap S = N$. This verifies N1. Also $S = X \cap S$ is a neighborhood of x since X is a neighborhood of x in the whole space, showing N2. Now suppose $A \subseteq B \subseteq S$ and A is a neighborhood of x in S. Then $A = A' \cap S$ for a neighborhood A' of x in X, and

$$B = B \cup A = B \cup (A' \cap S) = (B \cup A') \cap (B \cup S) = (B \cup A') \cap S.$$

And since $B \cup A'$ contains the neighborhood A' of x in X, it is also a neighborhood of x in X. Thus B is a neighborhood of x in S, verifying N3. Axiom N4 holds since $(N \cap S) \cap (M \cap S) = (N \cap M) \cap S$. Finally for N5, suppose $N = N' \cap S$ is a neighborhood of x in S for N' a neighborhood of x in X. Then by N5, there is an open set $U \subseteq N'$ of X containing x. Then $x \in U \cap S \subseteq N' \cap S$ and $U \cap S$ is open in S. To see this, take some $y \in U \cap S$ and note that U is a neighborhood of y in X, and thus $U \cap S$ is the intersection of a neighborhood of y in X with S, and thus by definition a neighborhood of y in S. □

The last argument in this proof proves that any intersection of an open set and S is open in S. In fact, this is usually how the subspace topology is defined. The converse of this also holds.

Lemma 2.11. Let U be open in S. Then there is a set U' open in X such that $U = U' \cap S$.

Proof. By definition of the subspace topology, for each $x \in U$ we have a neighborhood N_x of x in X such that $U = N_x \cap S$. Also, by N5 we have an open set $U_x \subseteq N_x$ containing x, so $U_x \cap S \subseteq N_x \cap S = U$. Thus

$$\left(\bigcup_{x\in U} U_x\right) \cap S = \bigcup_{x\in U} U_x \cap S \subseteq \bigcup_{x\in U} U = U.$$

But also since $x \in U_x$,

$$U = \bigcup_{x\in U} \{x\} \subseteq \bigcup_{x\in U} U_x.$$

And since also $U \subseteq S$, we have

$$U \subseteq \left(\bigcup_{x\in U} U_x\right) \cap S.$$

Which shows

$$U = \left(\bigcup_{x\in U} U_x\right) \cap S.$$

Finally by Lemma 2.1, $\bigcup_{x\in U} U_x$ is open in X since it is a union of open sets. □

This also gives us immediately the corresponding result for closed sets, since if C is closed in S, then $S \setminus C = U \cap S$ is open. Then $C = S \setminus (U \cap S) = S \setminus U = S \cap (X \setminus U)$ is the intersection of a closed set of X and S. Now note that the inclusion map $\iota_S : S \to X$ given by $\iota_S(x) = x$ is continuous, since $\iota^{-1}(U) = U \cap S$ for any open set U, and we know that this is open in S. Also the restriction of any continuous function $f : X \to Y$ to a subspace S, written $f\restriction_S$, is continuous. This comes from the fact that the composition of continuous functions is continuous and $f\restriction_S = f \circ \iota_S$. Dually, if $f : X \to Y$ and $\operatorname{im} f \subseteq B \subseteq Y$, then the restriction of the codomain of f to B is continuous, since if U is open in B, we must have $U = U' \cap B$ for some U' open in Y, and

$$f^{-1}(U) = f^{-1}(U') \cap f^{-1}(B) = f^{-1}(U') \cap X = f^{-1}(U').$$

And this is open by the continuity of f. Now that we can talk about subspaces, we are able to state the most important lemma of this section, which tells us how we can build up continuous functions from ones defined on smaller pieces of our space.

Lemma 2.12 (The Gluing Lemma). Suppose A, B are subspaces of X such that $A \cup B = X$ and we have continuous functions $f : A \to Y$ and $g : B \to Y$ which are equal on $A \cap B$. Then if $A \setminus B \subseteq \operatorname{Int} A$ and $B \setminus A \subseteq \operatorname{Int} B$, then

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

defines a continuous function on all of X.

Proof. Since f and g agree on the overlap of A and B, this is a well defined function. We use the definition of continuity given in Lemma 2.6. Suppose $x \in X$ and N is a neighborhood of $f(x)$ in Y. Then $h^{-1}(N) \cap A = f^{-1}(N)$ is a neighborhood of x in A by continuity of f, so there is some neighborhood L of x in X such that

$$h^{-1}(N) \cap A = f^{-1}(N) = L \cap A.$$

Similarly by continuity of g there is a neighborhood M of x in X such that

$$h^{-1}(N) \cap B = g^{-1}(N) = M \cap B.$$

We proceed by cases on whether $x \in A \cap B$ or $x \in A \setminus B$ or $x \in B \setminus A$. In this first case, since $A \cup B = X$,

$$L \cap M \subseteq (L \cap A) \cup (M \cap B) = (h^{-1}(N) \cap A) \cup (h^{-1}(N) \cap B) = h^{-1}(N).$$

But $L \cap M$ is the intersection of two neighborhoods of x, so it is a neighborhood of x, which means $h^{-1}(N)$ contains a neighborhood of x, and thus is itself a neighborhood of x.

In the case, if $x \in A \setminus B \subseteq \operatorname{Int} A$, A is a neighborhood of x, thus $A \cap L$ is, and so $h^{-1}(N)$ contains a neighborhood of x, and is thus a neighborhood of x. The final case follows by the same argument. $\square$

In particular, if A and B are closed sets then $A \setminus B = (A \cup B) \setminus B = X \setminus B$ is an open set contained in A, so it is contained in the interior of A (and the same but with A and B swapped), which means the conditions of the Gluing Lemma are automatically satisfied. We will often write "this is continuous by the Gluing Lemma" in to prove that a function defined by cases is continuous, and trust the reader to check that the sets satisfying those inequalities are closed.

We can also build new, larger spaces out of old ones.

Definition 2.8. If X and Y are spaces, there is a topological structure on their product $X \times Y$, where a set P is a neighborhood of a point (x, y) iff there exists neighborhoods N of x and M of y such that $N \times M \subseteq P$.

The reader should attempt to verify that this is a well defined topology. This definition immediately implies that the product of open sets is open, since if U and V are open then for any $(x, y) \in U \times V$, we have $N \times M \subseteq U \times V$ where $N = U$ is a neighborhood of x and $M = V$ is a neighborhood of y. The product of closed sets is closed since $(X \times Y) \setminus (C \times D) = (X \times (Y \setminus D)) \cup ((X \setminus C) \times Y)$. The product topology also comes with two important maps, $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$, called the projection maps, which are given by $\pi_X(x, y) = x$ and $\pi_Y(x, y) = y$. These are continuous since if U is open in X, then $\pi_X^{-1}(U) = U \times Y$ is the product of two open sets (and similarly for π_Y). Maps into and out of products behave nicely.

Lemma 2.13. Suppose $f : Z \to X$ and $g : Z \to Y$. Then the map $(f, g) : Z \to X \times Y$ given by $(f, g)(z) = (f(z), g(z))$ is continuous.

Proof. We use the definition of continuity from Lemma 2.6. Suppose $z \in Z$ and we have a neighborhood L of $(f(z), g(z))$ in $X \times Y$. Then there are neighborhoods N of $f(z)$ in X and M of $g(z)$ in Y such that $N \times M \subseteq L$, and thus $f^{-1}(N \times M) \subseteq f^{-1}(L)$. But also

$$f^{-1}(M \times N) = \{t \in Z : (f(t), g(t)) \in N \times M\} = f^{-1}(N) \cap g^{-1}(M)$$

And each of these sets is a neighborhood of z by continuity, so their intersection is as well. Thus $f^{-1}(L)$ contains a neighborhood of z, and so it is one too. $\square$

Although it may not be immediately apparent, this allows us to conclude that practically any function which feeds its inputs through a series of continuous functions is continuous. For example, if $f : X \to Z$ and $g : Y \to W$, then the function $f \times g : X \times Y \to Z \times W$ given by $(f \times g)(x, y) = (f(x), g(y))$ is continuous, since $f \times g = (f \circ \pi_X, g \circ \pi_Y)$. We will not bother to argue such functions are continuous in the future, since it is essentially just a matter of inserting combinations of projections, pairing, and composition.

We close this section by asking the question of when two spaces are the same. Consider the spaces $A = \mathbb{R}$ with the topology described earlier on and $B = \{(x, y) \in \mathbb{R}^2 : y = x^2\}$ with the subspace topology inherited from the product topology on $\mathbb{R}^2$. These two spaces are completely different from a geometrical standpoint, one being a line and the other a parabola, but they are "the same" topologically.

Definition 2.9. A **homeomorphism** between spaces X and Y is a continuous function $f : X \to Y$ such that there is some continuous map $g : Y \to X$ such that $g \circ f = id_X$ and $f \circ g = id_Y$, i.e. f is a continuous function with a continuous inverse. If such a homeomorphism exists, we say X and Y are **homeomorphic**, or **topologically equivalent**.

Lemma 2.14. The relation that two spaces are homeomorphic is an equivalence relation.

Proof. Let X, Y, Z be spaces. First note that X is homeomorphic to itself by the identity map. Then if $f : X \to Y$ is a homeomorphism, so is $f^{-1} : Y \to X$, so Y is homeomorphic to X. Finally if $g : Y \to Z$ is another homeomorphism, then $g \circ f$ is continuous with inverse $f^{-1} \circ g^{-1}$, and this is continuous since it is a composition of continuous functions. □

Under this definition, A and B are topologically equivalent, which is witnessed by the map $f : A \to B$ given by $f(x) = (x, x^2)$. This map is continuous (the proof that $x \mapsto x^2$ is continuous can be found in any analysis textbook), and it has inverse $\pi(x, y) = x$. Spacially, we're taking the real line and bending it up into a parabola. We can now, as the abstract of this paper claims "ignore distance".

Lemma 2.15. Let a, b, c, d be positive real numbers such that $a < b$ and $c < d$. Then, considered as subspaces of $\mathbb{R}$, the spaces $[a, b]$ and $[c, d]$ are homeomorphic.

Proof. By Lemma 2.14, it suffices to show the interval $[0, 1]$ is homeomorphic to any interval $[a, b]$. First define a homeomorphism $f : [0, 1] \to [0, b - a]$ by $f(t) = (b - a)t$, which is continuous with continuous inverse $f^{-1}(t) = \frac{t}{b-a}$ by standard analysis. Then define $g : [0, b-a] \to [a, b]$ by $g(t) = t + a$. This is continuous with continuous inverse $g^{-1}(t) = t - a$, so it is a homeomorphism. Thus $[0, 1]$ is homeomorphic to $[a, b]$. □

In the next section, we study an important topological property of certain spaces.

3 Compactness

Compactness is not an intuitive property. It allows us to turn infinite problems into finite ones, but the kinds of spaces which are compact are more varied than one would imagine. Our main use for compactness is in proving the Lebesgue Covering Lemma (Lemma 3.13). This lemma lets us break up intervals and rectangles into chunks small enough that certain desireable properties hold. Unfortunately, showing that the sets we care about are compact takes a fair amount of work.

Definition 3.1. Let K be a space. A **cover** of K is a family $\mathcal{C}$ of subsets of K such that $K = \bigcup_{S \in \mathcal{C}} S$. An **open cover** of K is a cover consistening of open sets. We say K is **compact** if for any open cover $\mathcal{U}$ of K there is a finite subfamily $\mathcal{F} \subseteq \mathcal{U}$ which is also a cover.

Lemma 3.1. Let X, Y be spaces and suppose we have a continuous surjection $f : X \to Y$. If X is compact, then Y is as well. As a corollary, if X is compact and Y is homeomorphic to X, then Y is compact.

Proof. Let $\mathcal{U}$ be an open cover of Y. Define $\mathcal{U} = \{f^{-1}(U) : U \in \mathcal{U}\}$. Then since f is continuous, this is an open cover of X, and thus by compactness there is a finite subcover $\{f^{-1}(U_1), f^{-1}(U_2), \ldots, f^{-1}(U_n)\}$. Then

$$Y = f(X) = f\left(\bigcup_{k=1}^{n} f^{-1}(U_k)\right) = \bigcup_{k=1}^{n} f(f^{-1}(U_k)) = \bigcup_{k=1}^{n} U_k.$$

Thus any open cover of Y has a finite subcover, showing compactness. □

We often wish to speak about compact subsets of an ambient space, and working with these subsets as subspaces can be clumsy. Luckily there is a nicer way of handling things.

Definition 3.2. Let K be a subset of the space X. Define a **cover of the subset** K to be a family $\mathcal{C}$ of subsets of X such that $K \subseteq \bigcup_{S \in \mathcal{C}} S$. We say K is a **compact subset of** X if every cover of the subset K has an open subcover.

Lemma 3.2. Let X be a space and K a subset. Then K is compact as a subspace if and only if it is compact as a subset.

Proof. Open sets of K are of the form $U \cap K$ for an open set U of X, so open covers of K are of the form $\{U_\lambda \cap K\}_{\lambda \in \Lambda}$ for some set of indices Λ. Then

$$K = \bigcup_{\lambda \in \Lambda} (U_\lambda \cap K) = \left(\bigcup_{\lambda \in \Lambda} U_\lambda \right) \cap K$$

So $K \subseteq \bigcup_{\lambda \in \Lambda} U_\lambda$, and thus $\{U_\lambda\}_{\lambda \in \Lambda}$ is an open cover of the subset K. Similarly, an open cover of the subset K can be turned into an open cover of the subspace by intersecting each member of the cover with K. Thus open covers of the subspace and subset are "the same", and the rest of the proof is just symbol pushing. □

Lemma 3.3. Let X be compact and let C be a closed subset of X. Then C is also compact.

Proof. Let $\mathcal{U}$ be an open cover of the subset C. Then $\mathcal{U} \cup \{X \setminus C\}$ is an open cover of X, since C is closed and thus $X \setminus C$ is open. Thus by compactness of X, there is a finite subcover $\mathcal{F}$ of $\mathcal{U} \cup \{X \setminus C\}$. Removing $X \setminus C$ from this subcover (if it is there) will still give us an open cover of the subset C, and will be a subcover of $\mathcal{U}$. Thus C is compact. □

The product of two compact spaces is also compact. In fact, the product of any number of compact spaces is compact. This celebrated result is known as Tychonoff's Theorem, and is equivalent to the axiom of choice. We do not prove it here, but it is the capstone of the kind of set-theoretical topology we do in this and the previous section.

Lemma 3.4. Let X and Y be spaces. Suppose K is a compact subset of Y and U an open subset of $X \times Y$. Then the set $V = \{x \in X : \{x\} \times K \subseteq U\}$ is open.

Proof. Let x be any point of V. Then since U is open, for any point $y \in K$, $(x, y) \in U$, so there are open neighborhoods D_y, E_y of x, y such that $D_y \times E_y \subseteq U$. Then $\{E_y\}_{y \in K}$ is an open cover of K, so by compactness there is a finite set of points $\{y_1, y_2, \dots, y_n\}$ such that $K \subseteq E_{y_1} \cup E_{y_2} \cup \dots \cup E_{y_n}$. Define $N = D_{y_1} \cap D_{y_2} \cap \dots D_{y_n}$. Since each D_{y_j} is open, so is N, and

$$N \times K = N \times \left(\bigcup_{k=1}^{n} E_{y_k} \right) = \bigcup_{k=1}^{n} N \times E_{y_k} \subseteq \bigcup_{k=1}^{n} D_{y_k} \times E_{y_k} \subseteq U.$$

Thus $N \subseteq V$. Since V contains a neighborhood of each of its points, and it is open. $\square$

Lemma 3.5. If X and Y are compact spaces, so is their product $X \times Y$.

Proof. Let $\mathcal{U}$ be an open cover of $X \times Y$. Then let x be any point of X. The subset $\{x\} \times Y$ of $X \times Y$ is compact, since it is easily seen to be homeomorphic to Y. Then since $\mathcal{U}$ is a cover of $X \times Y$, it is also a cover of $\{x\} \times Y$, so by compactness there are a finite number of sets $\{U_1, U_2, \dots, U_n\} \subseteq \mathcal{U}$ which cover $\{x\} \times Y$. Let $U = U_1 \cup U_2 \cup \dots \cup U_n$ and define $V_x = \{x' \in X : \{x\} \times Y \subseteq U\}$. By the previous lemma, V_x is open. Also, $V_x \times Y$ is covered by finitely many sets all belonging to $\mathcal{U}$. Then since $x \in V_x$ for all $x \in X$, we have an open cover $\{V_x : x \in X\}$ of X, and thus by compactness finitely many x_j such that $\bigcup_j V_{x_j}$ covers X. Thus $X \times Y$ is covered by the (finite) family $V_{x_j} \times Y$, and each of these sets is covered by finitely many elements of $\mathcal{U}$. Composing these coverings gives us a finite cover of $X \times Y$ by sets of $\mathcal{U}$. $\square$

It's not at all clear why we call this property compactness, but it turns out to correspond to the more intuitive notion of "closed and bounded" in standard Euclidean space. Before that, we define some extremely important spaces

Definition 3.3. The **circle** is the subspace of $\mathbb{R}^2$

$$\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.$$

Definition 3.4. The **interval** is the subspace of $\mathbb{R}$

$$\mathbb{I} = \{x \in \mathbb{R} : 0 \leq x \leq 1\}.$$

We show that the interval is compact, and conclude by Lemma 3.1 that the circle is as well, since we have a continuous surjection $\theta \mapsto (\cos(2\pi\theta), \sin(2\pi\theta))$ of $\mathbb{I}$ onto $\mathbb{S}^1$. It turns out that knowing $\mathbb{I}$ is compact is also strong enough to prove the theorem mentioned above about compact subsets of $\mathbb{R}^n$. To show $\mathbb{I}$ is compact, we first show it is "connected".

Definition 3.5. Let X be some space. We say X is **connected** iff there is no continuous surjective function $f : X \to \{0, 1\}$.

Note if X is connected and Y is homeomorphic to X, say with a homeomorphism $h : X \to Y$, then Y must be connected, since a surjective continuous function $f : Y \to \{0, 1\}$ would give rise to a continuous function $f \circ h : X \to \{0, 1\}$, which is surjective since both f and h are. This definition of connectedness is a somewhat intutive, since a continuous function onto a discrete space can only take on two different values if there are two different "parts" of X. However the following more abstract definition is also useful.

Lemma 3.6. A disconnection of a space X is a pair (A, B) of nonempty disjoint open subsets whose union is X. A space X is connected iff there does not exist a disconnection of it.

Proof. It suffices to show that there exists a continuous surjection $f : X \to \{0, 1\}$ iff there exists a disconnection of X. Given a surjection f, we can write $X = f^{-1}(\{0, 1\}) = f^{-1}(0) \cup f^{-1}(1)$. These sets are obviously disjoint, are open by continuity of f, and are nonempty because f is surjective.

Now suppose (A, B) is a disconnection of X. Then define

$$f(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \in B \end{cases}$$

This is well defined since A and B are disjoint, is continuous by the Gluing Lemma (since $A \setminus B \subseteq A = \operatorname{Int} A$ and the same for B), and is surjective because A and B are nonempty. Thus the result holds. □

Lemma 3.7. $\mathbb{I}$ is connected.

Proof. We show every continuous function $f : \mathbb{I} \to \{0, 1\}$ is constant. Let x, y be points of $\mathbb{I}$ and suppose $x < y$. Let $s = \sup\{z \in \mathbb{I} : x \le z \le y \text{ and } f(z) = f(x)\}$. Then $s = y$ or $s < y$. Suppose for contradiction that $s < y$. Then by the continuity of f, there is some $\delta > 0$ such that for all

z within δ of s, we have $|f(z) - f(s)| < 0.5$. But since $f(z)$ and $f(s)$ are integers, this implies that $f(z) = f(s)$ for z within δ of s. If $s < y$, then $\delta' = \frac{1}{2}\min(\delta, y - s)$ is such that $0 < \delta' < \delta$, so $f(s + \delta') = f(s) = f(x)$, and also $x \leq s < s + \delta' \leq y$. But this contradicts the fact that s is an upper bound for $\{z \in \mathbb{I} : x \leq z \leq y \text{ and } f(z) = f(x)\}$. Thus $\mathbb{I}$ is connected. □

Connectedness is a hugely important property, but we won't spend much time focusing on it (although arguably the material we will look at on homotopy is just a finer grained notion of connectedness).

Lemma 3.8. $\mathbb{I}$ is compact.

Proof. Let $\mathcal{A}$ be an open cover of the subset $\mathbb{I}$. Suppose we had some open cover $\mathcal{B}$ such that for each $U \in \mathcal{B}$, there is some V in $\mathcal{A}$ such that $U \subseteq V$. It suffices to find a finite subcover $\{U_1, \ldots, U_n\}$ of $\mathcal{B}$, since each element U_j of that cover must be contained in some element V_j of $\mathcal{A}$. Then $\mathbb{I} \subseteq \bigcup_{j=1}^{n} U_j \subseteq \bigcup_{j=1}^{n} V_j$, so we have a finite subcover $\{V_1, \ldots, V_n\}$ of $\mathcal{A}$. We define $\mathcal{B}$ to be the collection of open intervals I such that $I \subseteq U$ for some $U \in \mathcal{A}$. This is a cover of $\mathbb{I}$ since each element $x \in \mathbb{I}$ must be contained in some open $U \in \mathcal{A}$, and thus U is a neighborhood of x, which implies there is some open interval $I \subseteq U$ containing x.

Now define $f : \mathbb{I} \to \{0, 1\}$ by $f(x) = 1$ if $[0, x]$ can be covered by finitely many intervals in $\mathcal{B}$ and 0 otherwise. We show f is continuous Let $U = f^{-1}(1)$; we show U is open. Let x be some point of $\mathbb{I}$ such that $f(x) = 1$, so $[0, x] \subseteq I_1 \cup I_2 \cup \ldots I_n$ for $I_1, I_2, \ldots, I_n \in \mathcal{B}$. Since $\mathcal{B}$ covers $\mathbb{I}$, we must have $x \in I$ for some $I \in \mathcal{B}$. Then $I \subseteq U$ since for any $y \in I$, it is that

$$[0, y] \subseteq [0, x] \cup ([0, y] \setminus [0, x]) \subseteq [0, x] \cup I \subseteq I_1 \cup I_2 \cup \ldots \cup I_n \cup I$$

and thus $f(y) = 1$. Thus $f^{-1}(1)$ is open, and similarly $f^{-1}(0)$ is open, since if $f(x) = 0$, the above argument shows $f(y) = 0$ for all y in some neighborhood of x. Thus f is continuous, so by the connectedness of $\mathbb{I}$, f must be constant. But also $f(0) = 1$, since $[0, 0]$ can be covered by any interval in $\mathcal{B}$ containing 0. Thus $f(x) = 1$ for all $x \in \mathbb{I}$, and in particular $f(1) = 1$. Thus $\mathbb{I} = [0, 1]$ can be covered by a finite number of intervals in $\mathcal{B}$. □

Theorem 3.9 (Heine-Borel)**.** A subset K of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof. Suppose K is not closed. Then there is some point $x \in \mathbb{R}^n$ such that $x \in \overline{K} \setminus K$, so by basic analysis for all $\varepsilon > 0$ there is some $y \in K$ such that $|x - y| < \varepsilon$. Then the closed balls $B_\varepsilon = \{y \in \mathbb{R}^n : |x - y| \leq \varepsilon\}$ centered at x always contain some point in K, and so no finite subset of the family of complements of these balls $\{\mathbb{R}^n \setminus B_\varepsilon : \varepsilon > 0\}$ covers K. However since $\bigcup_{\varepsilon>0} \mathbb{R}^n \setminus B_\varepsilon = \mathbb{R}^n \setminus \bigcap_{\varepsilon>0} B_\varepsilon = \mathbb{R}^n \setminus \{x\}$ and $x \notin K$, the complements of these balls form an open cover of K with no finite subcover. Thus K is not compact.

Now suppose K is not bounded. Then we can cover $\mathbb{R}^n$, by open balls $B_r = \{y \in \mathbb{R}^n : |y| < r\}$, and thus cover K as well, but if K was contained in some finite subcover it would be bounded. Thus K is not compact.

Finally suppose K is both closed and bounded. Then for some $B > 0$ we have $K \subseteq [-B, B]^n$. By Lemma 2.15, $[-B, B]$ is homeomorphic to $\mathbb{I}$, so by Lemmas 3.1 and 3.8 it is compact. Then by Lemma 3.5 $[-B, B]^n$ is also compact. Since $K \subseteq [-B, B]^n$, also $K = K \cap [-B, B]^n$. Thus K is the intersection of a closed set of $\mathbb{R}^n$ (i.e. K) with the subspace $[-B, B]^n$, and so it is closed in $[-B, B]^n$. Then Lemma 3.3 implies K is compact. □

Closedness and boundedness are vastly easier to show than compactness, so this lemma is mostly useful in showing closed and bounded sets are compact. The Heine-Borel Theorem has some useful corollaries, like the Extreme Value Theorem from calculus.

Corollary 3.10. Suppose X is a compact space and $f : X \to \mathbb{R}$ is a continuous function. Then f attains a maximum value (and by symmetry a minimum value).

Proof. By Lemma 3.1, $f(A)$ is closed and bounded. Since it is bounded, it has a supremum, and since it is closed it contains that supremum. Thus the result holds. □

We close off this section with some last words on compactness, introducing an important lemma which we will use later. The reader should be aware that the following lemma holds in the more general context of compact metric spaces, but we will not pad the length of this paper even further by covering that.

Definition 3.6. Let S be a nonempty subset of $\mathbb{R}^n$. Define the **distance from a point x in $\mathbb{R}^n$ to S** to be $d_S(x) = \inf\{|x - y| : y \in S\}$

Lemma 3.11. The function $d_S : \mathbb{R}^n \to \mathbb{R}$ defined above is continuous.

Proof. By the triangle inequality, we have for any $a, b \in \mathbb{R}^n$ and $c \in S$ that $|a-c| \le |a-b| + |b-c|$. Then since taking infimums preserves weak inequalities,

$$d_S(a) = \inf_{c \in S} |a-c| \le \inf_{c\in S}(|a-b|+|b-c|) = |a-b| + \inf_{c \in S}(|b-c|) = |a-b| + d_S(b).$$

And so $d_S(a) - d_S(b) \le |a-b|$. Then by symmetry we also have $d_S(b) - d_S(a) \le |a-b|$, so $|d_S(b) - d_S(a)| \le |a-b|$. Standard analysis implies from here that d_S is continuous. □

Lemma 3.12. Suppose K is a compact subset of $\mathbb{R}^n$. Then $d_K(x) = 0$ implies $x \in K$.

Proof. By Corollary 3.10, if $\inf_{y \in K} |x-y| = 0$, then $|x - y_0| = 0$ for some $y_0 \in K$, since $f(y) = |x-y|$ is continuous. But then $x = y_0$, so $x \in K$. □

Lemma 3.13 (The Lebesgue Covering Lemma)**.** Let $\mathcal{A}$ be an open cover of a compact subset K of $\mathbb{R}^n$. There is some value $\delta > 0$ such that for any ball B of radius smaller than δ contained in K, there is some $U \in \mathcal{A}$ such that $B \subseteq U$.

Proof. By compactness, we can assume without loss of generality that $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ is finite. Then let $C_j = K \setminus A_j$. Since each A_j is open, the C_j are closed, and thus by Lemma 3.3 they are compact. Thus by Lemma 3.11 the function $f : K \to \mathbb{R}$ defined by

$$f(x) = \frac{1}{n} \sum_{j=1}^{n} d_{C_j}(x)$$

is continuous. By Corollary 3.10, this function attains some minimum value $f(x_0) = \delta$. If $\delta = 0$, then $d_{C_j}(x_0) = 0$ for all j, then by Lemma 3.12 $x \in C_j$ for all j, so $x \notin A_j$ for any j. But then $\mathcal{A}$ is not a cover, since $x \in K \setminus (A_1 \cup A_2 \cup \ldots \cup A_n)$. Thus $\delta > 0$. Then for any ball B centered at b of radius δ, we must have $f(b) \ge \delta$, so for some j we must have $d_{C_j}(b) \ge \delta$. But then for any c such that $|b-c| < \delta$, we can't have $c \in C_j$, since if we did we would have $\delta \le d_{C_j}(b) \le |b-c| < \delta$. Thus if $c \in B$, $c \notin C_j$, so $c \in A_j$. Thus $B \subseteq A_j$. □

4 Category Theory

Whereas topology seeks to abstract space, category theory seeks to abstract abstraction. The language of categories can be used to perform general constructions interpretable in a multitude of contexts, each specialization carrying different, context sensitive information. Unfortunately, this means it is hard to appreciate category theory without a wealth of examples to draw upon. With that said, we begin the section not by defining a category, but instead a group, so that we have one more point of comparison.

Definition 4.1. A **group** is a set G along with a function $m : G \times G \to G$ satisfying the conditions below. By tradition and for ease of reading we write m as multiplication, i.e. the expression xy is understood to mean $m(x, y)$.

G1. m is assosciative. That is $x(yz) = (xy)z$ for all $x, y, z \in G$.

G2. There is some identity element e_G such that $e_G x = x = x e_G$ for all $x \in G$.

G3. For all x, there is an inverse element y such that $xy = e_G = yx$.

We usually just write e for e_G. The two groups we care in this paper are "the" trivial group, which is what we call any group with a single element, and the group of integers. Any one element set $\{*\}$ is a group in a unique wayb by defining $m(*, *) = *$. The integers are a group with identity element 0 and "multiplication" given by $m(a, b) = a + b$, and where the inverse of a is $-a$. The reader will note we did not include the axiom of commutativity in our definiton of a group, i.e. that $xy = yx$. This is because many important groups have a noncommutative multiplication, like the group of invertible n by n matrices. A group with commutative multiplication is called **abelian**. The group axioms imply a few useful properties immediately.

Lemma 4.1. In a group G, the identity element is unique.

Proof. Suppose we have an element e' such that $xe' = x = e'x$ for all $x \in G$. Then taking $x = e$, we find $ee' = e$, but also by G2 $ee' = e'$. □

Lemma 4.2. For any $x, y, z \in G$, if $yx = e$ and $xz = e$, then $y = z$. As a corollary, inverses are unique, and so for all x we have a well defined inverse element x^{-1}.

Proof. $y = ye = y(xz) = (yx)z = ez = z$. □

When we work with groups, we don't care about what their elements look like, but instead the way they interact. That is to say, if the elements in two groups interact with one another in the same way, we want to think of them as "the same" group. As an example, consider the groups $A = \{-1, 1\}$ under multiplication and $B = \{\texttt{true}, \texttt{false}\}$ under boolean xor. Both of these consist of an identity element x and a nonidentity element y such that $m(y, y) = x$, i.e. y is its own inverse, and so to a group theorist they are the same. We make this precise by with the concept of an isomorphism.

Definition 4.2. An **isomorphism of groups** between groups G, H is an invertible function $\varphi : G \to H$ such that $\varphi(xy) = \varphi(x)\varphi(y)$ for all $x, y \in G$ and $\varphi^{-1}(xy) = \varphi^{-1}(x)\varphi^{-1}(y)$ for all $x, y \in H$.

The isomorphism from A to B as above is $\varphi(1) = \texttt{false}$, $\varphi(-1) = \texttt{true}$. Also, our condition that $\varphi(xy) = \varphi(x)\varphi(y)$ isn't just useful in saying when two groups are the same, but really tells us that φ respects the group structure. In the same sense that continuous maps are the "nice" functions of space, maps satisfying this property are the "nice" functions of groups.

Definition 4.3. A **homomorphism** between groups G, H is a function $\varphi : G \to H$ such that $\varphi(xy) = \varphi(x)\varphi(y)$ for all $x, y \in G$.

Note that we must have $\varphi(e_G)\varphi(e_G) = \varphi(e_G e_G) = \varphi(e_G)$, and multiplying both sides by $\varphi(e_G)^{-1}$ we see $\varphi(e_G) = e_H$. This along with Lemma 4.2 implies that $\varphi(x^{-1}) = \varphi(x)^{-1}$, since $\varphi(x^{-1})\varphi(x) = \varphi(xx^{-1}) = \varphi(e_G) = e_H$.

Lemma 4.3. Suppose G, H, K are groups and we have homomorphisms $\varphi : G \to H$, $\psi : H \to K$. Then $\psi \circ \varphi$ is a homomorphism.

Proof. $\psi(\varphi(xy)) = \psi(\varphi(x)\varphi(y)) = \psi(\varphi(x))\psi(\varphi(y))$ □

Lemma 4.4. Suppose G is a group. Then the identity $id_G(x) = x$ is a homomorphism.

Proof. $id_G(xy) = xy = id_G(x)id_G(y)$ □

Thus an isomorphism is a homomorphism with an inverse which is also a homomorphism. This sounds very similar to our definition of an homeomorphism in that it is a "nice" map between some kind of mathematical structures whose inverse is also "nice". Just like with topological spaces, we also have a kind of "product" of groups.

Definition 4.4. Let G and H be groups. Then the cartesian product $G \times H$ has a group structure, and we call this group the **direct product** of G and H. The identity element is (e_G, e_H), and multiplication componentwise, i.e. $(g, h)(g', h') = (gg', hh')$, and inverses are $(g, h)^{-1} = (g^{-1}, h^{-1})$.

This concludes the rapidfire introduction to groups. This paper is far more focused on *groupoids* (see the next section), and the upcoming material on categories is far more important to understand. We now state three completely unrelated lemmas about products of sets, spaces, and groups.

Lemma 4.5. Suppose X and Y are sets. Then there are functions $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ and for any other set Z with functions $f : Z \to X$, $g : Z \to Y$ there is a unique function $h : Z \to X \times Y$ such that $\pi_X \circ h = f$ and $\pi_Y \circ h = g$.

Proof. π_X and π_Y are the projection maps we've seen before. Then if we have such a Z, define $h(z) = (f(z), g(z))$ Then $\pi_X(h(z)) = \pi_X(f(z), g(z)) = f(z)$ and similarly $\pi_Y \circ h = g$. Now suppose there was some $h' : Z \to X \times Y$ such that $\pi_X \circ h' = f$ and $\pi_Y \circ h' = g$. Then for any $z \in Z$,

$$h'(z) = (\pi_X(h'(z)), \pi_Y(h'(z))) = (f(z), g(z)) = h(z).$$

So $h' = h$, and thus h is unique. □

Lemma 4.6. Suppose X and Y are spaces. Then there are continuous functions $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ and for any other space Z with continuous functions $f : X \to Z$, $g : Y \to Z$ there is a unique continuous function $h : X \times Y \to Z$ such that $\pi_X \circ h = f$ and $\pi_Y \circ h = g$.

Proof. The h defined in the previous lemma is continuous by Lemma 2.13, and uniqueness is as above. □

Lemma 4.7. Suppose X and Y are groups. Then there are homomorphisms $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ and for any other space Z with homomorphisms $f : X \to Z$, $g : Y \to Z$ there is a unique homomorphisms $h : X \times Y \to Z$ such that $\pi_X \circ h = f$ and $\pi_Y \circ h = g$.

Proof. We use the same h as in the last two lemmas. Then if f, g are homomorphisms, so is h, since

$$h(zw) = (f(zw), g(zw)) = (f(z)f(w), g(z)g(w)) = (f(z), g(z))(f(w), g(w)) = h(z)h(w).$$

□

We can state these three lemmas by the following diagram

$$\begin{tikzcd} & & X \\ Z \arrow[r, dashed, "h"] \arrow[urr, bend left, "f"] \arrow[drr, bend right, "g"'] & X \times Y \arrow[ur, "\pi_X"] \arrow[dr, "\pi_Y"] & \\ & & Y \end{tikzcd}$$

We say such a diagram is commutative if all paths between objects in it have the same composition. Then the past three lemmas say that there's always a unique map h making this diagram commute, if X, Y, Z are sets/spaces/group and f, g, π_X, π_Y, h are "nice" maps. And this is in fact the general definition of product that we use in a categorical setting. A category has just enough structure for us to be able to draw diagrams, and talk about whether they commute.

Definition 4.5. A **category** C has a collection consists of a collection $\mathrm{Obj}(\mathsf{C})$, the **objects** of C, and for each pair of objects $X, Y \in \mathrm{Obj}(\mathsf{C})$ a collection $\mathrm{Hom}_{\mathsf{C}}(X, Y)$ of **maps**† between the objects X and Y, called the hom set of X and Y. We write $f : X \to Y$ to mean $f \in \mathrm{Hom}_{\mathsf{C}}(X, Y)$. There is also a "composition" operator $\circ : \mathrm{Hom}_{\mathsf{C}}(Y, Z) \times \mathrm{Hom}_{\mathsf{C}}(X, Y) \to \mathrm{Hom}_{\mathsf{C}}(X, Z)$ for any three objects X, Y, Z and for each object X an identity $id_X \in \mathrm{Hom}_{\mathsf{C}}(X, X)$ for this composition. A category must satisfy following.

C1. Composition is assosciative. That is, if we have objects X, Y, Z, W in C and maps $f : X \to Y$, $g : Y \to Z$, $h : Z \to W$, we must have

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

C2. The identity map is an identity for $\circ$. That is, for all objects X, Y and maps $f : X \to Y$, $g : Y \to X$, we must have $f \circ id_X = f$ and $id_Y \circ g = g$.

This gives us just enough structure to state things in terms of commutative diagrams. There are three important examples of categories that you should think of whenever seeing a new concept.

† Also referred to as **morphisms** or **arrows**.

Definition 4.6. There is a category Set whose objects are sets where the maps in $\mathrm{Hom}_{\mathsf{Set}}(X, Y)$ are all functions $X \to Y$. Composition is regular function composition and the identity map the identity function $id(x) = x$.

Definition 4.7. There is a category Top whose objects are spaces where the maps in $\mathrm{Hom}_{\mathsf{Top}}(X, Y)$ are the *continuous* functions $X \to Y$. Composition is function composition and the identity map the identity function $id(x) = x$. This is well defined by Lemmas 2.7 and 2.8.

Definition 4.8. There is a category Grp whose objects are groups where the maps in $\mathrm{Hom}_{\mathsf{Grp}}(X, Y)$ are the *homomorphisms* $X \to Y$. Composition is function composition and the identity map the identity function $id(x) = x$. This is well defined by Lemmas 4.4 and 4.3.

To reiterate, categories give us the ability to draw diagrams, and diagrams let us define concepts. The main diagrammatical-concept we deal with is a "pushout", but before we tackle this we finish up our definition of products.

Definition 4.9. Suppose we have some category C and objects X, Y of C. If we have an object P of C and maps $f : P \to X$, $g : P \to Y$, we call (P, f, g) a **candidate product** of X and Y. If for any other product candidate (Z, α, β) of X and Y there is a unique map $h : Z \to P$ such that

$$\begin{tikzcd}
 & & X \\
Z \arrow[r, dashed, "h"] \arrow[urr, bend left, "\alpha"] \arrow[drr, bend right, "\beta"'] & P \arrow[ur, "g"] \arrow[dr, "f"] & \\
 & & Y
\end{tikzcd}$$

commutes, then we say (P, f, g) is the product of X and Y (often we drop the maps f, g and just say P is the product). We call this existence of the unique map the **universal property** of the product.

With this new language, we can restate lemmas 4.5, 4.6, and 4.7 as "Set, Top, and Grp have products". But the definition we've given is a little unclear. We've said that *any* object satisfying our universal property is *the* product. Definite articles generally shouldn't be attached to multiple things. However, it turns out that all products are related in a strong sense.

Lemma 4.8. Let C be a category with objects X, Y. If (P, f, g) and (P', f', g') are both the product of X and Y, there are maps $\varphi : P \to P'$ and $\psi : P' \to P$ such that $\varphi \circ \psi = id_{P'}$ and $\psi \circ \varphi = id_P$. Further, the following diagram commutes.

$$\begin{tikzcd}
& X & \\
P \arrow[ur, "f"] \arrow[rr, shift left, "\varphi"] \arrow[dr, "g"'] & & P' \arrow[ul, "f'"'] \arrow[ll, shift left, "\psi"] \arrow[dl, "g'"] \\
& Y &
\end{tikzcd}$$

Proof. We automatically obtain maps φ, ψ making the above commute by the universal property of the product. But then also

$$f \circ (\psi \circ \varphi) = (f \circ \psi) \circ \varphi = f' \circ \varphi = f.$$

And similarly $g \circ (\psi \circ \varphi) = g$. Thus we have the commutative diagram

$$\begin{tikzcd}
& & X \\
P \arrow[urr, bend left, "f"] \arrow[r, dashed, "h"] \arrow[drr, bend right, "g"'] & P \arrow[ur, "f"] \arrow[dr, "g"] & \\
& & Y
\end{tikzcd}$$

Where $h = \psi \circ \varphi$. But this also commutes when $h = id_P$, so by the uniqueness part of the universal property, $\psi \circ \varphi = id_P$. The same argument shows $\varphi \circ \psi = id_{P'}$, so the result holds. $\square$

What does this mean if C is Set? It says any two products of the same two sets are in bijection. In particular, $X \times Y$ and $Y \times X$ can both be shown to satisfy the universal property for the product of X and Y, so this gives a bijection between them. How about when C is Top? Any two products are homeomorphic. And if C is Grp, this says any two products are isomorphic. We reuse this last term and say an invertible map $\varphi \in \mathrm{Hom}_{\mathsf{C}}(X, Y)$ is an isomorphism between objects X and Y for any category C. In category theory, we essentially only care about things up to isomorphism. Any categorical properties that hold for an object X also hold for any Y isomorphic to X, like for example the property of being a product of two objects (do you see how to make the maps?). We close the section by introducing the pushout square, a central concept of this paper, and proving a useful lemma about it.

Definition 4.10. Let X, Y, Z be objects of a category C, and suppose we have maps $f : X \to Y$ and $g : X \to Z$. Call an object W with maps $f' : Z \to W$, $g' : Y \to W$ a **pushout candidate** of (f, g) if the following square commutes

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "g"'] & Y \arrow[d, "g'"] \\
Z \arrow[r, "f'"'] & W
\end{tikzcd}$$

If W is such that for any other pushout candidate W' with maps $\tilde{f} : Z \to W$, $\tilde{g} : Y \to W$ there is a unique map $h : W \to W'$ making the following commute,

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "g"'] & Y \arrow[d, "g'"] \arrow[ddr, bend left, "\tilde{g}"] & \\
Z \arrow[r, "f'"'] \arrow[drr, bend right, "\tilde{f}"'] & W \arrow[dr, dashed, "h"] & \\
 & & W'
\end{tikzcd}$$

then we say W is the **pushout** of (f, g). In this case we call the square given above a **pushout square**.

An anologue of Lemma 4.8 holds for pushouts, but the proof is essentially the same so we leave it as an exercise to the reader. Arbitrary pushouts exist in all three categories we've looked at so far, but their construction in Top and Grp requires a little more knowledge than we have now. In Set, we can construct a pushout of (f, g) by first embedding Y and Z in the disjoint union $Y \sqcup Z$, then quotienting this disjoint union out by the equivalence relation $f(x) \sim g(x)$.[†] Intuitively a pushout consists of the original two objects Y and Z put together with the images of f and g "glued". We prove a quick lemma about pushouts of groups, which we'll use in Section 9 to conclude some groups of interest are actually trivial.

[†]If we generalize the disjoint union to a **coproduct**, whose universal property is like that of product but with all arrows reversed, and the quotient operation to the **coequalizer** of the maps f, g, and coproducts/coequalizers exist for all pairs of objects/maps in a given category, then that category has pushouts which are the coequalizer of a coproduct.

Lemma 4.9. Suppose we have a pushout of groups

$$\begin{tikzcd}
H \arrow[r, "f"] \arrow[d, "g"'] & G \arrow[d, "g'"] \\
K \arrow[r, "f'"'] & W
\end{tikzcd}$$

If g' and f' are trivial, in the sense that $g'(x) = e_W$ and $f'(y) = e_W$ for $x \in G$, $y \in K$, then W is a trivial group. If G and K are trivial groups, this condition is automatically satisfied, so W is a trivial group with no assumption on f', g'.

Proof. By the universal property of the pushout, we have a unique map $h : W \to W$ making

$$\begin{tikzcd}
H \arrow[r, "f"] \arrow[d, "g"'] & G \arrow[d, "g'"] \arrow[ddr, bend left, "g'"] & \\
K \arrow[r, "f'"'] \arrow[drr, bend right, "f'"'] & W \arrow[dr, dashed, "h"] & \\
 & & W
\end{tikzcd}$$

commute. Then $id_W \circ g' = g'$ since $id_W(g'(x)) = id_W(e_W) = e_W$ for any $x \in G$, and by the same argument $id_W \circ f' = f'$. Thus $h = id_W$. Also if if $k : W \to W$ is the homomorphism $k(x) = e_W$, we have $k \circ g' = g'$ since $k(g'(x)) = k(e_W) = e_W$, and similarly $k \circ f' = f'$. Thus $k = h = id_W$, so $x = id_W(x) = k(x) = e_W$ for any $x \in W$. Thus K is trivial. $\square$

In the rest of this section, we prove some techinical lemmas about pushouts.

Definition 4.11. Suppose we have commutative squares S and S'

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "g"'] & Y \arrow[d, "g'"] \\
Z \arrow[r, "f'"'] & W
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
X' \arrow[r, "h"] \arrow[d, "k"'] & Y' \arrow[d, "k'"] \\
Z' \arrow[r, "h'"'] & W'
\end{tikzcd}$$

A **map of squares** $S \to S'$ is a family of maps $\alpha : X \to X'$, $\beta : Y \to Y'$,

$\gamma : Z \to Z'$, and $\delta : W \to W'$ such that the following cube commutes.

$$\begin{tikzcd}
X \arrow[rr, "f"] \arrow[dr, "\alpha"'] \arrow[dd, "g"'] & & Y \arrow[dr, "\beta"] \arrow[dd, "g'"'] & \\
& X' \arrow[rr, "h"] \arrow[dd, "k"'] & & Y' \arrow[dd, "k'"] \\
Z \arrow[rr, "f'"] \arrow[dr, "\gamma"'] & & W \arrow[dr, "\delta"'] & \\
& Z' \arrow[rr, "h'"] & & W'
\end{tikzcd}$$

It's worth noting that given a category C, the collection of commutative squares drawn in that category forms a new cateogry $\mathsf{C}_\square$ with the morphisms maps of square. This paper doesn't cover this further, however.

Lemma 4.10. Suppose in a category C we have squares as in Definition 4.11. Also suppose we have maps of squares $(\alpha, \beta, \gamma, \delta) : S \to S'$ and $(a, b, c, d) : S' \to S$ such that $a \circ \alpha = id_X$, $b \circ \beta = id_Y$, $c \circ \gamma = id_Z$, and $d \circ \delta = id_W$. Then if S' is a pushout square, so is S.

Proof. Suppose we have a pushout candidate $(\tilde{W}, \tilde{f}, \tilde{g})$, i.e. a square

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "g"'] & Y \arrow[d, "\tilde{g}"] \\
Z \arrow[r, "\tilde{f}"'] & \tilde{W}
\end{tikzcd}$$

Then we have maps $\tilde{f} \circ c : Z' \to \tilde{W}$ and $\tilde{g} \circ b : Y' \to \tilde{W}$. Then

$$(\tilde{f} \circ c) \circ k = \tilde{f} \circ c \circ k = \tilde{f} \circ g \circ a = \tilde{g} \circ f \circ a = \tilde{g} \circ b \circ h = (\tilde{g} \circ b) \circ h.$$

So we have the commutative diagram

$$\begin{tikzcd}
X' \arrow[r, "h"] \arrow[d, "k"'] & Y' \arrow[d, "\tilde{g} \circ b"] \\
Z' \arrow[r, "\tilde{f} \circ c"'] & \tilde{W}
\end{tikzcd}$$

And by the universal property of the pushout, this induces a unique map $q : W' \to \tilde{W}$ such that

$$\begin{tikzcd}
X' \arrow[r, "h"] \arrow[d, "k"'] & Y' \arrow[d, "k'"] \arrow[ddr, bend left, "\tilde{g} \circ b"] & \\
Z' \arrow[r, "h'"'] \arrow[drr, bend right, "\tilde{f} \circ c"'] & W' \arrow[dr, dashed, "q"] & \\
 & & \tilde{W}
\end{tikzcd}$$

commutes. Then if $p = q \circ \delta : W \to \tilde{W}$,

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "g"'] & Y \arrow[d, "g'"] \arrow[ddr, bend left, "\tilde{g}"] & \\
Z \arrow[r, "f'"'] \arrow[drr, bend right, "\tilde{f}"'] & W \arrow[dr, dashed, "p"] & \\
 & & \tilde{W}
\end{tikzcd}$$

also commutes. This is because

$$p \circ f' = q \circ \delta \circ f' = q \circ h' \circ \gamma = \tilde{f} \circ c \circ \gamma = \tilde{f}.$$

Similarly,

$$p \circ g' = q \circ \delta \circ g' = q \circ k' \circ \beta = \tilde{g} \circ b \circ \beta = \tilde{g}.$$

This verifies existence. To show uniqueness, suppose there was some $p' : W \to W'$ making the above commute. Then let $q' = p' \circ d : W' \to \tilde{W}$. Then

$$q' \circ k' = p' \circ d \circ k' = p' \circ g' \circ b = \tilde{g} \circ b$$

and

$$q' \circ h' = p' \circ d \circ h' = p' \circ f' \circ c = \tilde{f} \circ c.$$

Which means we have the commutative diagram

$$\begin{tikzcd}
X' \arrow[r, "h"] \arrow[d, "k"'] & Y' \arrow[d, "k'"] \arrow[ddr, bend left, "\tilde{g} \circ b"] & \\
Z' \arrow[r, "h'"'] \arrow[drr, bend right, "\tilde{f} \circ c"'] & W' \arrow[dr, dashed, "q'"] & \\
 & & \tilde{W}
\end{tikzcd}$$

And thus by uniqueness of q, this means $q = q' = p' \circ d$. Thus

$$p = q \circ \delta = (p' \circ d) \circ \delta = p'$$

showing uniqueness of p. □

This diagram chase is rather technical (and requires a cube to state!), and the reader should feel no sadness if they don't understand the proof. We close this section by showing that pushouts can be glued together (this might suggest a notion of composition to the reader, but that way lies madness and double categories, an idea too complicated for this paper to broach).

Lemma 4.11. Suppose in a category C we have the pushout squares

$$\begin{tikzcd} X \arrow[r, "\alpha"] \arrow[d, "\beta"'] & Y \arrow[d, "\delta"] \\ W \arrow[r, "\overline{\alpha}"'] & P \end{tikzcd} \qquad \begin{tikzcd} Y \arrow[r, "\gamma"] \arrow[d, "\delta"'] & Z \arrow[d, "\overline{\delta}"] \\ P \arrow[r, "\overline{\gamma}"'] & Q \end{tikzcd}$$

Then the following is a pushout square

$$\begin{tikzcd} X \arrow[r, "\gamma\circ\alpha"] \arrow[d, "\beta"'] & Z \arrow[d, "\overline{\delta}"] \\ W \arrow[r, "\overline{\gamma}\circ\overline{\alpha}"'] & Q \end{tikzcd}$$

Proof. It's immediate that this is a pushout candidate by the following diagram

$$\begin{tikzcd} X \arrow[r, "\alpha"] \arrow[d, "\beta"'] & Y \arrow[r, "\gamma"] \arrow[d, "\delta"] & Z \arrow[d, "\overline{\delta}"] \\ W \arrow[r, "\overline{\alpha}"'] & P \arrow[r, "\overline{\gamma}"'] & Q \end{tikzcd}$$

Now suppose we have another candidate, i.e. we have the diagram

$$\begin{tikzcd} X \arrow[r, "\alpha"] \arrow[d, "\beta"'] & Y \arrow[r, "\gamma"] \arrow[d, "\delta"] & Z \arrow[d, "\overline{\delta}"] \arrow[ddr, bend left, "f"] & \\ W \arrow[r, "\overline{\alpha}"'] \arrow[drrr, bend right, "g"'] & P \arrow[r, "\overline{\gamma}"'] & Q & \\ & & & R \end{tikzcd}$$

Then $f \circ \gamma : Y \to R$ and $(f \circ \gamma) \circ \alpha = g \circ \beta$, so we have a unique map $h : P \to R$ such that

$$\begin{tikzcd}
X \arrow[r, "\alpha"] \arrow[d, "\beta"'] & Y \arrow[r, "\gamma"] \arrow[d, "\delta"] & Z \arrow[d, "\overline{\delta}"] \arrow[ddr, bend left, "f"] & \\
W \arrow[r, "\overline{\alpha}"'] \arrow[drrr, bend right, "g"'] & P \arrow[r, "\overline{\gamma}"'] \arrow[drr, dashed, bend right, "h"] & Q & \\
 & & & R
\end{tikzcd}$$

commutes. Then since the second square is a pushout this induces a unique map $k : Q \to R$ such that

$$\begin{tikzcd}
X \arrow[r, "\alpha"] \arrow[d, "\beta"'] & Y \arrow[r, "\gamma"] \arrow[d, "\delta"] & Z \arrow[d, "\overline{\delta}"] \arrow[ddr, bend left, "f"] & \\
W \arrow[r, "\overline{\alpha}"'] \arrow[drrr, bend right, "g"'] & P \arrow[r, "\overline{\gamma}"'] \arrow[drr, dashed, bend right, "h"] & Q \arrow[dr, dashed, "k"] & \\
 & & & R
\end{tikzcd}$$

commutes. To be clear, k is the unique map such that $k \circ \overline{\gamma} = h$ and $k \circ \overline{\delta} = f$, not such that $k \circ \overline{\gamma} \circ \overline{\alpha} = g$ and $k \circ \overline{\delta} = f$, so uniquenes remains to be shown. Now suppose we have a map k' making

$$\begin{tikzcd}
X \arrow[r, "\alpha"] \arrow[d, "\beta"'] & Y \arrow[r, "\gamma"] \arrow[d, "\delta"] & Z \arrow[d, "\overline{\delta}"] \arrow[ddr, bend left, "f"] & \\
W \arrow[r, "\overline{\alpha}"'] \arrow[drrr, bend right, "g"'] & P \arrow[r, "\overline{\gamma}"'] & Q \arrow[dr, "k'"] & \\
 & & & R
\end{tikzcd}$$

commute. Then let $h' = k' \circ \overline{\gamma}$. Then the commutativity of the above implies $h' \circ \overline{\alpha} = g$ and $h' \circ \delta = f$, and so by the uniqueness of h we see $h' = h$. Thus $k' \circ \overline{\gamma} = h' = h$, so the uniqueness of k implies $k = k'$. Thus the composed square is a pushout. $\square$

5 Groupoids, Functors, and More Category Theory

In section 3 we saw categories as organizing principles, ways to define highly general concepts and prove highly general theorems. In this section, we study categories as algebraic objects themselves. First, we see that groups can arise out of categories.

Definition 5.1. Let C be a category, and let X be some object of C. Define $\mathrm{Aut}_{\mathsf{C}}(X)$ to be the set of isomorphisms $X \to X$ (called **automorphisms** of X).

Lemma 5.1. For any category C and object X, the set $\mathrm{Aut}_{\mathsf{C}}(X)$ is a group under composition.

Proof. We know composition is associative in any category, and the identity map is an identity for this operation. Then since we only consider isomorphisms, and all isomorphisms are invertible, every element $\varphi \in \mathrm{Aut}_{\mathsf{C}}(X)$ also has an inverse $\varphi^{-1} \in \mathrm{Aut}_{\mathsf{C}}(X)$. Thus it forms a group. □

For instance, the group of all bijections from a set back to itself forms a group (this is called the symmetric group, and is incredibly important in group theory). But not only do categories determine groups, every group also determines a category.

Definition 5.2. Let G be a group. The **categorification** of G is the category $\mathsf{C}(G)$ with a single object $*$ and where the maps $* \to *$ are elements of G. Composition is given by $g \circ h = gh$ and the identity map id_* is the identity element e_G of G.

We can recover G from its categorication by looking at $\mathrm{Aut}_{\mathsf{C}(G)}(*)$. The categorification of a group is more special than other categories we've looked at so far, though. Since every element of a group is invertible, every map in such a categorification is an isomorphism.

Definition 5.3. A category G is a **groupoid** if every map between objects in G is an isomorphism.

Also, the inverse of an isomorphism is unique; to see this, reread the proof of Lemma 4.2 in this new context (also note that the statement "the inverse of an isomorphism is unique" implies Lemma 4.2 by categorifying the group). Thus for any map $p : X \to Y$ in a groupoid, we can speak of its inverse $p^{-1} : Y \to X$. Essentially, groupoids are a generalization of groups, and groups are groupoids with a single object. And just like how we have special maps between groups, we have special maps between groupoids, or more generally between categories.

Definition 5.4. A **functor** $F : \mathsf{C} \to \mathsf{D}$ between categories C and D consists of an object $F(X)$ in D for each object X in C and a map $Ff : FX \to FY$ in D for any map $f : X \to Y$ in C. We also require $F(id_X) = id_{F(X)}$ and $F(g \circ f) = F(g) \circ F(f)$.

The reader should verify that homomorphisms between groups correspond with functors between their categorifications in a natural way. The reader should also not be surprised that we have an identity functor $id_{\mathsf{C}} : \mathsf{C} \to \mathsf{C}$ which is the identity on both objects and morphisms, and a notion of composition of functors.

Definition 5.5. Cat is the category whose objects are categories and whose maps are functors. Grpd is the category whose objects are groupoids and whose maps are functors.

Functors between categories really do preserve the structure we care about. For example, a functor out of a category induces a map on each of the automorphism groups of that category.

Lemma 5.2. For any functor $F : \mathsf{C} \to \mathsf{D}$ and object X of C, there is an group homomorphism $\varphi : \mathrm{Aut}_{\mathsf{C}}(X) \to \mathrm{Aut}_{\mathsf{D}}(F(X))$. If F is an isomorphism of categories, φ is an isomorphism of groups.

Proof. An element f of $\mathrm{Aut}_{\mathsf{C}}(X)$ is an invertible morphism $f : X \to X$. Then $F(f)$ is also invertible, since $F(f) \circ F(f^{-1}) = F(f \circ f^{-1}) = F(id_X) = id_{F(X)}$, and similarly $F(f^{-1}) \circ F(f) = id_{F(X)}$. Thus $\varphi(f) = F(f)$ defines a function $\mathrm{Aut}_{\mathsf{C}}(X) \to \mathrm{Aut}_{\mathsf{D}}(F(X))$. It is a group homomorphism since $\varphi(f \circ g) = F(f \circ g) = F(f) \circ F(g) = \varphi(f) \circ \varphi(g)$ by the fact that F is a functor. If F is invertible, we can apply the same process to F^{-1} and obtain a map $\varphi^{-1}(f) = F^{-1}(f)$, which is clearly an inverse for φ. □

The automorphism group is also useful because the object whose automorphisms we're considering doesn't matter all that much. If X and Y are isomorphism objects of the category C, then $\mathrm{Aut}_{\mathsf{C}}(X)$ and $\mathrm{Aut}_{\mathsf{C}}(Y)$ are isomorphic as groups.

Lemma 5.3. Let X and Y be objects in a category C. Suppose $\varphi : X \to Y$ is an isomorphism in C. Then there is a group isomorphism $\varphi^* : \mathrm{Aut}_{\mathsf{C}}(X) \to \mathrm{Aut}_{\mathsf{C}}(Y)$.

Proof. For an automorphism σ of X, define $\varphi^*(\sigma) = \varphi \circ \sigma \circ \varphi^{-1}$. This composition is well defined by the following (commutative) diagram

$$\begin{tikzcd} X \arrow[r, "\sigma"] & X \arrow[d, "\varphi"] \\ Y \arrow[u, "\varphi^{-1}"] \arrow[r, "\varphi^*(\sigma)"] & Y \end{tikzcd}$$

Also, if σ and τ are automorphisms of X then

$$\varphi^*(\tau \circ \sigma) = \varphi \circ \tau \circ \sigma \circ \varphi^{-1} = \varphi \circ \tau \circ \varphi^{-1} \circ \varphi \circ \sigma \circ \varphi^{-1} = \varphi^*(\tau) \circ \varphi^*(\sigma).$$

Thus $\varphi^*(\sigma)$ is in fact an automorphism, since

$$\varphi^*(\sigma^{-1}) \circ \varphi^*(\sigma) = \varphi^*(\sigma^{-1} \circ \sigma) = \varphi^*(id_X) = \varphi \circ id_X \circ \varphi^{-1} = \varphi \circ \varphi^{-1} = id_Y$$

and similarly $\varphi^*(\sigma) \circ \varphi^*(\sigma^{-1}) = id_Y$. Thus $\varphi^*(\sigma)$ is an isomorphism with inverse $\varphi^*(\sigma)$. φ^* is a homomorphism since, as we checked above, $\varphi^*(\tau \circ \sigma) = \varphi^*(\tau) \circ \varphi^*(\sigma)$. It is an isomorphism because the function $(\varphi^{-1})^* : \mathrm{Aut}_{\mathsf{C}}(Y) \to \mathrm{Aut}_{\mathsf{C}}(X)$ defined by $(\varphi^{-1})^*(\sigma) = \varphi^{-1} \circ \sigma \circ \varphi$ satisfies

$$(\varphi^{-1})^*(\varphi^*(\sigma)) = (\varphi^{-1})^*(\varphi \circ \sigma \circ \varphi^{-1}) = \varphi^{-1} \circ (\varphi \circ \sigma \circ \varphi^{-1}) \circ \varphi = \sigma.$$

Thus $(\varphi^{-1})^* \circ \varphi^* = id_{\mathrm{Aut}_{\mathsf{C}}(X)}$, and by the same argument $\varphi^* \circ (\varphi^{-1})^* = id_{\mathrm{Aut}_{\mathsf{C}}(Y)}$. □

Groupoids, pushouts of groupoids, and the fundamental groupoid (as defined in the next section) will be our most critical tools in showing the circle is not homeomorphic to the interval.

6 Paths, Homotopy, and the Fundamental Groupoid

The intuitive reason that a circle and an interval are different is that one has a hole, while the other does not. Homeomorphisms allow us to shrink, stretch, and generally deform our spaces in wild ways, but they do not allow us to create or remove holes. This is imprecise, but we will make these ideas formal. Firstly, how do we define a hole? A hole exists when we can loop from some point back to itself in some nontrivial way. But before we tackle the question of what "nontrivial" means here, we must define what a loop is.

Definition 6.1. Let X be a space. A **path** between points $x, y \in X$ is a continuous function $p : [0, r] \to X$ for some $r \geq 0$ such that $p(0) = x$ and $p(1) = y$. A **loop** based at a point $x \in X$ is a path from x to x.

Definition 6.2. Let X be a space and suppose we have points $x, y, z \in X$. Further suppose we have paths $p : [0, r] \to X$ between x and y and $q : [0, r'] \to X$ between y and z. Then there is a path $q \bullet p : [0, r + r'] \to X$ between x and z, given by

$$(q \bullet p)(t) = \begin{cases} p(t) & \text{if } 0 \leq t \leq r \\ q(t - r) & \text{if } r \leq t \leq r + r' \end{cases}$$

This path is called the concatenation of p and q. It is well defined since $p(r) = y = q(0) = q(r - r)$, and continuous by the Gluing Lemma.

Definition 6.3. Let X be a space and suppose we have a point $x \in X$. Then there is a path $c_x : [0, 0] \to X$ given by $c_x(t) = x$, called the constant path at x.

It is immediate from the definition of path concatenation that $p \bullet c_x = p$ and $c_y \bullet p = p$ for any path p between points x and y.

Lemma 6.1. Path concatenation is assosciative. That is, for any space X and points $x, y, z, w \in X$, if we have paths $\alpha : [0, r] \to X$ from x to y, $\beta : [0, s] \to X$ from y to z, and $\gamma : [0, k] \to X$ from z to w, $\gamma \bullet (\beta \bullet \alpha) = (\gamma \bullet \beta) \bullet \alpha$.

Proof. We see

$$(\beta \bullet \alpha)(t) = \begin{cases} \alpha(t) & \text{if } 0 \leq t \leq r \\ \beta(t - r) & \text{if } r \leq t \leq r + s \end{cases}$$

Then

$$(\gamma \bullet (\beta \bullet \alpha))(t) = \begin{cases} \alpha(t) & \text{if } 0 \leq t \leq r \\ \beta(t - r) & \text{if } r \leq t \leq r + s \\ \gamma(t - (r + s)) & \text{if } r + s \leq t \leq r + s + k \end{cases}$$

But also

$$(\gamma \bullet \beta)(t) = \begin{cases} \beta(t) & \text{if } 0 \leq t \leq s \\ \gamma(t - s) & \text{if } s \leq t \leq s + k \end{cases}$$

And
$$((\gamma \bullet \beta) \bullet \alpha)(t) = \begin{cases} \alpha(t) & \text{if } 0 \leq t \leq r \\ \beta(t-r) & \text{if } 0 \leq t-r \leq s \\ \gamma((t-r)-s) & \text{if } s \leq t-r \leq s+k \end{cases}$$
And with a little bit of algebra on the bounds, this shows $\gamma \bullet (\beta \bullet \alpha) = (\gamma \bullet \beta) \bullet \alpha$. □

To take stock, we've defined a certain kind of thing with a beginning and end and shown that there is a way of combining these things which has an identity and is assosciative. But that's exactly the data of a category! The previous categories we've seen have had objects as sets with extra structure, and maps as "nice" functions, but the objects in this new category are points, not sets, and the maps are paths, which compose in a completely different way from how functions compose.

Definition 6.4. Let X be a space. Define the path category $\mathsf{P}X$ to be the category whose objects are points in X and where $\mathrm{Hom}_{\mathsf{P}X}(x, y)$ is the set of paths from x to y in X.

Lemma 6.2. The assignment $X \mapsto \mathsf{P}$ is a functor $\mathsf{Top} \to \mathsf{Cat}$.

Proof. We've seen how P acts on objects, so we just need to define it on maps. Suppose we have a continuous function $f : X \to Y$. Then $\mathsf{P}f$ should be a map between $\mathsf{P}X$ and $\mathsf{P}Y$, i.e. a functor. Objects in $\mathsf{P}X$ are just points in X, and we can define $(\mathsf{P}f)(x) = f(x)$. A map $p \in \mathrm{Hom}_{\mathsf{P}X}(x, y)$ is a path from x to x', i.e. a continuous function $p : [0, r] \to X$, and we can make this into a path from $f(x)$ to $f(x')$ by looking at $(\mathsf{P}f)(p) = f \circ p$. Clearly $(\mathsf{P}f)(c_x) = f \circ c_x = c_{f(x)}$, and verifying $(\mathsf{P}f)(q \bullet p) = (\mathsf{P}f)(q) \bullet (\mathsf{P}f)(p)$ is just casework. Thus for any continuous function $f : X \to Y$, we have a functor $\mathsf{P}f : \mathsf{P}X \to \mathsf{P}Y$. This assignment is also functorial since $(\mathsf{P}id_X)(x) = id_X(x) = x = id_{\mathsf{P}X}(x)$ and $(\mathsf{P}id_X)(p) = id_X \circ p = p$. The fact that $\mathsf{P}(g \circ f) = (\mathsf{P}g) \circ (\mathsf{P}f)$ is annoying to verify but requires no trickery. □

Note that $\mathsf{P}X$ is not a groupoid (except for the trivial case where X is empty). In fact, the only isomorphisms are the identity maps. If there's a path p whose domain is $[0, r]$ for $r > 0$, it's inverse would be a path q with domain $[0, s]$ for some $s \geq 0$, and $p \bullet q$ being a constant path would imply $r + s = 0$. This is a little awkward, since travelling along some path p and then going backwards along p is in some sense the same as doing nothing at all. To turn $\mathsf{P}X$ into a groupoid, we need the notion of homotopy.

Definition 6.5. Let X and Y be spaces and suppose we have continuous functions $f, g : X \to Y$. A **homotopy** between f and g is a continuous function $H : X \times \mathbb{I} \to Y$ such that $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$ for all $x \in X$. If a homotopy exists we say f and g are homotopic.

Intuitively, a homotopy between two functions is a continuous family of functions which interpolate between the two. For example, if $X = \mathbb{R}$ and $Y = \mathbb{R}$, we have a homotopy H between $f(x) = x$ and $g(x) = x + 1$ given by $H(x, s) = x + s$. This notion of homotopy isn't quite right for paths, though. For example, the loops p, q defined by

$$\begin{aligned} p, q &: [0, 2\pi] \to \mathbb{S}^1 \\ p(t) &= (0, 0) \\ q(t) &= (\cos(t), \sin(t)) \end{aligned}$$

are homotopic, by homotopy

$$H(t, s) = \begin{cases} (0, 0) & \text{if } 0 \le t \le 2\pi - s \\ (\cos(t), \sin(t)) & \text{if } 2\pi - s \le t \le 2\pi \end{cases}$$

which pulls the right endpoint of q back along the circle. Note that H is continuous by the Gluing Lemma (Lemma 2.12). This homotopy points to an error in our definition, since we're able to contract paths around holes. We're going to try and detect holes in a space by looking at the paths in it which can't be continuously shrunk back to a point. The key idea The problem is that we've allowed ourselves to shift around the endpoints of our paths. The reader should attempt to pull continuously deform the path q into p while holding the endpoints fixed, as this visualization is core to understanding how path-homotopy helps detect holes.

Definition 6.6. Let X be a space and suppose $p, q : [0, r] \to X$ are paths of the same length which both start at x and end at y. Then x and y are **path homotopic**, or **homotopic rel endpoints** if there is some homotopy $H : [0, r] \times \mathbb{I} \to X$ such that $H(t, 0) = x$ and $H(t, 1) = y$ for all $t \in [0, r]$.

The following displays a path homotopy between two curves in the plane.

Homotopy is a useful way to identify multiple different paths. As we noted above, the category $\mathsf{P}X$ is not very nice, since raw paths are too unwieldy. However if we consider path which can be continuously deformed into one another the same, i.e. identify homotopic paths, we obtain a much nicer groupoid called the fundamental groupoid. To perform this quotient, we first need to check that "path-homotopic" is in fact an equivalence relation.

Lemma 6.3. Let $f \approx g$ mean there is some homotopy between f and g. Then $\approx$ is an equivalence relation. If $p \approx' q$ means that p and q are path homotopic, then $\approx'$ is also an equivalence relation.

Proof. Let X, Y be spaces and suppose we have maps $f, g, h : X \to Y$. Then $f \approx f$ by the homotopy $H(x, s) = f(x)$. Also if $f \approx g$, there must be some homotopy G such that $G(x, 0) = f(x)$ and $G(x, 1) = g(x)$. Then let $G'(x, s) = G(x, 1 - s)$. Then $G'(x, 0) = G(x, 1) = g(x)$ and $G'(x, 1) = G(x, 0) = f(x)$, so G is a homotopy between g and f, and thus $g \approx f$. Finally suppose we have G as above and $g \approx h$ by a homotopy F. Then define

$$F \bullet G : X \times \mathbb{I} \to Y$$
$$(F \bullet G)(x, s) = \begin{cases} G(x, 2s) & \text{if } 0 \leq s \leq 0.5 \\ F(x, 2s - 1) & \text{if } 0.5 \leq s \leq 1 \end{cases}$$

This is continuous by the Gluing Lemma (Lemma 2.12), and $(F \bullet G)(x, 0) = G(x, 0) = f(x)$ and $(F \bullet G)(x, 1) = F(x, 1) = h(x)$. Thus $F \bullet G$ is a homotopy between f and h, and so $f \approx h$. Thus homotopy is an equivalence relation. Also, it is easy to check that if f, g, h are paths, then all homotopies defined above are path-homotopies, so path-homotopy is also an equivalence relation. □

We've almost got the "correct" notion of when two paths are equivalent, but we need to be able to compare paths with different domains, i.e. the path c_x should be the same as the path $p : \mathbb{I} \to X$, $p(t) = x$. We also want the it to be that if we trace out a path and then go along that same path in reverse, the resulting concatenated path is equivalent to a constant path. This will mean that every path has an "inverse path", up to equivalence, and so our very large category $\mathsf{P}X$ can be turned into a more reasonable groupoid.

Definition 6.7. Let X be a space. If $p : [0, r] \to X$ and $q : [0, s] \to X$ are paths with the same endpoints x, y, we say p and q are equivalent, written

$p \sim q$, if there are constants a, b such that $r + a = s + b$ and the paths

$$p' : [0, r + a] \to X$$
$$p'(t) = \begin{cases} p(t) & \text{if } 0 \leq t \leq r \\ y & \text{if } r \leq t \leq r + a \end{cases}$$
$$q' : [0, s + b] \to X$$
$$q'(t) = \begin{cases} q(t) & \text{if } 0 \leq t \leq s \\ y & \text{if } s \leq t \leq s + b \end{cases}$$

are path-homotopic. This defines an equivalence relation by a short argument using Lemma 6.3.

Lemma 6.4. Let $p : [0, r] \to X$ be a path from x to y. Then the path q from y to x given by $q(t) = p(r - t)$ satisfies $q \bullet p \sim c_x$ and $p \bullet q \sim c_y$.

Proof. Suppose $p : [0, r] \to X$ is a path from x to y. Define $q : [0, r] \to X$ by $q(t) = p(r - t)$. Then $q(0) = p(r) = y$ and $q(r) = p(0) = x$. Further, we have a homotopy $H : [0, r] \times \mathbb{I} \to X$ between $q \bullet p$ and c'_x, where $c'_x : [0, r + r] \to X$ and $c'_x(t) = x$, defined by

$$H(t, s) = \begin{cases} p(t) & \text{if } 0 \leq t \leq r(1 - s) \\ p(r(1 - s)) & \text{if } r(1 - s) \leq t \leq r(1 + s) \\ q(t - r) & \text{if } r(1 + s) \leq t \leq 2r \end{cases}$$

This is well defined since

$$q(r(1 + s) - r) = q(r(1 + s - s)) = q(rs) = p(r - rs) = p(r(1 - s)).$$

As always, this is continuous by the Gluing Lemma. At some fixed s, the path $H(-, s)$ goes along p for $(1-s)$th of its whole length, then stays constant at the end of this portion of p, and finally goes back to x by q. Visually, we're bringing the endpoint of p back to its initial point.

When $s = 0$, the middle section $r(1 - s) \leq t \leq r(1 + s)$ is only satisfied by $t = r$, so we'll just trace out p and then q. If $s = 1$, then $r(1 - s) = 0$ and $r(1 + s) = 2r$, so the path is always in this middle section, which will be constant at $p(r(1 - s)) = p(0) = x$. Also this is a path-homotopy since $H(0, s) = p(0) = x$ and $H(2r, s) = q(r) = x$ for all s. The equivalence $p \bullet q \sim c_y$ holds similarly. □

Lemma 6.5. Suppose X is a space and we have points $x, y, z \in X$. Further suppose we have paths p, p' from x to y and q, q' from y to z. If $p \sim p'$ and $q \sim q'$ then $q \bullet p \sim q' \bullet p'$.

Proof. For ease of explanation, we assume $p, p' : [0, r] \to X$ have the same domain and $q, q' : [0, r'] \to X$ do as well. The only added difficulty in the case where p and p' have different domains is in bookeeping. Let $H : [0, r] \times \mathbb{I} \to X$ be the path-homotopy between p and p' and $H' : [0, r'] \times \mathbb{I} \to X$ be the one between q and q'. Define $H'' : [0, r + r'] \times \mathbb{I} \to X$

$$H''(t, s) = \begin{cases} H(t, s) & \text{if } 0 \leq t \leq r \\ H'(t - r, s) & \text{if } r \leq t \leq r + r'. \end{cases}$$

First note that $H''(0, s) = H(0, s) = x$ and $H''(r + r', s) = H'(r', s) = z$. Then also

$$H''(t, 0) = \begin{cases} H(t, 0) & \text{if } 0 \leq t \leq r \\ H'(t - r, 0) & \text{if } r \leq t \leq r + r'. \end{cases}$$

Then $H(t, 0) = p(t)$ as H is a homotopy from p to p', and also $H'(t - r, 0) = q(t - r)$. Thus

$$H''(t, 0) = \begin{cases} p(t) & \text{if } 0 \leq t \leq r \\ q(t - r) & \text{if } r \leq t \leq r + r'. \end{cases}$$

Which is $q \bullet p$. A similar computation shows $H(t, 1) = (q' \bullet p')(t)$. $\square$

This shows us that we have well defined concatenation and inverses of paths *up to homotopy*. Since we also showed equivalence of paths is an equivalence relation, this means that we can turn the category $\mathsf{P}X$ into a much nicer category, and in fact a groupoid, by considering each hom set only up to homotopy.

Definition 6.8. Let X be a space. Then we have a groupoid πX, called the **fundamental groupoid** of X, whose objects are the same as those of $\mathsf{P}X$ (i.e. points of X) and where $\mathrm{Hom}_{\pi X}(x, y) = \mathrm{Hom}_{\mathsf{P}X}(x, y)/\sim$. Identity maps are given by $id_x = [c_x]$ and composition by $[q] \circ [p] = [q \bullet p]$, where brackets denote equivalence classes.

Composition is well defined by Lemma 6.5, and πX is a groupoid by Lemma 6.4. Even better, π is still a functor! This is almost immediate from Lemma 6.2, but we need to know that if $p \sim p'$, then $(\mathsf{P}f)(p) \sim (\mathsf{P}f)(p')$.

Lemma 6.6. Let X and Y be spaces and suppose we have a continuous fucntion $f : X \to Y$. Let x, y be points in X and suppose $p : [0, r] \to X$, $p' : [0, r'] \to X$ are paths from x to y. Then if $p \sim p'$ we have $f \circ p \sim f \circ p'$.

Proof. We once again suppose for notational convenience that $r = r'$. Let $H : [0, r] \times \mathbb{I} \to X$ be a homotopy between p and p'. Then $f \circ H$ is a path-homotopy between $f \circ p$ and $f \circ p'$. We check this explicitly

$$\begin{aligned}
(f \circ H)(0, s) &= f(H(0, s)) = f(x) \\
(f \circ H)(r, s) &= f(H(r, s)) = f(y) \\
(f \circ H)(t, 0) &= f(H(t, 0)) = f(p(t)) = (f \circ p)(t) \\
(f \circ H)(t, 1) &= f(H(t, 1)) = f(p'(t)) = (f \circ p')(t)
\end{aligned}$$

Thus the result holds, and so $\mathsf{P}f$ is well defined on equivalence classes of paths. □

Lemma 6.7. The assignment $X \mapsto \pi X$ is a functor $\mathsf{Top} \to \mathsf{Cat}$.

Proof. This follows from Lemmas 6.2 and 6.6 by plumbing around equivalence classes. Explicitly, for a map $f : X \to Y$ the functor $\pi f : \pi X \to \pi Y$ is given by $(\pi f)(x) = f(x)$ on objects and $(\pi f)([p]) = [f \circ p]$ on maps (i.e. equivalence classes of paths). □

The functors π and P are obviously very related. Both send a topological space to a category involving the paths on that space. In a sense, π is just a version of P with fewer messy details. We can capture how they are related by the following map.

Lemma 6.8. For any space X, the mapping $\mathrm{proj}_X : \mathsf{P}X \to \pi X$ which leaves points unchanged and sends paths p to their equivalence classes $[p]$ is a functor.

Proof. By definition, $\mathrm{proj}_X(c_x) = [c_x]$ is the identity in πX for any $x \in X$, and

$$\mathrm{proj}_X(q \bullet p) = [q \bullet p] = [q] \circ [p].$$

□

Lemma 6.9. For any two spaces X, Y and continuous functions $f : X \to Y$, we have the following commutative square in Cat[†]

$$\begin{tikzcd} \mathsf{P}X \arrow[r, "\mathsf{P}f"] \arrow[d, "\mathrm{proj}_X"'] & \mathsf{P}Y \arrow[d, "\mathrm{proj}_Y"] \\ \pi X \arrow[r, "\pi f"'] & \pi Y \end{tikzcd}$$

Proof. For a point $x \in X$

$$(\pi f)(\mathrm{proj}_X(x)) = (\pi f)(x) = f(x) = \mathrm{proj}_Y(f(x)) = \mathrm{proj}_Y((\mathsf{P}f)(x)).$$

For a path $p : x \to y$ in $\mathsf{P}X$,

$$(\pi f)(\mathrm{proj}_X(p)) = (\pi f)([p]) = [f \circ p] = \mathrm{proj}_Y(f \circ p) = \mathrm{proj}_Y((\mathsf{P}f)(p)).$$

Thus the square commutes. □

We often wish to "zoom in" on these fundamental groupoids, picking a certain (usually finite) family of basepoints, and declaring that we only care about paths between those basepoints.

Definition 6.9. Let X be a space and let $A \subseteq X$ be a set a points. Define the groupoid $\pi_A X$ to have the points of A as objects, and maps between those objects equivalence classes of paths in X.

Note that for any two points $x, y \in A$, the set of maps $\mathrm{Hom}_{\pi_A X}(x, y)$ is exactly equal to $\mathrm{Hom}_{\pi X}(x, y)$. We only restrict the number of objects. In the case that A consists of a single point, we obtain a one point groupoid, aka a group. In fact, this will be the automorphism group of πX at some point. If $A = \{a\}$, we call this the fundamental group at the basepoint a, and write it $\pi(X, a)$. This group consists of all loops from a back to itself, up to homotopy. Intuitively, if the fundamental group at a point a is nontrivial, then there is hole in the space X, since we can find loops which can't be continuously deformed into a constant path. A more traditional approach to homotopy focuses only on the fundamental group, ignoring the

[†]This kind of square, with P and π replaced by any two functors with the same domain and codomain and a family of maps between their output like proj, is called the naturality square. The fact that it commutes in this case says proj is a **natural trannsformation** between P and π.

richer groupoid-structure. Such an approach focuses only on path-homotopy classes of loops, instead of path-homotopy classes of paths. This loop-centric perspective can be simpler in general; for example, we can describe loops in X from a point p to p as maps $\ell : \mathbb{S}^1 \to X$ such that $\ell((0,0)) = p$.

In the last part of this section, we prove several lemmas which help determine when two paths are equivalent. Then, using these lemmas, we prove that the fundamental group of the interval is trivial.

Lemma 6.10. Suppose C is a convex subset of $\mathbb{R}^n$. Then for any two points $x, y \in C$ and any paths $p, p' : [0,r] \to X$ from x to y, there is a path-homotopy between p and p'.

Proof. Intuitively, we just connect $p(t)$ and $q(t)$ by a straight line at each time t, and deform the paths into one another along those lines. Formally, define $H(t,s) = (1-s)p(t) + sq(t)$. Then H is well defined as a function with codomain C since $p(t)$ and $q(t)$ are always in C, and C is convex so any point on the line $(1-s)p(t) + sq(t)$ for $s \in \mathbb{I}$ is also in C. We further see $H(t,0) = p(t)$ and $H(t,1) = q(t)$ and $H(0,s) = (1-s)p(0) + sq(0) = (1-s)x + sx = x$ and similarly $H(r,s) = (1-s)p(r) + sq(r) = (1-s)y + sy = y$. □

Lemma 6.11. For any r, $[0,r]$ is convex. In particular, $\mathbb{I}$ is convex.

Proof. Let a, b be such that $0 \leq a \leq r$ and $0 \leq b \leq r$. Then $0 \leq ta \leq tr$ and $0 \leq (1-t)b \leq (1-t)r$. But then adding these inequalities, $0 \leq ta + (1-t)b \leq tr + (1-t)r = r$. □

Lemma 6.12. For any path $p : [0,r] \to X$, if $\overline{p} : \mathbb{I} \to X$ is the path $\overline{p} = p(rt)$ then $p \sim p'$.

Proof. Suppose $r \geq 1$. Then if $p^* : [0,r] \to X$ is $p(rt)$ when $0 \leq t \leq 1$ and $p(r)$ for $t \geq 1$, it suffices to give a homotopy $H : [0,r] \times \mathbb{I} \to X$ from p to p^*. Define this by

$$H(t,s) = \begin{cases} p(((r-1)s+1)t) & \text{if } 0 \leq t \leq \frac{r}{(r-1)s+1} \\ p(r) & \text{if } \frac{r}{(r-1)s+1} \leq t \leq r \end{cases}$$

Then this is endpoint preserving since $H(0,s) = p(((r-1)s+1) \cdot 0) = p(0)$ and $H(r,s) = p(r)$. It is a homotopy since when $s = 0$, $(r-1)s + 1 = 1$, and thus the paths' entire time is spent tracing out $p(((r-1)s+1)t) = p(t)$. If $s = 1$, then $(r-1)s + 1 = r$, so the first chunk is just the function $p(rt)$ from $t = 0$ to $t = 1$ and the rest is $p(r)$, i.e. we trace out $p^*(t)$. If $r < 1$, the proof is essentially the same. □

Lemma 6.13. Let X be a space and $p : [0, r] \to X$ be a path in X. Then for $r' > 0$ and continuous map $\sigma : [0, r'] \to [0, r]$ such that $\sigma(0) = 0$ and $\sigma(r') = r$, we have an equivalence $p \sim p \circ \sigma$. In particular, taking $r = 1$ and $\sigma(t) = \frac{t}{r'}$, see see each path is equivalent to one out of the unit interval.

Proof. We can consider σ as a path from 0 to r in $[0, r]$, and then by Lemmas 6.10 and 6.11, $\sigma \sim \alpha$, where $\alpha : [0, r'] \to [0, r]$, $\alpha(t) = \frac{r}{r'} \cdot t$. Thus by Lemma 6.6, $p \circ \sigma \sim p \circ \alpha$, so it suffies to show $p \circ \alpha \sim p$. With notation as in the last lemma, $p \circ \alpha \sim \overline{p \circ \alpha}$ and $p \sim \overline{p}$. But $\overline{p \circ \alpha} = \overline{p}$, since for any $t \in \mathbb{I}$,

$$\overline{p \circ \alpha}(t) = (p \circ \alpha)(r't) = p(\alpha(r't)) = p\left(\frac{r}{r'} \cdot r't\right) = p(rt) = \overline{p}(t).$$

And thus $p \circ \sigma \sim p \circ \alpha \sim \overline{p \circ \alpha} = \overline{p} \sim p$. □

Lemma 6.14. Let x be any point in $\mathbb{I}$. Then $\pi(\mathbb{I}, x)$ is the trivial group with one element.

Proof. $\mathbb{I}$ is convex, so by Lemma 6.10 all loops from x to x are equivalent. Thus $\mathrm{Hom}_{\pi X}(x, x) = \mathrm{Hom}_{\mathsf{P}X}(x, x)/\sim$ has a single equivalence class. □

This points at a way to show that $\mathbb{I}$ and $\mathbb{S}^1$ are topologically different: show that $\mathbb{S}^1$ has a nontrivial fundamental group at some point. We embark on this in the next section. We close the section by proving two lemmas which help to prove equivalence in more complicated situations. The first requires is more natural in a more traditional loop-centric development of homotopy theory. Such an approach can avoid the complexity of groupoids, instead working only in terms of the fundamental group of a space. Like in the following lemma, loops can be simpler than arbitrary paths.

Lemma 6.15. Suppose that X is a space, p is some point of X, and $f : [0, r] \to X$ a loop at p. Define $\tilde{f} : \mathbb{S}^1 \to X$ by $\tilde{f}(\cos(t), \sin(t)) = f\left(\frac{r}{2\pi} t\right)$ for $t \in [0, 2\pi)$. This is continuous since $\tilde{f}(1, 0) = f(0) = p = f(r) = \lim_{t \to 2\pi^-} f\left(\frac{r}{2\pi} \cdot t\right) = \lim_{t \to 2\pi^-} \tilde{f}(\cos(t), \sin(t))$. If $\tilde{f}$ is homotopic to the constant map $x \mapsto c$ for some $c \in X$, then f is equivalent to the trivial loop c_p at p.

Proof. We can assume w.l.o.g. by Lemma 6.12 that $r = 1$. Let $H : \mathbb{S}^1 \times \mathbb{I} \to X$ be a homotopy from $\tilde{f}$ to $x \mapsto c$. Now define $G : [0, 3] \times \mathbb{I} \to X$ by

$$G(t, s) = \begin{cases} H((1, 0), st) & \text{if } 0 \le t \le 1 \\ H((\cos(2\pi(t-1)), \sin(2\pi(t-1))), s) & \text{if } 1 \le t \le 2 \\ H((1, 0), s(1 - (t-2)) & \text{if } 2 \le t \le 3. \end{cases}$$

It's easy to check that these cases line up when t is 1 or 2, so this is well defined and it is continuous by the Gluing Lemma. We also see that $G(0,s) = H((1,0),0) = \tilde{f}(1,0) = p$ and $G(3,s) = H((1,0),0) = p$, so $t \mapsto G(t,s_0)$ is a loop at p for each $s_0 \in \mathbb{I}$. Thus G is a path-homotopy between the loops $t \mapsto G(t,0)$ and $t \mapsto G(t,1)$, so it suffices to show $t \mapsto G(t,0)$ is equivalent to f and $t \mapsto G(t,1)$ equivalent to the constant loop at p. Define the paths $k_1, k_2, k_3, h_1, h_2, h_3 : \mathbb{I} \to X$ by

$$\begin{aligned}
k_1(t) &= G(t,0) = H((1,0),0) = \tilde{f}(1,0) = p \\
k_2(t) &= G(t+1,0) = H((\cos(2\pi t), \sin(2\pi t)),0) = \tilde{f}(\cos(2\pi t), \sin(2\pi t)) = f(t) \\
k_3(t) &= G(t+2,0) = H((1,0),0) = \tilde{f}(1,0) = p \\
h_1(t) &= G(t,1) = H((1,0),t) \\
h_2(t) &= G(t+1,1) = H((\cos(2\pi t), \sin(2\pi t)),1) = c \\
h_3(t) &= G(t+2,1) = H((1,0),1-t).
\end{aligned}$$

Thus G is a path-homotopy between $k_3 \bullet k_2 \bullet k_1$ and $h_3 \bullet h_2 \bullet h_1$. It's immediate by the definition of equivalence that k_1 and k_3 are equivalent to c_p, so by Lemma 6.5, $k_3 \bullet k_2 \bullet k_1 \sim c_p \bullet k_2 \bullet c_p = k_2 = f$. By the same argument, $h_3 \bullet h_2 \bullet h_1 \sim h_3 \bullet h_1$. But $h_3(t) = h_1(1-t)$, so by Lemma 6.4, $h_3 \bullet h_1 \sim c_p$. Thus

$$f = k_2 \sim k_3 \bullet k_2 \bullet k_1 \sim h_3 \bullet h_2 \bullet h_1 \sim h_3 \bullet h_1 \sim c_p.$$

□

The next lemma will be used in our proof of the Seifert-van Kampen Theorem. Essentially, if we have an embedding F of a rectangle R into some space X, then the restriction of F to the left and upper sides is path, as is its restriction to the bottom and right sides. Then we can deform these lines in to one another in R, and pushing this deformation through F gives us a path-homotopy between the paths given by restriction of F in X. The proof is mostly bookeeping and fidgeting with homotopies.

Lemma 6.16 (The Rectangle Lemma)**.** *Let X be a space, and suppose for real numbers $a < b$ and $c < d$ we have a function $F : [a,b] \times [c,d] \to X$.*

Define the paths

$$\begin{aligned} p, q' &: [0, d-c] \to X \\ p', q &: [0, b-a] \to X \\ p(t) &= F(a, t+c) \\ p'(t) &= F(t+a, d) \\ q(t) &= F(t+a, c) \\ q'(t) &= F(b, t+c). \end{aligned}$$

Geometrically, p goes up the left edge of the rectangle, p' goes rightwards on the top, q goes rightwards on the bottom, and q' goes up along the right. Then $p' \bullet p \sim q' \bullet q$.

Proof. Consider the maps $\varphi : [0, b-a] \to [0, d-c]$ and $\psi : [0, d-c] \to [0, b-a]$ given by $\varphi(t) = \left(\frac{d-c}{b-a}\right) t$ and $\psi(t) = \left(\frac{b-a}{d-c}\right) t$. By Lemmas 6.13 and 6.5, $p' \bullet p \sim (p' \circ \psi) \bullet (p \circ \varphi)$, so it suffices to give a path-homotopy between $(p' \circ \psi) \bullet (p \circ \varphi)$ and $q' \bullet q$.

Define $H_1 : [0, b-a] \times \mathbb{I} \to [a, b] \times [c, d]$ by

$$H_1(t, s) = \left(st + a, \left(\frac{d-c}{b-a}\right)(1-s)t + c\right).$$

Though it is hard to tell from the algebra, we're essentially choosing a point $w(s) = ((b-a)s + a, (d-c)(1-s) + c)$ along the antidiagonal $\{(x, y) \in [a, b] \times [c, d] : \frac{x-a}{b-a} + \frac{y-c}{d-c} = 1\}$ of the rectangle, and for each fixed s tracing out the line from (a, c) to $w(s)$. Now define $H_2 : [0, d-c] \times \mathbb{I} \to [a, b] \times [c, d]$ by

$$H_2(t, s) = \left(\left(\frac{b-a}{d-c}\right)((d-c)s + (1-s)t) + a, (d-c)(1-s) + st + c\right).$$

Similarly this traces out a path from $w(s)$ to $(b-a, d-c)$. Finally, let $H : [0, (b-a) + (d-c)] \times \mathbb{I} \to X$ be given by

$$H(t, s) = \begin{cases} F(H_1(t, s)) & \text{if } 0 \le t \le b-a \\ F(H_2(t-(b-a), s)) & \text{if } b-a \le t \le (b-a) + (d-c) \end{cases}$$

This is well defined because $H_1(b-a, s) = w(s) = H_2(0, s)$, and continuous by the Gluing Lemma. Then for any s,

$$H_1(0, s) = \left(s \cdot 0 + a, \left(\frac{d-c}{b-a}\right)(1-s) \cdot 0 + c\right) = (a, c).$$

And

$$H_2(d-c,s) = \left(\left(\frac{b-a}{d-c} \right) ((d-c)s + (1-s)(d-c)) + a, (d-c)(1-s) + s(d-c) + c \right) = (b,d).$$

Since $(d-c)s + (1-s)(d-c) = d-c$. This shows that

$$H(0,s) = F(H_1(0,s)) = F(a,c)$$

and

$$H((d-c)+(b-a),s) = F(H_2(d-c),s) = F(b,d).$$

So H preserves endpoints. Then also

$$\begin{aligned}
F(H_1(t,0)) &= F\left(0 \cdot t + a, \left(\frac{d-c}{b-a}\right)(1-0)t + c\right) \\
&= F\left(a, \left(\frac{d-c}{b-a}\right)t + c\right) \\
&= p(\varphi(t)) \\
F(H_2(t,0)) &= F\left(\left(\frac{b-a}{d-c}\right)((d-c)\cdot 0 + (1-0)t) + a, (d-c)(1-0) + 0 \cdot t + c\right) \\
&= F\left(\left(\frac{b-a}{d-c}\right)t + a, d\right) \\
&= p'(\psi(t)).
\end{aligned}$$

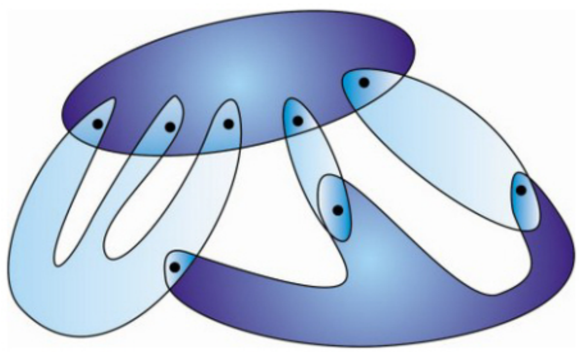
Which means $H(t,0) = ((p' \circ \psi) \bullet (p \circ \varphi))(t)$. Then also

$$\begin{aligned}
F(H_1(t,1)) &= F\left(1 \cdot t + a, \left(\frac{d-c}{b-a}\right)(1-1)t + c\right) \\
&= F(t+a, c) \\
&= q(t) \\
F(H_2(t,1)) &= F\left(\left(\frac{b-a}{d-c}\right)((d-c)\cdot 1 + (1-1)t) + a, (d-c)(1-1) + 1 \cdot t + c\right) \\
&= F(b, t+c) \\
&= q'(t).
\end{aligned}$$

Thus H is a path-homotopy from $(p' \circ \psi) \bullet (p \circ \varphi)$ to $q' \bullet q$, and so we are done. $\square$

7 Some Seifert-van Kampen Theorems

Showing that the circle has a hole in it is not simple task. A more traditional approach would be to define the winding number of a loop around the circle, but (as the reader may have noticed) this paper prefers groupoid-oriented techniques. There is a classical theorem of algebraic topology called the Seifert-van Kampen theorem which, given an open cover $\{U, V\}$ of a space, determines the fundamental group of that sepac in terms of the fundamental groups U, V, and $U \cap V$. However the classical formulation has an issue when the intersection of these subsets is not connected, as illustrated in the following diagram.



If the lighter shaded area is X_1 and the darker one X_2, and we know the fundamental groupoid of X_1 and X_2 at the indicated points, the classical Seifert-van Kampem theorem is not strong enough to determine the fundamental group of the whole space. We can remedy this by focusing not on fundamental *groups*, but on *groupoids*, since groupoids can keep track of the fundamental group at any number of points.

For the rest of this section let X be a space with subspaces X_1, X_2 such that $\operatorname{Int} X_1 \cup \operatorname{Int} X_2 = X$, and define $X_0 = X_1 \cap X_2$. This basic setup immediately gives us some categorical structure.

Lemma 7.1. Let $\iota_1 : X_0 \to X_1$, $\iota_2 : X_0 \to X_2$, $i_1 : X_1 \to X$, and $i_2 : X_2 \to X$, all be inclusion maps. Then we have a pushout square

$$\begin{tikzcd} X_0 \arrow[r, "\iota_1"] \arrow[d, "\iota_2"'] & X_1 \arrow[d, "i_1"] \\ X_2 \arrow[r, "i_2"'] & X \end{tikzcd}$$

Proof. This square is automatically commutative, since for any $x \in X_0$, $i_1(\iota_1(x)) = x = i_2(\iota_2(x))$. Now suppose we have another pushout candidate

$$\begin{tikzcd} X_0 \arrow[r, "\iota_1"] \arrow[d, "\iota_2"'] & X_1 \arrow[d, "f"] \\ X_2 \arrow[r, "g"'] & P \end{tikzcd}$$

Then $X_1 \setminus X_2 \subseteq X \setminus X_2 = (\operatorname{Int} X_1 \cup \operatorname{Int} X_2) \setminus X_2 = \operatorname{Int} X_1 \setminus X_2 \subseteq \operatorname{Int} X_1$, and similarly $X_2 \setminus X_1 \subseteq \operatorname{Int} X_2$. By commutativity of the diagram and the Gluing Lemma, we have a continuous function $h : X \to P$ where h is f on X_1 and g on X_2. Thus we have the square

$$\begin{tikzcd} X_0 \arrow[r, "\iota_1"] \arrow[d, "\iota_2"'] & X_1 \arrow[d, "i_1"] \arrow[ddr, bend left, "f"] & \\ X_2 \arrow[r, "i_2"'] \arrow[drr, bend right, "g"'] & X \arrow[dr, dashed, "h"] & \\ & & P \end{tikzcd}$$

We prove this h is unique. If there were another $h' : X \to P$ making the above commute, we would have $h' \circ i_1 = f$ and $h' \circ i_2 = g$, so $h' = h$ on both X_1 and X_2, which implies $h = h'$ as $X_1 \cup X_2 = X$. Thus the map h is unique, and so the square is a pushout. $\square$

We now begin our proof of the Seifert-van Kampen theorem. Throughout, we will refer to the squares

$$\begin{tikzcd} X \arrow[r, "\iota_1"] \arrow[d, "\iota_2"'] & X_1 \arrow[d, "i_1"] \\ X_2 \arrow[r, "i_2"'] & X \end{tikzcd} \qquad \begin{tikzcd} \mathsf{P}X_0 \arrow[r, "\mathsf{P}\iota_1"] \arrow[d, "\mathsf{P}\iota_2"'] & \mathsf{P}X_1 \arrow[d, "\mathsf{P}i_1"] \\ \mathsf{P}X_2 \arrow[r, "\mathsf{P}i_2"'] & \mathsf{P}X \end{tikzcd} \qquad \begin{tikzcd} \pi X_0 \arrow[r, "\pi\iota_1"] \arrow[d, "\pi\iota_2"'] & \pi X_1 \arrow[d, "\pi i_1"] \\ \pi X_2 \arrow[r, "\pi i_2"'] & \pi X \end{tikzcd}$$

as $\mathbf{X}, \mathsf{P}\mathbf{X}$, and $\pi\mathbf{X}$, respectively. We know that this first square is commutative, and since functors preserve commutatitivty, so are the second two.

Lemma 7.2. The square $\mathsf{P}\mathbf{X}$ is a pushout (in Cat).

Proof. Suppose we have the diagram

$$\begin{tikzcd} \mathsf{P}X_0 \arrow[r, "\mathsf{P}\iota_1"] \arrow[d, "\mathsf{P}\iota_2"'] & \mathsf{P}X_1 \arrow[d, "f"] \\ \mathsf{P}X_2 \arrow[r, "g"'] & \mathsf{C} \end{tikzcd}$$

We construct a map $h : \mathsf{P}X \to \mathsf{C}$. We define h on objects in X_1 as f and by objects in X_2 as g, as in Lemma 7.1. Now if p is a path $x \to y$ in X such that $\operatorname{im} p \subseteq X_1$ or $\operatorname{im} p \subseteq X_2$, we must have either $p = i_1 \circ p_1 = (\mathsf{P}i_1)(p_1)$ for a path p_1 in X_1 or $p = i_1 \circ p_2 = (\mathsf{P}i_2)(p_2)$ for a path p_2 in X_2. In the first case, let $h(p) = f(p_1)$, and in the second let $h(p) = g(p_2)$. This is well defined since if $(\mathsf{P}i_1)(p_1) = p = (\mathsf{P}i_2)(p_2)$, then p_1, p_2 must in fact both be the inclusion $(\mathsf{P}\iota_1)(p')$ and $(\mathsf{P}\iota_2)(p')$ of a path p' in X_0. Then

$$f(p_1) = f((\mathsf{P}\iota_1)(p')) = g((\mathsf{P}\iota_2)(p')) = g(p_2)$$

Now suppose p is an arbitrary path $p : x \to y$, say $p : [0, r] \to X$. If we have a finite sequence of points $0 = a_0 < a_1 < \ldots < a_{n-1} < a_n = r$, we can subdivide p into $p = q_{n-1} \bullet q_{n-2} \bullet \ldots \bullet q_1 \bullet q_0$, where

$$\begin{aligned} q_j &: [0, a_{j+1} - a_j] \to X \\ q_j(t) &= p\,(t + a_j) \end{aligned}$$

Essentially we've broken up $[0, r]$ into $[a_0, a_1] \cup [a_1, a_2] \cup \ldots [a_{n-1}, a_n]$, then restricted p to each of those subintervals. In particular, if we have a partition $\{a_0, a_1, \ldots, a_{n-1}, a_n\}$ of $[0, r]$ such that $p([a_j, a_{j+1}]) \subseteq X_1$ or $p([a_j, a_{j+1}]) \subseteq X_2$ for each j, then we can define $h(p) = h(q_{n-1}) \circ h(q_{n-2}) \circ \ldots \circ h(q_1) \circ h(q_0)$. Such a subdivision always exists by the Lebesgue Covering Lemma (Lemma 3.13); take the open cover $\{p^{-1}(\operatorname{Int} X_1), p^{-1}(\operatorname{Int} X_1)\}$ and choose a_j such that $a_{j+1} - a_j < \delta$. However it's not immediately obvious that the definition is independent of which partition we choose. So suppose we had two partition $\{a_0, a_1, \ldots, a_{n-1}, a_n\}$ and $\{b_0, b_1, \ldots, b_{k-1}, b_k\}$, giving rise to families $\{q_j\}_{j=0}^{n-1}$ and $\{q'_j\}_{j=0}^{k-1}$ of paths as above. Now consider the partition

$$\{c_0, c_1, \ldots, c_{m-1}, c_m\} = \{a_0, a_1, \ldots, a_{n-1}, a_n\} \cup \{b_0, b_1, \ldots, b_{k-1}, b_k\}.$$

Where we take the union on the right and then enumerate it by the c_j such that $c_0 < c_1 < \ldots < c_{m-1} < c_m$. Define $\{q''_j\}_{j=0}^{m}$ to be the paths induced by this partition. Since the partition $\{c_j\}$ is the union of $\{a_j\}$ and $\{b_j\}$, for each j there is some ℓ_j such that $a_j = c_{\ell_j}$, and so

$$q_j = q''_{\ell_{j+1}-1} \bullet q''_{\ell_{j+1}-2} \bullet \ldots \bullet q''_{\ell_j+1} \bullet q''_{\ell_j}.$$

Now by assumption q_j is either contained entirely within X_1 or within X_2, so $h(q_j) = f(q_j)$ or $h(q_j) = g(q_j)$. In this first case,

$$h(q_j) = f(q_j) = f(q''_{\ell_{j+1}-1}\bullet q''_{\ell_{j+1}-2}\bullet\ldots\bullet q''_{\ell_j+1}\bullet q''_{\ell_j}) = f(q''_{\ell_{j+1}-1})\circ f(q''_{\ell_{j+1}-2})\circ\ldots\circ f(q''_{\ell_j+1})\circ f(q''_{\ell_j})$$

since f is a functor. But also if $\ell_j \leq i < \ell_{j+1}$, then $\operatorname{im} q''_i \subseteq \operatorname{im} p_j \subseteq X_1$, so $f(q''_i) = h(q''_i)$. Thus

$$h(q_j) = h(q''_{\ell_{j+1}-1}) \circ h(q''_{\ell_{j+1}-2}) \circ \ldots \circ h(q''_{\ell_j+1}) \circ h(q''_{\ell_j})$$

for each j. The same argument applies when $\operatorname{im} q_j \subseteq X_2$, and thus

$$\bigcirc_{j=0}^{n-1} h(q_j) = \bigcirc_{j=0}^{n-1} h\left(\bigbullet_{i=\ell_j}^{\ell_{j+1}-1} q''_i\right) = \bigcirc_{j=0}^{n-1}\bigcirc_{i=\ell_j}^{\ell_{j+1}-1} h(q''_i) = \bigcirc_{i=\ell_0}^{\ell_n-1} h(q''_i)$$

Then since $a_0 = 0 = c_0$ and $a_n = r = c_m$, we have $\ell_0 = 0$ and $\ell_n = m$, and so this equality says $\bigcirc_{j=0}^{n-1} h(q_j) = \bigcirc_{i=0}^{m-1} h(q''_i)$. This same argument shows $\bigcirc_{j=0}^{n-1} h(q'_j) = \bigcirc_{i=0}^{m-1} h(q''_i)$, so $\bigcirc_{j=0}^{n-1} h(q_j) = \bigcirc_{i=0}^{m} h(q'_i)$. Thus our definition of $h(p)$ is independent of choice of subdivision, and hence is well defined.

We still need to verify that h is a functor, though. For any point $x \in X$, either $x \in X_1$ or $x \in X_2$. In this first case, $h(c_x) = f(c_x) = id_{f(x)} = id_{h(x)}$, and similarly for the second. Then if p, q are paths, we can subdivide them by $p = p_{n-1} \bullet \cdots \bullet p_1 \bullet p_0$ and $q = q_{m-1} \bullet \cdots \bullet q_1 \bullet q_0$ such that the image of each of these subpaths is contained in X_1 or X_2, and so that

$$q \bullet p = q_{m-1} \bullet \cdots \bullet q_1 \bullet q_0 \bullet p_{n-1} \bullet \cdots \bullet p_1 \bullet p_0.$$

Is an appropriate subdivision of $q \bullet p$. But then by definition of h,

$$h(q\bullet p) = h(q_{m-1})\circ\cdots\circ h(q_1)\circ h(q_0)\circ h(p_{n-1})\circ\cdots\circ h(p_1)\circ h(p_0) = h(q)\circ h(p)$$

Thus h preserves the identity and composition, and so is a well defined functor. Also if the image of a path p is contained in X_1 or X_2, i.e. if p in a morphism in $\mathsf{P}X_1$ or $\mathsf{P}X_2$, then $h(p) = f(p)$ or $h(p) = g(p)$ since p is a subdivision of itself. Thus $h \circ \mathsf{P}i_1 = f$ and $h \circ \mathsf{P}i_1 = g$, giving us the commutative diagram

$$\begin{tikzcd}
\mathsf{P}X_0 \arrow[r, "\mathsf{P}\iota_1"] \arrow[d, "\mathsf{P}\iota_2"'] & \mathsf{P}X_1 \arrow[d, "\mathsf{P}i_1"] \arrow[ddr, bend left, "f"] & \\
\mathsf{P}X_2 \arrow[r, "\mathsf{P}i_2"'] \arrow[drr, bend right, "g"'] & \mathsf{P}X \arrow[dr, dashed, "h"] & \\
& & \mathsf{C}
\end{tikzcd}$$

Now suppose there was another $h' : \mathsf{P}X \to \mathsf{C}$ making this commute. Since $X_1 \cup X_2 = X$, for any point $x \in X$ either $x \in X_1$ or $x \in X_2$, so by commutativity $h'(x) = f(x) = h(x)$ or $h'(x) = g(x) = h(x)$. Thus $h = h'$ on objects. Then for any path p, if we subdivide p into $p_{n-1} \bullet \cdots \bullet p_1 \bullet p_0$ with each p_j having image contained in X_1 or X_2, then by functorality

$$h'(p) = h'(p_{n-1}) \circ \cdots \circ h'(p_1) \circ h'(p_0)$$

And each of these paths is either in $\mathsf{P}X_1$ or in $\mathsf{P}X_2$, so by commutativity $h'(p_j) = f(p_j) = h(p_j)$ or $h'(p_j) = g(p_j) = h(p_j)$. Thus

$$h'(p) = h'(p_{n-1}) \circ \cdots \circ h'(p_1) \circ h'(p_0) = h(p_{n-1}) \circ \cdots \circ h(p_1) \circ h(p_0) = h(p)$$

Which means $h = h'$ both on objects and maps, so they are equal as functors. Thus h is unique, and so $\mathsf{P}X$ is a pushout. $\square$

This lemma is almost what we want, but we need to show that when C is a groupoid G, this induced map h on paths agrees on homotopy classes of paths.

Theorem 7.3 (The Seifert van-Kampen Theorem for Groupoids)**.** The square $\pi\mathbf{X}$ is a pushout of groupoids.

Proof. Suppose we have the diagram of groupoids

$$\begin{tikzcd}
\pi X_0 \arrow[r, "\pi\iota_1"] \arrow[d, "\pi\iota_2"'] & \pi X_1 \arrow[d, "f"] \\
\pi X_2 \arrow[r, "g"'] & \mathsf{G}
\end{tikzcd}$$

Define $f' : \mathsf{P}X_1 \to \mathsf{G}$ by $f' = f \circ \mathrm{proj}_{X_1}$, and $g' : \mathsf{P}X_2 \to \mathsf{G}$ similarly. Then by Lemma 6.9, we have the two commutative squares

$$\begin{tikzcd}
\mathsf{P}X_0 \arrow[r, "\mathsf{P}\iota_1"] \arrow[d, "\mathrm{proj}_{X_0}"'] & \mathsf{P}X_1 \arrow[d, "\mathrm{proj}_{X_1}"] \\
\pi X_0 \arrow[r, "\pi\iota_1"] & \pi X_1
\end{tikzcd} \quad \text{and} \quad \begin{tikzcd}
\mathsf{P}X_0 \arrow[r, "\mathsf{P}\iota_2"] \arrow[d, "\mathrm{proj}_{X_0}"'] & \mathsf{P}X_2 \arrow[d, "\mathrm{proj}_{X_2}"] \\
\pi X_0 \arrow[r, "\pi\iota_2"] & \pi X_2
\end{tikzcd}$$

Then gluing these along the proj_{X_0} edge and gluing the square above to $\pi X_2 \leftarrow \pi X_0 \to \pi X_1$, we get the commutative diagram

$$\begin{tikzcd}
\mathsf{P}X_0 \arrow[rr, "\mathsf{P}\iota_2"'] \arrow[dd, "\mathsf{P}\iota_2"] \arrow[dr, "\mathrm{proj}_{X_0}"] & & \mathsf{P}X_1 \arrow[d, "\mathrm{proj}_{X_1}"] \\
& \pi X_0 \arrow[r, "\pi\iota_1"] \arrow[d, "\pi\iota_2"] & \pi X_1 \arrow[d, "f"] \\
\mathsf{P}X_2 \arrow[r, "\mathrm{proj}_{X_2}"] & \pi X_2 \arrow[r, "g"] & \mathsf{G}
\end{tikzcd}$$

Let $f' = f \circ \mathrm{proj}_{X_1}$ and $g' = g \circ \mathrm{proj}_{X_1}$. Then considering the outside square and applying our previous lemma, there is a unique map $h : \mathsf{P}X \to \mathsf{G}$ such that the following commutes

$$\begin{tikzcd}
\mathsf{P}X_0 \arrow[r, "\mathsf{P}\iota_1"] \arrow[d, "\mathsf{P}\iota_2"'] & \mathsf{P}X_1 \arrow[d, "\mathsf{P}i_1"] \arrow[ddr, bend left, "f'"] & \\
\mathsf{P}X_2 \arrow[r, "\mathsf{P}i_2"'] \arrow[drr, bend right, "g'"'] & \mathsf{P}X \arrow[dr, dashed, "h"] & \\
& & \mathsf{G}
\end{tikzcd}$$

We define $\tilde{h} : \pi X \to \mathsf{G}$ in terms of h. Since $\mathsf{P}X$ and πX have the same objects, we can set $\tilde{h}(x) = h(x)$ for any $x \in X$. Now to define $\tilde{h}$ on maps in πX, it suffices to show that $h(p) = h(q)$ whenever $p \sim q$, since the maps in πX are equivalence classes of maps in $\mathsf{P}X$. First note that if $p : [0, r] \to X$ is a path from x to y and $c : [0, r'] \to X$ is a path constant at y, then $h(c \bullet p) = h(c) \circ h(p)$, and since c is contant, either $h(c) = f'(c) = f(\mathrm{proj}_{X_1}(c)) = f([c_x]) = id_{h(y)}$ or $h(c) = g'(c) = g([c_x]) = id_{h(y)}$, since $c \sim c_y$. Thus $h(c \bullet p) = h(c) \circ h(p) = id_{h(y)} \circ h(p) = h(p)$. Thus it suffices to show

$h(p) = h(q)$ when p and q have the same domain and are path homotopic (as otherwise we can preppend such a path c to each).

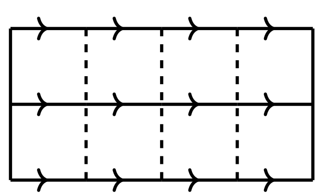
Let $p, q : [0, r] \to X$ be such paths and suppose $H : [0, r] \times \mathbb{I} \to X$ is a path-homotopy. For each n define g

$$R^n_{i,j} = \left[\frac{i}{n}r, \frac{(i+1)}{n}r\right] \times \left[\frac{j}{n}, \frac{(j+1)}{n}\right].$$

So that for any n,

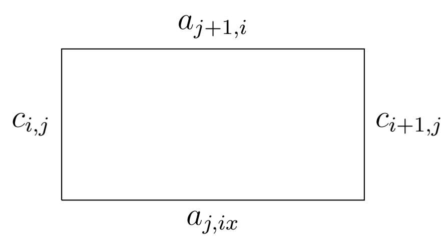
$$[0, r] \times \mathbb{I} = \bigcup_{i=0}^{n-1} \bigcup_{j=0}^{n-1} R^n_{i,j}.$$

Geometrically, we've broken up the domain of H into n^2 equal sized rectangles. Now note that $\{H^{-1}(\operatorname{Int} X_1), H^{-1}(\operatorname{Int} X_2)\}$ is an open cover of X, and so by the Lebesgue Covering Lemma (Lemma 3.13), there is some $\delta > 0$ such that any ball of radius less than δ is contained in $H^{-1}(\operatorname{Int} X_1)$ or $H^{-1}(\operatorname{Int} X_2)$. Now if we choose n such that each $R^n_{i,j}$ is contained in a ball of radius smaller than δ, this means that for each n and $0 \leq i, j < n$, either $H(R^n_{i,j}) \subseteq X_1$ or $H(R^n_{i,j}) \subseteq X_2$. Now for $0 \leq j \leq n$ define the paths $a_j : [0, r] \to X$ by $a_j(t) = H\left(t, \frac{j}{n}\right)$, and for $0 \leq i < n$ define $a_{j,i} : [0, r/n] \to X$ by $a_{j,i}(t) = H\left(\frac{i}{n}r + t, \frac{j}{n}\right)$. Essentially we've taken each horizontal line in our partition $\{R^n_{i,j}\}$ and called the path that H traces out along it a_j, then partitioned a_j at the intersection with each vertical line.



Also note that $a_0(t) = H(t, 0) = p(t)$ and $a_n = H(t, 1) = q(t)$, since H is a homotopy. Now instead of looking at H on the horizontal lines, we look at it on the vertical segments.

For $0 \leq i \leq n$ and $0 \leq j < n$, define $c_{i,j} : [0, 1/n] \to X$ by $c_{i,j}(t) = H\left(\frac{i}{n}r, t + \frac{j}{n}\right)$. This is essentially the the restriction of H to the segment on the ith vertical line, going from the jth to the $(j+1)$st horizontal line. Now consider the restriction $H_{i,j} : R^n_{i,j} \to X$ of H to $R^n_{i,j}$. Then $H_{i,j}(R^n_{i,j})$ looks something like



Thus by the Rectangle Lemma (Lemma 6.16), $a_{j+1,i} \bullet c_{i,j} \sim c_{i+1,j} \bullet a_{j,i}$. But also since $a_{j,i}$ and $c_{i,j}$ are H applied to points within $R^n_{i,j}$, their images are contained in $H(R^n_{i,j})$, and we picked n such that $H(R^n_{i,j})$ are contained in X_1 or X_2. If this is contained in X_1, then

$$h(a_{j+1,i}) \circ h(c_{i,j}) = h(a_{j+1,i} \bullet c_{i,j}) = f([a_{j+1,i} \bullet c_{i,j}]) = f([c_{i+1,j} \bullet a_{j,i}]) = h(c_{i+1,j}) \circ h(a_{j,i}).$$

The other case is similar. Thus in any case, $h(a_{j,i}) = h(c_{i+1,j})^{-1} \circ h(a_{j+1,i}) \circ h(c_{i,j})$, since the codomain G of h is a groupoid. Then since $a_j = a_{j,n-1} \bullet \cdots \bullet a_{j,1} \bullet a_{j,0}$,

$$h(a_j) = \bigcirc_{i=0}^{n-1} h(a_{j,i}) = \bigcirc_{i=0}^{n-1} \left(h(c_{i+1,j})^{-1} \circ h(a_{j+1,i}) \circ h(c_{i,j})\right).$$

But this sum telescopes, since each term begins with $h(c_{i+1,j})^{-1}$ and ends with $h(c_{i,j})$. Thus

$$h(a_j) = h(c_{n,j})^{-1} \circ \left(\bigcirc_{i=0}^{n-1} h(a_{j+1,i})\right) \circ h(c_{0,j}) = h(c_{n,j})^{-1} \circ h(a_{j+1}) \circ h(c_{0,j}).$$

Then we see $c_{0,j}(t) = H\left(\frac{0}{n} \cdot r, t + \frac{j}{n}\right) = H\left(0, t + \frac{j}{n}\right) = p(0)$ and $c_{n,j}(t) = H(r, t + \frac{j}{n}) = p(r)$, since H is a path-homotopy. Thus by a lemma above, $h(a_j) = h(a_{j+1})$, and so by induction

$$h(p) = h(a_0) = h(a_1) = \ldots = h(a_{n-1}) = h(a_n) = h(q).$$

Thus h gives the same value when two paths are homotopy equivalent, and so we can set $\tilde{h}([p]) = h(p)$. This is clearly a functor, since h is. Then the diagram

$$\begin{tikzcd}
\pi X_0 \arrow[r, "\pi\iota_1"] \arrow[d, "\pi\iota_2"'] & \pi X_1 \arrow[d, "\pi i_1"] \arrow[ddr, bend left, "f"] & \\
\pi X_2 \arrow[r, "\pi i_2"'] \arrow[drr, bend right, "g"'] & \pi X \arrow[dr, "\tilde{h}"] & \\
 & & \mathsf{G}
\end{tikzcd}$$

commutes, because for any class of paths $[p] \in \pi X_1$,

$$\tilde{h}(\pi i_1([p])) = \tilde{h}([i_1 \circ p]) = h(i_1 \circ p) = h((\mathsf{P}i_1)(p)) = f'(p) = f([p]).$$

So $\tilde{h} \circ \pi i_1 = f$, and similarly $\tilde{h} \circ \pi i_2 = g$.
Now suppose we have another map $w : \pi X \to \mathsf{G}$ making the following commute

$$\begin{tikzcd}
\pi X_0 \arrow[r, "\pi\iota_1"] \arrow[d, "\pi\iota_2"'] & \pi X_1 \arrow[d, "\pi i_1"] \arrow[ddr, bend left, "f"] & \\
\pi X_2 \arrow[r, "\pi i_2"'] \arrow[drr, bend right, "g"'] & \pi X \arrow[dr, "w"] & \\
& & \mathsf{G}
\end{tikzcd}$$

Define $w' : \mathsf{P}X \to \mathsf{G}$ by $w' = w \circ \mathrm{proj}_X$. Then the diagram

$$\begin{tikzcd}
\mathsf{P}X_0 \arrow[r, "\mathsf{P}\iota_1"] \arrow[d, "\mathsf{P}\iota_2"'] & \mathsf{P}X_1 \arrow[d, "\mathsf{P}i_1"] \arrow[ddr, bend left, "f'"] & \\
\mathsf{P}X_2 \arrow[r, "\mathsf{P}i_2"'] \arrow[drr, bend right, "g'"'] & \mathsf{P}X \arrow[dr, "w'"] & \\
& & \mathsf{G}
\end{tikzcd}$$

commutes by Lemma 6.9, as

$$w' \circ \mathsf{P}i_1 = w \circ \mathrm{proj}_X \circ \mathsf{P}i_1 = w \circ \pi i_1 \circ \mathrm{proj}_{X_1} = f \circ \mathrm{proj}_{X_1} = f'$$

and similarly $w' \circ \mathsf{P}i_2 = g'$. Thus by uniqueness of h, we have $w \circ \mathrm{proj}_X = w' = h$. Thus for all classes $[p] \in \pi X$, $w([p]) = (w \circ \mathrm{proj}_X)(p) = h(p) = \tilde{h}(p)$. Thus $w = \tilde{h}$, showing that $\pi\mathbf{X}$ is a pushout square. $\square$

This is a truly beautiful theorem, and will be our main computational tool when dealing with fundamental groupoids. The theorem gets even better, though. The following generalization will allow us to compute the fundamental groupoid in situations like the diagram depicted at the start of this section, where we have several points of interest. We define some categorical preliminaries first.

Definition 7.1. Let C be a category and A' a subset of $\mathrm{Obj}(\mathsf{C})$. A subset A' is **representative** in C if for every object x there is some $y \in A'$ such that x and y are isomorphic.

In the case that $\mathsf{C} = \pi_A X$ for some set A of basepoints, a subset $A' \subseteq A$ is representative iff for each point $x \in A$ there is a point $y \in A'$ such that x and y are connected by a path lying in X.

Lemma 7.4. Suppose X is a space with subsets X', A and we have a subset A' of A which is representative in $\pi_A X$ such that $A' \cap X'$ is representative in $\pi_{A\cap X'}X'$. Then we have maps $r' : \pi_{A\cap X'}X' \to \pi_{A'\cap X'}X', r : \pi_A X \to \pi_{A'}X$ and a commutative diagrams

$$\begin{tikzcd}
\pi_{A'\cap X'}X' \arrow[r, "\iota'"] \arrow[d, "j"] \arrow[rr, bend left, "id"] & \pi_{A\cap X'}X' \arrow[r, "r'"] \arrow[d, "i"] & \pi_{A'\cap X'}X' \arrow[d, "j"] \\
\pi_{A'}X \arrow[r, "\iota"] \arrow[rr, bend right, "id"'] & \pi_A X \arrow[r, "r"] & \pi_{A'}X
\end{tikzcd}$$

Where $\iota' : \pi_{A'\cap X'}X' \to \pi_{A\cap X'}X'$ and $\iota : \pi_{A'\cap X'}X \to \pi_A X$ are inclusion maps, $j : \pi_{A'\cap X'}X' \to \pi_{A'}X$ is the inclusion on points and sends the equivalence class of a path $p : [0, r] \to X'$ to the equivalence class of the path $\overline{p} : [0, r] \to X$ given by $\overline{p}(t) = p(t)$, and $i : \pi_{A\cap X'}X' \to \pi_A X$ is defined similar to j.

Proof. It's immediate that the diagram

$$\begin{tikzcd}
\pi_{A'\cap X'}X' \arrow[r, "\iota'"] \arrow[d, "j"] & \pi_{A\cap X'}X' \arrow[d, "i"] \\
\pi_{A'}X \arrow[r, "\iota"] & \pi_A X
\end{tikzcd}$$

commutes, since all the maps are the inclusion on objects and i, j affect maps in the same way. We first define $r' : \pi_{A\cap X'}X' \to \pi_{A'}X'$. First, we choose for each $x \in A \cap X'$ an element $a_x \in A' \cap X'x$ and isomorphism $\varphi_x : x \to a_x$. If $x \in A' \cap X'$, define $a_x = x$ and $\varphi_x = id_x$. Otherwise, since $A' \cap X'$ is representative in $\pi_{A\cap X'}X'$ we know there exists such an a_x and φ_x, and we pick any such element and path. On objects (i.e. points) define

$r'(x) = a_x$ and for an equivalence class $p \in \mathrm{Hom}_{\pi_{A\cap X'}}(x,y)$ of paths, set $r'(p) = \varphi_y \circ p \circ \varphi_x^{-1} : a_x \to a_y$. This is a functor since $r'(id_x) = \varphi_x \circ id_x \circ \varphi_x^{-1} = \varphi_x \circ \varphi_x^{-1} = id_{a_x} = id_{r'(x)}$, and given $p : x \to y$ and $q : y \to z$,

$$r'(q \circ p) = \varphi_z \circ q \circ p \circ \varphi_x^{-1} = \varphi_z \circ q \circ \varphi_y^{-1} \circ \varphi_y \circ p \circ \varphi_x^{-1} = r'(q) \circ r'(p).$$

Also, we show $r' \circ \iota' = id$. Immediately $r'(\iota'(x)) = r'(x) = x$ for any object x of A', and for any $p : x \to y$ for $x, y \in A' \cap X'$,

$$r'(\iota'(p)) = r'(p) = \varphi_y \circ p \circ \varphi_x^{-1} = id_y \circ p \circ id_x^{-1} = p.$$

This gives us the top half of the desired diagram. We can define $r : \pi_A X \to \pi_{A'} X$ similarly. Choose for each $x \in A$ a point $b_x \in A'$ and morphism $\psi_x : x \to b_x$. If $x \in A \cap X'$, set $b_x = a_x$ as before (so in particular $b_x = x$ for $x \in A' \cap X'$) and $\psi_x = i(\varphi_x)$. Otherwise, we just pick b_x and ψ_x arbitrarily. Now set $r(x) = b_x$ and $r(p) = \psi_y \circ p \circ \psi_x^{-1}$ as before. The proof that r is a functor and that $r \circ \iota = id$ is just like it is for r'. Thus all that remains to be shown is the commutativity of

$$\begin{tikzcd}
\pi_{A\cap X'} X' \arrow[r, "r'"] \arrow[d, "i"] & \pi_{A'\cap X'} X' \arrow[d, "j"] \\
\pi_A X \arrow[r, "r"] & \pi_{A'} X
\end{tikzcd}$$

Given a point $x \in A \cap X'$, we have that $j(r'(x)) = j(a_x) = a_x = b_x = r(x) = r(i(x))$. Now suppose $x, y \in A \cap X'$ and p is a morphism $x \to y$. Then

$$r(i(p)) = \psi_y \circ i(p) \circ \psi_x^{-1} = i(\varphi_y) \circ i(p) \circ i(\varphi_x)^{-1} = i(\varphi_y \circ p \circ \varphi_x) = i(r'(p)) = j(r'(p))$$

Where we used the fact that i is the restriction of j to $\pi_{A\cap X'} X'$, as both just send equivalence classes of a path with codomain X' to the equivalence class of that path with the codomain X. This shows that the diagram of interest commutes, so we are done. $\square$

Lemma 7.5. Suppose A is a subset of X and let $A_j = A \cap X_j$ for $j = 0, 1, 2$. If A is representative in πX and each A_j is representative in πX_j, then

$$\begin{tikzcd}
\pi_{A_0} X_0 \arrow[r, "\pi\iota_1"] \arrow[d, "\pi\iota_2"'] & \pi_{A_1} X_1 \arrow[d, "\pi i_1"] \\
\pi_{A_2} X_2 \arrow[r, "\pi i_2"'] & \pi_A X
\end{tikzcd}$$

is a pushout in Grpd.

Proof. By Lemmas 4.10 and 7.3, it suffices to give commutative cubes

$$\begin{tikzcd}
\pi X_0 \arrow[rr, "\pi\iota_1"] \arrow[dr, "r_0"] \arrow[dd, "\pi\iota_2"'] & & \pi X_1 \arrow[dd, "\pi i_1"] \arrow[dr, "r_1"] & \\
& \pi_{A_0} X_0 \arrow[rr, "\pi\iota_1"] \arrow[dd, "\pi\iota_2"] & & \pi_{A_1} X_1 \arrow[dd, "\pi i_1"] \\
\pi X_2 \arrow[rr, "\pi i_2"] \arrow[dr, "r_2"] & & \pi X \arrow[dr, "r_3"] & \\
& \pi_{A_2} X_2 \arrow[rr, "\pi i_2"] & & \pi_A X
\end{tikzcd}$$

and

$$\begin{tikzcd}
\pi X_0 \arrow[rr, "\pi\iota_1"] \arrow[dd, "\pi\iota_2"'] & & \pi X_1 \arrow[dd, "\pi i_1"] & \\
& \pi_{A_0} X_0 \arrow[ul, "s_0"] \arrow[rr, "\pi\iota_1"] \arrow[dd, "\pi\iota_2"] & & \pi_{A_1} X_1 \arrow[ul, "s_1"'] \arrow[dd, "\pi i_1"] \\
\pi X_2 \arrow[rr, "\pi i_2"] & & \pi X & \\
& \pi_{A_2} X_2 \arrow[ul, "s_2"] \arrow[rr, "\pi i_2"] & & \pi_A X \arrow[ul, "s_3"']
\end{tikzcd}$$

Such that $r_j \circ s_j$ is the identity for $j \in \{0, 1, 2, 3\}$. By applying Lemma 7.4 with $X = X_1$, $X' = X_0$, $A = X_1$, and $A' = A_1$, we get the commutative diagrams

$$\begin{tikzcd}
\pi_{A_0} X_0 \arrow[r, "\iota_0"] \arrow[d, "j_1"] \arrow[rr, bend left, "id"] & \pi X_0 \arrow[r, "r_0"] \arrow[d, "i_1"] & \pi_{A_0} X' \arrow[d, "j_1"] \\
\pi_{A_1} X_1 \arrow[r, "\iota_1"] \arrow[rr, bend right, "id"'] & \pi X_1 \arrow[r, "r_1"] & \pi_{A_1} X_1
\end{tikzcd}$$

And by the same process for $X = X_2$, we obtain

$$\begin{tikzcd}
\pi_{A_0}X_0 \arrow[r, "\iota_0"] \arrow[d, "j_2"] \arrow[rr, bend left, "id"] & \pi X_0 \arrow[r, "r_0"] \arrow[d, "i_2"] & \pi_{A_0}X' \arrow[d, "j_2"] \\
\pi_{A_2}X_2 \arrow[r, "\iota_2"] \arrow[rr, bend right, "id"'] & \pi X_2 \arrow[r, "r_2"] & \pi_{A_2}X_2
\end{tikzcd}$$

Now if we define $\iota_3 : \pi_A X \to \pi X$ to be the inclusion functor, and define $s_j = \iota_j$, we have that $r_j \circ s_j$ is the identity for $j = 0, 1, 2$, and have the commutative diagrams

$$\begin{tikzcd}
\pi X_0 \arrow[rr, "\pi\iota_1"] \arrow[dr, "r_0"] \arrow[dd, "\pi\iota_2"'] & & \pi X_1 \arrow[dr, "r_1"] \arrow[dd, "\pi i_1"' near start] & \\
& \pi_{A_0}X_0 \arrow[rr, "\pi\iota_1" near start] \arrow[dd, "\pi\iota_2"' near start] & & \pi_{A_1}X_1 \arrow[dd, "\pi i_1"] \\
\pi X_2 \arrow[rr, "\pi i_2" near end] \arrow[dr, "r_2"] & & \pi X & \\
& \pi_{A_2}X_2 \arrow[rr, "\pi i_2"] & & \pi_A X
\end{tikzcd}$$

and

$$\begin{tikzcd}
\pi X_0 \arrow[rr, "\pi\iota_1"] \arrow[dd, "\pi\iota_2"'] & & \pi X_1 \arrow[dd, "\pi i_1"' near start] & \\
& \pi_{A_0}X_0 \arrow[ul, "s_0"] \arrow[rr, "\pi\iota_1" near start] \arrow[dd, "\pi\iota_2"' near start] & & \pi_{A_1}X_1 \arrow[ul, "s_1"'] \arrow[dd, "\pi i_1"] \\
\pi X_2 \arrow[rr, "\pi i_2" near end] & & \pi X & \\
& \pi_{A_2}X_2 \arrow[ul, "s_2"] \arrow[rr, "\pi i_2"] & & \pi_A X \arrow[ul, "s_3"']
\end{tikzcd}$$

By the process in Lemma 7.4 (adapted so that our choices for a_x on both agree with those in both r_1 and r_2) we obtain a map $r_3 : \pi_A X \to \pi X$ which makes the above cube commute. Thus we are done $\square$

This lemma allows to restrict our attention to only the basepoints we care about, and when combined with the next lemma gives us a tool for computing the fundamental group in terms of the fundamental groupoid with respect to a smaller set of basepoints.

Lemma 7.6. Suppose as in Lemma 7.4 that we have a space X, subsets X' and $A' \subseteq A$ where A' is representative in $\pi_A X$ and $A' \cap X'$ is representative in $\pi_{A\cap X'} X'$. Then if $A'' = (A' \cap X') \cup (A \setminus X')$, we have the diagram

$$\begin{tikzcd}
\pi_{A'\cap X'} X' \arrow[r, "\iota'"] \arrow[d, "j"] \arrow[rr, bend left, "id"] & \pi_{A\cap X'} X' \arrow[r, "r'"] \arrow[d, "i"] & \pi_{A'\cap X'} X' \arrow[d, "j"] \\
\pi_{A''} X \arrow[r, "\iota"] \arrow[rr, bend right, "id"'] & \pi_A X \arrow[r, "r"] & \pi_{A''} X
\end{tikzcd}$$

and the inner right square

$$\begin{tikzcd}
\pi_{A\cap X'} X' \arrow[r, "r'"] \arrow[d, "i"] & \pi_{A'\cap X'} X' \arrow[d, "j"] \\
\pi_A X \arrow[r, "r"] & \pi_{A''} X
\end{tikzcd}$$

is a pushout square.

Proof. We obtain the commutative diagram

$$\begin{tikzcd}
\pi_{A'\cap X'} X' \arrow[r, "\iota'"] \arrow[d, "j"] \arrow[rr, bend left, "id"] & \pi_{A\cap X'} X' \arrow[r, "r'"] \arrow[d, "i"] & \pi_{A'\cap X'} X' \arrow[d, "j"] \\
\pi_{A''} X \arrow[r, "\iota"] \arrow[rr, bend right, "id"'] & \pi_A X \arrow[r, "r"] & \pi_{A''} X
\end{tikzcd}$$

through the same process as in Lemma 7.4]. Now given a commutative square of groupoids

$$\begin{tikzcd}
\pi_{A\cap X'}X' \arrow[r, "r'"] \arrow[d, "i"] & \pi_{A'\cap X'}X' \arrow[d, "f"] \\
\pi_A X \arrow[r, "g"] & \mathsf{G}
\end{tikzcd}$$

we can define $h : \pi_{A'}X \to \mathsf{G}$ by $h = g \circ \iota$. Then the diagram

$$\begin{tikzcd}
\pi_{A\cap X'}X' \arrow[r, "r'"] \arrow[d, "i"] & \pi_{A'\cap X'}X' \arrow[d, "j"] \arrow[ddr, bend left, "f"] & \\
\pi_A X \arrow[r, "r"] \arrow[drr, bend right, "g"] & \pi_{A''}X \arrow[dr, "h"] & \\
 & & \mathsf{G}
\end{tikzcd}$$

commutes, since

$$h \circ j = g \circ \iota \circ j = g \circ i \circ \iota' = f \circ r' \circ \iota' = f$$

and $h \circ r = g \circ \iota \circ r$. We show $g \circ \iota \circ r = g$ as follows. For any $x \in A$, either $x \in X'$ or $x \notin X'$. If $x \notin X'$, then $r(x) = x$ since $A \setminus X' \subseteq A''$, so

$$g(\iota(r(x))) = g(r(x)) = g(x).$$

Otherwise $x \in A \cap X'$ and so $x = i(x)$, thus

$$g \circ \iota \circ \underline{r \circ i} = g \circ \underline{\iota \circ j} \circ r' = \underline{g \circ i} \circ \iota' \circ r' = f \circ \underline{r' \circ \iota'} \circ r' = \underline{f \circ r'} = g \circ i.$$

Where we have used underlines to indicate which terms change during the next equality. Thus if $x \in A \cap X'$, $g(\iota(r(x))) = g(\iota(r(i(x)))) = g(i(x)) = g(x)$. Now for x in $\pi_A X$,

$$(h \circ r)(x) = ((g \circ \iota) \circ r)(x) = g(x).$$

So to show that our diagram of interest is a pushout square, it suffices to show that $(h \circ r)(p) = g(p)$ for any morphism p in $\pi_A X$. To do so, we need to dig into the construction of r. In particular, for any $x \in \pi_{A\cap X'}X'$ we have a map $\varphi_x : x \to r'(x)$ and for each $x \in \pi_{A\cap X}X$ a map $\psi_x : x \to r(x)$ such that

$\psi_x = i(\varphi_x)$ when $x \in A \cap X'$ and $\psi_x = id_x$ when $x \in A''$. Also, for a map $p : x \to y$, we defined $r(p) = \psi_y \circ p \circ \varphi_x^{-1}$. We show $g(\psi_x) = id_{g(x)}$ for any x. As before, any x is either in A'' or in $A \cap X'$. In the first case, $\psi_x = id_x$, so $g(\psi_x) = id_{g(x)}$ by the fact that g is a functor. Now if $x \in A \cap X'$, $\psi_x = i(\varphi_x)$, and thus

$$g(\psi_x) = g(i(\varphi_x)) = f(r'(\varphi_x)) = f(\varphi_{r'(x)} \circ \varphi_x \circ \varphi_x^{-1}) = f(id_{r'(x)}) = id_{f(r'(x))} = id_{g(i(x))} = id_{g(x)}.$$

Thus $g(\psi_x) = id_{g(x)}$ for all x. Then given a morphism $p : x \to y$ in $\pi_A X$,

$$h(r(p)) = g(\iota(r(p))) = g(r(p)) = g(\psi_y \circ p \circ \psi_x^{-1}) = g(\psi_y) \circ g(p) \circ g(\psi_x)^{-1} = id_{g(y)} \circ g(p) \circ id_{g(x)}^{-1} = g(p).$$

This shows $h \circ r = g$ on maps as well as objects, and so the diagram commutes, showing it is a pushout. $\square$

This gives us the tools needed to prove our relative Seifert-van Kampen Theorem, which shall compute the fundamental group of the circle.

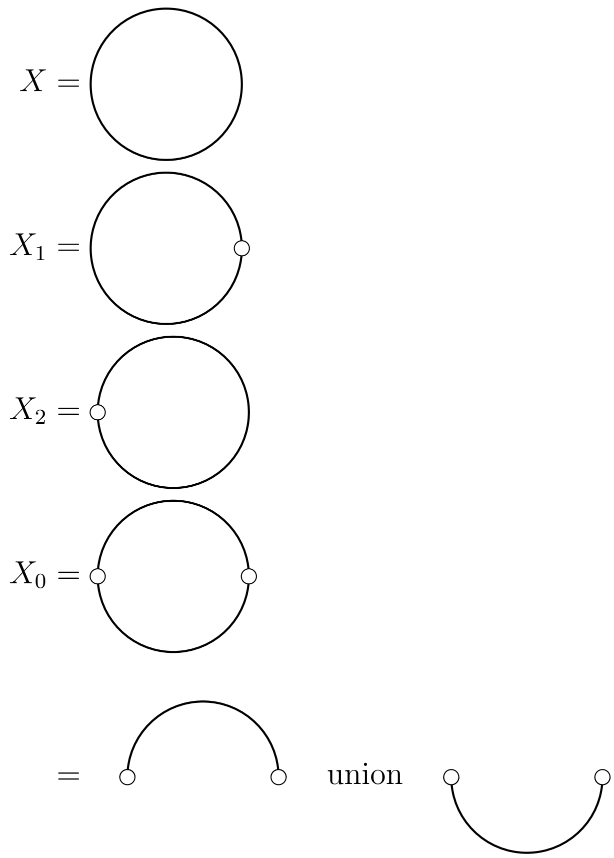
Theorem 7.7 (The Relative Seifert-van Kampen Theorem)**.** Suppose we have A as in Lemma 7.5 and a subset A' of A such that $A' \cap X'$ is representative in $\pi_{A \cap X'} X'$. Then if $A'' = (A' \cap X') \cup (A \setminus X')$, the outer rectangle of

$$\begin{array}{ccccc}
\pi_{A_0} X_0 & \xrightarrow{\pi\iota_1} & \pi_{A_1} X_1 & \xrightarrow{r'} & \pi_{A'} X_1 \\
\downarrow{\scriptstyle \pi\iota_2} & & \downarrow{\scriptstyle \pi i_1} & & \downarrow{\scriptstyle j} \\
\pi_{A_2} X_2 & \xrightarrow{\pi i_2} & \pi_A X & \xrightarrow{r} & \pi_{A''} X
\end{array}$$

is a pushout diagram.

Proof. Both inner squares are pushouts by Lemmas 7.5 and 7.6, and by Lemma 4.11 this implies the outer rectangle is too. $\square$

We can now determine the fundamental group of the circle. Define



More formally, let $X = \mathbb{S}^1$, $X_1 = \mathbb{S}^1 \setminus \{(1,0)\}$, $X_2 = \mathbb{S}^1 \setminus \{(-1,0)\}$, and $X_0 = X_1 \cap X_2$. Define our set of basepoints by $A = \{(0,1),(0,-1)\}$ (i.e. top and bottom of the circle) and let $A' = \{(1,0)\}$ be just the top. Then $A \subseteq X_0$ so $A'' = A'$, giving us the pushout diagram

$$\begin{tikzcd}
\pi_A X_0 \arrow[r, "\pi\iota_1"] \arrow[d, "\pi\iota_2"] & \pi_A X_1 \arrow[r, "r'"] \arrow[d, "\pi i_1"] & \pi_{A'} X_1 \arrow[d, "j"] \\
\pi_A X_2 \arrow[r, "\pi i_2"] & \pi_A X \arrow[r, "r"] & \pi_{A'} X
\end{tikzcd}$$

Also note that we can parameterize X_1 and X_2 by

$$\begin{aligned} p_1 &: (0, 2\pi) \to X_1 \\ p_1(\theta) &= (\cos(\theta), \sin(\theta)) \\ p_2 &: (-\pi, \pi) \to X_2 \\ p_2(\theta) &= (\cos(\theta), \sin(\theta)). \end{aligned}$$

These functions are clearly continuous, and on the circle minus a point they have continuous inverses (in the language of complex analysis, these inverses are branches of the argument function). Similarly, X_0 consists of two copies disjoint of spaces homeomorphic to an open interval, i.e. the sets of those points on the circle with angle in $(-\pi/2, \pi/2)$ or in $(\pi/2, 3\pi/2)$. Thus X_1 and X_2 are homeomorphic to convex sets, on which any path between two points is equivalent to any other such path, so also any two paths with the same source and target are equivalent in X_1 and X_2. The same is true for X_0, since any path will be between two objects in a subset homeomorphic to a convex set. Thus $\pi_A X_2$ is the groupoid with two objects and a single invertible map between them, $\pi_{A'} X_1$ is the single object groupoid with no nonidentity maps, and $\pi_A X_0$ is the two object groupoid with no intermediate maps. Let $\mathbf{I}$ be the groupoid $id_{\mathbf{0}} \circlearrowright \mathbf{0} \underset{\kappa^{-1}}{\overset{\kappa}{\rightleftarrows}} \mathbf{1} \circlearrowleft id_{\mathbf{1}}$, let $\mathbf{Z}$ be the groupoid $* \circlearrowleft id_*$, and let $\mathbf{D}$ be the groupoid $id_{\mathbf{0}} \circlearrowright \mathbf{0} \quad \mathbf{1} \circlearrowleft id_{\mathbf{1}}$. Then as we argued above, $\pi_A X_0 = \mathbf{D}$, $\pi_A X_2 = \mathbf{I}$, and $\pi_{A''} X_1 = \mathbf{Z}$, so by Theorem 7.7, we get a pushout square

$$\begin{tikzcd} \mathbf{D} \arrow[r, "\alpha"] \arrow[d, "\beta"] & \mathbf{Z} \arrow[d] \\ \mathbf{I} \arrow[r] & \pi_{A'}\mathbb{S}^1 \end{tikzcd}$$

Note that there is only one map $\mathbf{D} \to \mathbf{Z}$, namely the functor which squashes two objects into one, and also only one $\mathbf{D} \to \mathbf{I}$, the inclusion, so we just label them α and β. We determine $\pi_A \mathbb{S}^1$ by showing that a certain, more familiar group is also a pushout of $\mathbf{Z} \to \mathbf{D} \leftarrow \mathbf{I}$, and concluding that these groups are isomorphic by uniqueness of pushouts.

Theorem 7.8. Let $\mathbb{Z}$ denote the group of integers under addition, with the identity 0. Then

$$\pi(\mathbb{S}^1, (1, 0)) \cong \mathbb{Z}.$$

Proof. As in our comments above, it suffices to give a pushout square

$$\begin{array}{ccc} \mathbf{D} & \xrightarrow{\alpha} & \mathbf{Z} \\ \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle f} \\ \mathbf{I} & \xrightarrow{g} & \mathbb{Z} \end{array}$$

There is only one such f, namely the one which sends $*$ to the unique object in the groupoid $\mathbb{Z}$ and the identity map to 0. Let g be the map which squashes $\mathbf{0}$ and $\mathbf{1}$ into the only object of $\mathbb{Z}$, sends the identity maps to 0, and sends κ to 1 (so it also sends κ^{-1} to -1). Immediately the square commutes on the level of objects, since both compositions $f \circ \alpha$ and $g \circ \beta$ squash everything into the only object of $\mathbb{Z}$. Since there are no (nontrivial) maps in $\mathbf{D}$, it also commutes on the level of maps. Now suppose we have a pushout candidate

$$\begin{array}{ccc} \mathbf{D} & \xrightarrow{\alpha} & \mathbf{Z} \\ \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle p} \\ \mathbf{I} & \xrightarrow{q} & \mathsf{G} \end{array}$$

Note that since $\mathbf{Z}$ only has a single object, p must pick out some object K of G, and by commutativity of the diagram both $p \circ \alpha$ and $q \circ \beta$ must squash both objects in $\mathbf{D}$ into K. In particular, $q(\mathbf{0}) = K = q(\mathbf{1})$, so $q(\kappa) \in \mathrm{Hom}_{\mathsf{G}}(K, K)$. Define $h : \mathbb{Z} \to \mathsf{G}$ to send the object of $\mathbb{Z}$ to K and let $h(1) = q(\kappa)$. This defines a functor on all maps of $\mathbb{Z}$ since we must have $h(n) = h(1 + 1 + \cdots + 1) = h(1) \circ h(1) \circ \cdots \circ h(1)$, and $h(-n) = h(n)^{-1}$. We show that $h \circ f = p$ and $h \circ g = q$, and conclude that the integers are a pushout of $\mathbf{I} \leftarrow \mathbf{D} \to \mathbf{Z}$. Clearly $h \circ f = p$, since both send $*$ to K and id_* to id_K. Then we constructed h so that $h(g(\kappa)) = h(1) = q(\kappa)$, so $h \circ g = q$ as well. $\square$

Now since all points x in $\mathbb{S}^1$ can be connected to $(1, 0)$ by a path (just go around the circle), all objects x of $\pi\mathbb{S}^1$ are isomorphic to $(0, 1)$, so by Lemma 5.3, $\pi(\mathbb{S}^1, x) \cong \pi(\mathbb{S}^1, (1, 0)) \cong \mathbb{Z}$ for any x. Having shown that the circle has a hole in it, we conclude the section by proving the circle and the interval are distinct.

Corollary 7.9. The circle and the interval are not homeomorphic.

Proof. Suppose we had a homeomorphism $h : \mathbb{S}^1 \to \mathbb{I}$. Then since $\pi : \mathsf{Top} \to \mathsf{Grpd}$ is a functor, the map $\pi h : \pi\mathbb{S}^1 \to \pi\mathbb{I}$ would be an isomorphism.

Explicitly, $(\pi h) \circ (\pi h^{-1}) = \pi(h \circ h^{-1}) = \pi id_{\mathbb{I}} = id_{\pi\mathbb{I}}$ by the fact that functors preserve composition, and the same argument shows $(\pi h^{-1}) \circ (\pi h) = id_{\pi\mathbb{S}^1}$. Now let $x = h((1,0))$. By Lemma 5.2, the isomorphism πh restricts to an isomorphism $h' : \pi(\mathbb{S}^1, (1,0)) \to \pi(\mathbb{I}, x)$. But by Lemma 6.14, $\pi(\mathbb{I}, x)$ has only one element, and by Theorem 7.8 $\pi(\mathbb{S}^1, (1,0))$ has infinitely many, a contradiction. □

8 Interlude: Homotopy Equivalence and the Sphere

The previous section accomplished the core goal of this paper: proving the circle has a hole in it. The Seifert-van Kampen Theorem is far more powerful than this, however. In the last section of the paper, we use it to prove the result that every closed curve in the plane has an inside and an outside. While the theorem is simple to state, it eluded proof for hundreds of years. Before we do this, however, we need to do a little geometry. The proof of the Jordan Curve Theorem requires computing the fundmental groupoids of certain spaces related to the sphere, and this penultimate secton accomplishes that.

Definition 8.1. The sphere is the space

$$\mathbb{S}^2 = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}.$$

The north pole is the point $N = (0,0,1)$ and the south pole $S = (0,0,-1)$.

Lemma 8.1. For any point $a \in \mathbb{S}^2$, there's a homeomorphism of the sphere sending a to N.

Proof. Rotate the sphere along the great circle on which a lies. This is an invertible linear transformation (by elementary linear algebra) so it's a homeomorphism (as its inverse is also a linear transformation and all linear transformations are continuous). □

Lemma 8.2. There is a homeomorphism $h : \mathbb{S}^2 \setminus \{N\} \to \mathbb{R}^2$ between "punctured sphere" and the plane. This homeomorphism satisfies $h(S) = (0,0)$ and $\lim_{x \to N, x \in \mathbb{S}^2} |h(x)| = \infty$.

Proof. Define $h(x,y,z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$. Since $1 - z = 0$ iff (x,y,z) is the north pole, this map is well defined, and since it's a rational function it's continuous. It has the inverse

$$h^{-1}(u,v) = \left(u\left(1 - \frac{u^2+v^2-1}{u^2+v^2+1}\right), v\left(1 - \frac{u^2+v^2-1}{u^2+v^2+1}\right), \frac{u^2+v^2-1}{u^2+v^2+1}\right).$$

By the same argument as above, h^{-1} is continuous. Then also if $(u,v) = h(x,y,z)$,

$$\frac{u^2+v^2-1}{u^2+v^2+1} = \frac{(x/(1-z))^2 + (y/(1-z))^2 - 1}{(x/(1-z))^2 + (y/(1-z))^2 + 1} = \frac{x^2+y^2-(1-z)^2}{x^2+y^2+(1-z)^2} = \frac{1-z^2-(1-z)^2}{1-z^2+(1-z)^2} = z.$$

Thus $h^{-1}(h(x,y,z)) = \left(\frac{x}{1-z} \cdot (1-z), \frac{y}{1-z} \cdot (1-z), z\right) = (x,y,z)$. Then also

$$\begin{aligned} h(h^{-1}(u,v)) &= h\left(u\left(1 - \frac{u^2+v^2-1}{u^2+v^2+1}\right), v\left(1 - \frac{u^2+v^2-1}{u^2+v^2+1}\right), \frac{u^2+v^2-1}{u^2+v^2+1}\right) \\ &= \left(u\left(1 - \frac{u^2+v^2-1}{u^2+v^2+1}\right) \div \left(1 - \frac{u^2+v^2-1}{u^2+v^2+1}\right), v\left(1 - \frac{u^2+v^2-1}{u^2+v^2+1}\right) \div \left(1 - \frac{u^2+v^2}{u^2+v^2}\right.\right. \\ &= (u,v). \end{aligned}$$

And so h^{-1} is in fact an inverse of h, which means h is a homeomorphism. Then immediately $h(0,0,-1) = \left(\frac{0}{2}, \frac{0}{2}\right) = 0$. To see the claim about the limit, note that for (x,y,z) on the sphere, $|h(x,y,z)|^2 = \frac{x^2+y^2}{(1-z)^2} = \frac{1-z^2}{(1-z)^2} = \frac{1+z}{1-z}$ which goes to ∞ as z goes to 1. $\square$

These lemmas above immediately imply that the fundamental group $\pi(\mathbb{S}^2 \setminus \{a\}, b)$ is trivial for any two distinct points $a, b \in \mathbb{S}^2$. Let R be the homeomorphism in Lemma 8.1 sending a to N restricted to a homeomorphism between $\mathbb{S}^2 \setminus \{a\}$ and $\mathbb{S}^2 \setminus \{N\}$. Further, let h be the homeomorphism in Lemma 8.2, so $h \circ R$ is a homeomorphism between $\mathbb{S}^2 \setminus \{a\}$ and $\mathbb{R}^2$. Then $\pi(h \circ R)$ is an isomorphism of groupoids between $\pi(\mathbb{S}^2 \setminus \{a\})$ and $\pi\mathbb{R}^2$. By Lemma 5.2, this restricts to an isomorphism of groups $\pi(\mathbb{S}^2 \setminus \{a\}, b) \cong \pi(\mathbb{R}^2, h(R(b)))$, the latter of which is trivial since $\mathbb{R}^2$ is convex. This also implies that the fundamental group of $\mathbb{S}^2$ is trivial, by applying the Seifert van-Kampen Theorem to the set $X_1 = \mathbb{S}^2 \setminus \{N\}$ and $X_2 = \mathbb{S}^2 \setminus \{S\}$ and invoking Lemma 4.9. More complicated is determining the fundamental group of $\mathbb{S}^2 \setminus \{a,b\}$ at a point $x \in \mathbb{S}^2 \setminus \{a,b\}$. First define $T : \mathbb{S}^2 \setminus \{a,b\} \to \mathbb{R}^2 \setminus \{0\}$ by $T(x) = h^{-1}(h(R(x)) - h(R(b)))$. This is a composition of homeomorphisms (rotate, stereographically project, and translate), so it is a homeomorphism as well. Then by Lemma 5.2, $\pi(\mathbb{S}^2 \setminus \{a,b\}, x) \cong \pi(\mathbb{R}^2 \setminus \{0\}, T(x))$.

Thus it suffices to determine the fundamental group of $\mathbb{R}^2 \setminus \{N, S\}$ at an arbitrary point. Even better, it doesn't matter what point we choose to compute the fundamental group at, since all points in $\mathbb{R}^2 \setminus \{0\}$ can be connected by paths (take the straight line and make a semicircle around the origin if needed). Choose x_0 to be a point on $\mathbb{S}^1$. Intuitively $\pi(\mathbb{R}^2 \setminus \{0\}, x_0)$ should be the integers, since the only hole in our space is like the circle. Can we use our previous result that $\pi(\mathbb{S}^1, x_0) \cong \mathbb{Z}$ to deduce this? The Seifert-van Kampen theorem isn't useful for this, since any two open sets which cover $\mathbb{R}^2 \setminus \{0\}$ will be too fat to look like circles. Instead, we *retract* the punctured plane onto the unit circle by sending each point x to the point $x/|x|$ on the unit circle.

Definition 8.2. If A is a subspace of B, then a **retraction** of B onto A is a continuous map $r : B \to A$ such that $r(a) = a$ for all $a \in A$. If $\iota : A \to B$ is the inclusion map, we can state this condition as $r \circ \iota = id_A$[†]. cond

Not all retractions preserve the fundamental group(oid), however. Consider the retraction r of the circle $\mathbb{S}^1$ onto its upper hemisphere $H = \{(x, y) \in \mathbb{S}^1 : y \geq 0\}$ given by $r(x, y) = (x, |y|)$. Since the absolute value is continuous, this is continuous, and $r(x, y) = (x, |y|) = (x, y)$ when $y \geq 0$. Thus r is a retraction, but H is homeomorphic to the interval, so its fundamental group is trivial everywhere (whereas the fundamental group of the circle is $\mathbb{Z}$). The condition on r we're looking for is most easily stated in terms of *homotopy equivalence*, a notion of "equivalence" of topological spaces even weaker than homeomorphism.

Definition 8.3. Suppose (X, x_0) and (Y, y_0) are pointed spaces, i.e. X and Y are spaces and x_0, y_0 are points in them. Let $f, g : X \to Y$ be continuous functions which both send x_0 to y_0. A **homotopy rel basepoints** between f and g is a homotopy $H : X \times I \to Y$ between f and g such that $H(x_0, t) = y_0$ for all t. If such a homotopy exists, we say f and g are **homotopic rel basepoints**.

Definition 8.4. Let (X, x_0) and (Y, y_0) be pointed spaces and suppose we have maps $f : X \to Y$, $g : Y \to X$ such that $f(x_0) = y_0$ and $g(y_0) = f(x_0)$. We say f and g are **homotopy inverses** if $f \circ g$ is homotopic to id_Y rel basepoints and $g \circ f$ is homotopic to id_X rel basepoints. In this case, we say

[†]This equivalent formulation works in any category. We saw retractions of groupoids in Lemma 7.6 and of commutative diagrams in Lemma 4.10.

that f is a **homotopy equivalence** (and the same for g) and that (X, x_0) and (Y, y_0) are **homotopy equivalent**. Note that every homeomorphism is a homotopy equivalence.

Homotopy rel basepoints is an incredibly useful idea, since it turns out that you only need to know a map up to homotopy rel basepoints to determine the map it induces on the fundamental groups. That is to say, if f, g are maps (X_0, x_0) to (Y_0, y_0) which are homotopic rel basepoints, then f and g induce the same homomorphism $\pi(X, x_0) \to \pi(Y, y_0)$.

Lemma 8.3. Suppose (X, x_0) and (Y, y_0) are pointed spaces and we have maps $f, g : X \to Y$ which send x_0 to y_0 that are homotopic rel basepoints. Then the homomorphisms $\pi f, \pi g : \pi(X, x_0) \to \pi(Y, y_0)$ are equal.

Proof. It suffices to show that for any loop ℓ based at x_0, the loops $f \circ \ell$ and $g \circ \ell$ (based at y_0) are path-homotopic, since then $(\pi f)([\ell]) = [f \circ \ell] = [g \circ \ell] = (\pi g)([\ell])$ for all loops ℓ. W.l.o.g. we can assume the domain of ℓ is $\mathbb{I}$ by Lemma 6.12. Let $H : X \times \mathbb{I} \to Y$ be a homotopy rel basepoints between f and g. Then define $G : \mathbb{I} \times \mathbb{I} \to Y$ by $G(t, s) = H(\ell(t), s)$. Clearly G is continuous, and $G(0, s) = H(\ell(0), s) = H(x_0, s) = y_0$ since H is a homotopy rel basepoints (similarly $G(1, s) = y_0$). Then $G(t, 0) = H(\ell(t), 0) = f(\ell(t))$ and $G(t, 1) = H(\ell(t), 1) = g(\ell(t))$ since H is a homotopy between f and g. Thus G is a path-homotopy between $f \circ \ell$ and $g \circ \ell$, as desired. □

This is espescially powerful when applied to homotopy equivalences. Given maps k, h which are inverses only up to homotopy, the induced homomorphisms $\pi(k)$ and $\pi(h)$ are genuine inverses, and so homotopy equivalences turn into isomorphisms.

Lemma 8.4. Suppose (X, x_0) and (Y, y_0) are homotopy equivalent. Then $\pi(X, x_0)$ and $\pi(Y, y_0)$ are isomorphic.

Proof. By assumption, we have homotopy inverses $f : X \to Y$ and $g : Y \to X$. Then $f \circ g$ is homotopic rel basepoints to id_Y, so $\pi(f \circ g) = \pi id_Y$. But since π is a functor, this means $\pi f \circ \pi g = \pi(f \circ g) = \pi id_Y = id_{\pi(Y, y_0)}$. Similarly, $\pi g \circ \pi f = id_{\pi(X, x_0)}$. Thus πf is an invertible group homomorphism $\pi(X, x_0) \to \pi(Y, y_0)$, so $\pi(X, x_0)$ is isomorphic to $\pi(Y, y_0)$. □

Now that we've learned the basics of homotopy equivalence, we can return to the problem of retractions. Let x_0 be a point in A. When does a retraction

$r : B \to A$ induce an isomorphism between $\pi(A, x_0)$ and $\pi(B, x_0)$? Well by the above lemma it's certainly sufficient to ask that r is a homotopy equivalence between (B, x_0) and (A, x_0). But we already know $r \circ \iota = id_A$, where $\iota : A \to B$ is the inclusion, so we automaticlly get a homotopy rel basepoints between $r \circ \iota$ and id_A. Thus a "nice" class of retractions are those for which $\iota \circ r$ is homotopic rel basepoints to id_B.

Definition 8.5. Let A be a subspace of B and suppose we have a retraction $r : B \to A$. Let $\iota : A \to B$ be the inclusion and pick some point $x_0 \in A$. We call r a **deformation retraction** if $\iota \circ r : (B, x_0) \to (B, x_0)$ is basepoint-homotopic to id_B. As we argued above, if such a deformation retraction exists then (A, x_0) and (B, x_0) are homotopy equivalent, so $\pi(B, x_0) \cong \pi(A, x_0)$.

We now return to the retraction described on page 51. Let $\iota : \mathbb{S}^1 \to \mathbb{R}^2 \setminus \{0\}$ be the inclusion map. Now define $r : \mathbb{R}^2 \setminus \{0\} \to \mathbb{S}^1$ by $r(x) = x/|x|$. We have that $r(x) = x/|x| = x$ when $|x| = 1$, i.e. when x is on the unit circle, so $r \circ \iota = id_{\mathbb{S}^1}$. Thus if we can show r is a deformation retraction, we may conclude $\pi(\mathbb{R}^2 \setminus \{0\}, x_0) \cong \pi(\mathbb{S}^1, x_0) \cong \mathbb{Z}$.

Lemma 8.5. The fundamental group of $\mathbb{S}^2 \setminus \{a, b\}$ at any point is $\mathbb{Z}$.

Proof. As described above, it suffices to prove $r(x) = x/|x|$ is a deformation retraction $\mathbb{R}^2 \setminus \{0\} \to \mathbb{S}^1$. Define $H : (\mathbb{R}^2 \setminus \{0\}) \times \mathbb{I} \to \mathbb{R}^2 \setminus \{0\}$ by

$$H(x, s) = \begin{cases} x & \text{if } s \le |x| \text{ and } s \le \frac{1}{|x|} \\ \frac{s}{|x|}x & \text{if } s \ge |x| \text{ or } s \ge \frac{1}{|x|} \end{cases}.$$

This is continuous by the Gluing Lemma. Essentially at time s we have a retration $x \mapsto H(x, s)$ of $\mathbb{R}^2 \setminus \{0\}$ onto the closed annulus of inner radius s and outer radius $1/s$. We check that it is a homotopy between $\iota \circ r$ and $id_{\mathbb{R}^2 \setminus \{0\}}$ rel basepoints. First if $s = 0$, then $s < |x|$ and $s < \frac{1}{|x|}$ for any x, and so $H(x, 0) = x$ for all x. If $s = 1$, then $s \ge |x|$ or $s \ge \frac{1}{|x|}$ for any x, and so $H(x, 1) = \frac{x}{|x|}$. Thus H is a homotopy between $\iota \circ r$ and $id_{\mathbb{S}^2 \setminus \{N,S\}}$. It is basepoint preserving since our basepoint is on the unit circle, and $s \le 1 = |x_0| = \frac{1}{|x_0|}$ for all s. Thus

$$\pi(\mathbb{R}^2 \setminus \{0\}, x_0) \cong \pi(\mathbb{S}^1, x_0) \cong \mathbb{Z}.$$

$\square$

We close the section with one last bit of geometry of the sphere. It's easy to see that any two point on the sphere can be connected to one another by a path, just by going along great circles (or stereographically projecting). But in the next section we'll need this to be true even if we zoom into the sphere, i.e. that this property holds locally.

Lemma 8.6. Let M be a neighborhood of a point x on the sphere $\mathbb{S}^2$. Then M contains a neighborhood M' of x such that all points of M' can be connected to x by a path.

Proof. By definition of the subspace topology, M must be the intersection of K and $\mathbb{S}^2$ for some neighborhood K of x in $\mathbb{R}^3$. Then for some $\varepsilon > 0$, K contains all point $y \in \mathbb{R}^3$ such that $|x - y| < \varepsilon$, so the neighborhood M of x on $\mathbb{S}^2$ contains the set $B = \{y \in \mathbb{S}^2 : |x - y| < \varepsilon\}$. If $x \neq N$, then w.l.o.g. we can assume $N \notin B$, possibly by making ε smaller. In this case, let h be the stereographic homeomorphism from Lemma 8.2. If $x = N$, do the same thing but project relative to S. In either case, $h(B)$ is a neighborhood of $h(x)$ in the plane, and so for some $\delta > 0$, $h(B)$ contains some disk $D = \{u \in \mathbb{R}^2 : |h(x) - u| < \delta\}$. Then D contains the straight line from u to $h(x)$ for any $u \in D$, so for any $y \in h^{-1}(D)$, there is a path p_y in D between $h(y)$ and $h(x)$, and so a path $h^{-1} \circ p_y$ between y and x in $h^{-1}(D)$. Since D is a neighborhood of $h^{-1}(x)$ and h is a homeomorphism, $h^{-1}(D)$ is a neighborhood of x, and since $D \subseteq h(B) \subseteq h(M)$, also $M' = h^{-1}(D) \subseteq M$, as desired. $\square$

The material we'll use from this section in the proof of the Jordan Curve Theorem are the maps constructed in Lemmas 8.1 and 8.2 and the statements of Lemmas 8.5 and 8.6.

9 The Jordan Curve Theorem

We close this paper with another application of the Seifert-van Kampen theorem. In basic vector calculus, a student learns about Green's theorem, which relates the integral along a simple closed curve to the integral over the interior region of that curve. But what does it mean for a curve to bound the interior of curve? Like the question of whether the circle has a hole in it, this seems obvious. Of course a closed curve bounds an region! However, the theorem is much harder to prove than it looks (and nontrivial to state precisely). The reader should take a moment to try and fail to prove the theorem. We begin by discussing how we can defining two partitions of a space into large connected subsets, and one way of doing something analogous for a groupoid.

Definition 9.1. Let X be a space. A subset C of X is a **connected component** of X if C is connected and no sets properly containing C are connected. In symbols, C is a connected component if for whenever C' is connected and $C \subseteq C' \subseteq X$ implies $C = C'$.

Definition 9.2. Call a space X path connected if for any pair of points $x, y \in X$, there is a path in X from x to y.

Definition 9.3. Let X be a space. Analogous to definition 9.1, the **path components** of X are the maximal path connected subsets of X.

Note that the union of the path components of X is all of X. To see this, let x be any point of X and define $P = \{y \in X : x \text{ can be connected to } y \text{ by a path }\}$. Then P is path connected, since given $y, z \in P$ we have a path p connecting x to y and a path q connecting x to z, and $q \bullet p^{-1}$ then connects y to z. P is maximal because if $P \subseteq P'$ and P' is path connected, every point in P' can be connected to x by a path, and so $P' \subseteq P$. Thus every point is contained in some path component, so $\bigcup_{P \text{ a path component}} P = X$.

Definition 9.4. Let G be a space. For any object x in G, the **component** of G containing x is the set of all objects which x is isomomorphic to. In the case that G is a fundamental groupoid, this component is the path component of x.

The Jordan Curve Theorem states that for any subset C of $\mathbb{R}^2$ which is homeomorphic to the circle (called a simple closed curve), $\mathbb{R}^2 \setminus C$ has two connected components. Further, if we let $\partial S = \overline{S} \setminus S$ be the boundary of a set S, then the boundary of both components is C. We spend the majority of this section proving that the complement of any simple closed curve on $\mathbb{S}^2$ (the unit sphere in $\mathbb{R}^3$) has two path components, each with that curve as the boundary. Then to close out the paper we reduce the Jordan Curve Theorem to this result. First, we state but do not prove a technical lemma about groupoids. For a proof, the reader can see Theorem 1.1 of [2].

Lemma 9.1. Suppose the groupoids $\mathsf{A}, \mathsf{B}, \mathsf{C}, \mathsf{G}$ have the same set J of objects and we have a pushout

$$\begin{tikzcd} \mathsf{C} \arrow[r, "i"] \arrow[d, "j"] & \mathsf{A} \arrow[d, "u"] \\ \mathsf{B} \arrow[r, "v"] & \mathsf{G} \end{tikzcd}$$

where i, j are the identity on objects. Suppose C has no isomorphisms between distinct objects, and that all objects of G are isomorphic. Let $n_{\mathsf{A}}, n_{\mathsf{B}}, n_{\mathsf{C}}$ count the number of components in the corresponding groupoids. Then for any $p \in J$, there is a surjection from $\mathrm{Aut}_{\mathsf{G}}(p)$ to F, where F is a free group on $n_{\mathsf{C}} - n_{\mathsf{A}} - n_{\mathsf{B}} + 1$ generators, assuming n_A and n_B are both finite. If there are distinct element a, b of J such that $\mathrm{Hom}_{\mathsf{A}}(a, b)$ and $\mathrm{Hom}_{\mathsf{B}}(a, b)$ are both nonempty, then F has at least one generator.

The reader may consult Section 2.5 of [3] for a definition of the free group. The only results that we need about free groups are that the free group on 1 or more generator is infinite, and the free group on 2 or more generators is nonabelian. Since the image of an abelian group is abelian, if in the situation of Lemma 9.1 the set J has more than two elements, then the automorphism group at any point of G is nonabelian. We apply this to the Jordan Curve Theorem in Lemma 9.4, when proving that the complement of a simple closed curve on the sphere has at most 2 path components. First, we prove some simpler lemmas about path components of subspaces of the sphere.

Lemma 9.2. Let x be a point on the sphere and define $X = \mathbb{S}^2 \setminus \{x\}$. Suppose D, E are disjoint closed subsets of X and a, b points of X. If a and b can be connected by a path in $X \setminus D$ and $X \setminus E$, then they can be connected by a path in $X \setminus (D \cup E)$. †

Proof. Suppose that the desired property does not hold, so that a and b can be connected by a path in $X \setminus D$ and $X \setminus E$ but not in $X \setminus (D \cup E)$. We can assume that a and b are distinct, since otherwise they can be connected by the constant path. Let $X_1 = X \setminus D$, $X_2 = X \setminus E$, $X_0 = X \setminus (D \cup E)$, so that X_1 and X_2 are open and $X_0 = X_1 \cap X_2$. Pick J to contain one point in each path component of X_0, and pick it such that a and b are in J (this is possible since they are in different path components by assumption). Choose a point p in J. Then by the Seifert-van Kampen theorem and Lemma 9.1, there is a surjection $\pi(X, p)$ to F, where F is the free group on at least one generator (here we use the fact that a and b can be connected by a path in X_1 and X_2). There are no surjections from a one element set to an infinite set, so this is a contradiction. □

Lemma 9.3. Let I be a subset of of $\mathbb{S}^2$ homeomorphic to the interval. Then $\mathbb{S}^2 \setminus I$ is path connected.

† This is called the Phragmen-Brouwer property.

Proof. Suppose we had points a and b in $\mathbb{S}^2 \setminus I$ which could not be connected by a path in $\mathbb{S}^2 \setminus I$. Fix a homeomorphism $h : \mathbb{I} \to A$. Let $x = h(0.5)$ be the midpoint of A and write $A = A' \cup A''$ where $A' = h([0, 0.5])$ and $A'' = h([0.5, 1])$, so $A' \cap A'' = \{x\}$. Since A' and A'' are the continuous images of compact sets, they are compact. Then by Heine-Borel (Lemma 3.9), they are closed in $\mathbb{R}^2$, and so closed in $\mathbb{S}^2$. Thus $A' \setminus \{x\}$ and $A'' \setminus \{x\}$ are disjoint and closed in $\mathbb{S}^2 \setminus \{x\}$, and their union is $A \setminus \{x\}$. Because a and b cannot be connected in $(\mathbb{S}^2 \setminus \{x\}) \setminus (A \setminus \{x\}) = \mathbb{S}^2 \setminus A$, the previous lemma implies they cannot be connected by a path in at least one of $A' \setminus \{x\}$ or $A'' \setminus \{x\}$. Let A_1 be one of A', A'' such that a and b cannot be connected by a path in $\mathbb{S}^2 \setminus A_1$. Repeat this process and obtain a sequence A_i of bisections of A such that a and b lie in different path components of $\mathbb{S}^2 \setminus A_i$ for each i. Let $I_i = h^{-1}(A_i)$ be the sequence of sub-intervals of $\mathbb{I}$ corresponding to the A_i. By construction, the length of each I_i is half that of the previous one, and $I_{i+1} \subseteq I_i$, so the intersection of all I_i is a single point t. Thus

$$\bigcap_{i \geq 1} A_i = \bigcap_{i \geq 1} h(I_i) = h\left(\bigcap_{i \geq 1} I_i\right) = h(\{t\}) = \{h(t)\}.$$

Since $\mathbb{S}^2 \setminus \{h(t)\}$ is homeomorphic to $\mathbb{R}^2$ (Lemma 8.2), it is path connected, and so there is a path p joining a to b in $\mathbb{S}^2 \setminus \{h(t)\}$. Since a and b cannot be joined by a path in any A_i, we can pick a point x_i in both A_i and the image of p for each i. Since the A_i shrink towards $\{h(t)\}$, this sequence x_i converges to $h(t)$. But the image of p is compact (as it is the continuous image of a compact set), so by Heine-Borel (Lemma 3.9) it is closed, which means it contains its limit point $h(t)$, contradicting the fact that the codomain of p excludes $h(t)$. $\square$

From here, we are able to prove that the complement of a simple closed curve on the sphere has at most 2 path components. We'll split up the curve into two arcs each homeomorphic to the interval, then apply the Seifert-van Kampen Theorem and Lemmas 9.1, 9.2, and 9.3.

Lemma 9.4. The complement of a simple closed on the sphere has at most two path components.

Proof. Let C be a simple closed curve on $\mathbb{S}^2$. Decompose it as $A \cup B$ where A and B are homeomorphic to $\mathbb{I}$ and A and B meet only at the points a and b. Let $X = \mathbb{S}^2 \setminus \{a, b\}$ and define $X_1 = \mathbb{S}^2 \setminus A$, $X_2 = \mathbb{S}^2 \setminus B$, $X_0 = X_1 \cap X_2 =$

$\mathbb{S}^2\setminus(A\cup B)=\mathbb{S}^2\setminus C$. By Heine-Borel (Lemma 3.9), A, B are closed in $\mathbb{R}^3$ and so in $\mathbb{S}^2$, and thus X_1, X_2 are open in $\mathbb{S}^2\setminus\{a,b\}$. If we take J to contain one element in each path component of X_0, we get by the Seifert van-Kampen Theorem a pushout

$$\begin{tikzcd} \pi_J X_0 \arrow[r] \arrow[d] & \pi_J X_1 \arrow[d] \\ \pi_J X_2 \arrow[r] & \pi_J X \end{tikzcd}$$

in which all arrows are the identity on objects. By Lemma 9.3, X_1 and X_2 are path connected, and so by Lemma 9.1, the fundamental group of X at any point in J surjects onto a free group on $n_C-n_A-n_B+1=n_C-1-1+1=n_c-1=|J|-1$ generators. If J had more than 2 elements, this free group would be nonabelian, but by Lemma 8.5, $c\pi(X,p)\cong\mathbb{Z}$ is abelian. Since there are no surjections from an abelian group to a nonabelian one, this implies $|J|\leq 2$. □

This establishes the first step of the Jordan Curve Theorem: the complement of any simple closed curve on the unit sphere has at most two path components. We prove below that it has more than one path component.

Lemma 9.5. The complement of a simple closed curve in $\mathbb{S}^2$ is not path connected.

Proof. Let C be a simple closed curve and suppose $\mathbb{S}^2\setminus C$ had only one path component. Decompose C as $A\cup B$ for A, B homeomorphic $\mathbb{I}$ and meeting only at a, b. Now note that $\mathbb{S}^2\setminus C$ is nonempty, since otherwise we would have $\mathbb{S}^2=C$, and so the sphere and the circle would be homeomorphic, contradicting the fact that the fundamental group of the circle is the integers and that of the sphere is trivial. Pick any $x_0\in\mathbb{S}^2\setminus C$. Since $\mathbb{S}^2\setminus C$ is path connected by assumption, the set $\{x_0\}$ is representative in $\mathbb{S}^2\setminus C$, and so by the Seifert-van Kampen theorem we have a pushout (of groupoids)

$$\begin{tikzcd} \pi_{\{x_0\}}(\mathbb{S}^2\setminus C) \arrow[r] \arrow[d] & \pi_{\{x_0\}}(\mathbb{S}^2\setminus A) \arrow[d, "i^*"] \\ \pi_{\{x_0\}}(\mathbb{S}^2\setminus B) \arrow[r, "j^*"] & \pi_{\{x_0\}}(\mathbb{S}^2\setminus\{a,b\}) \end{tikzcd}$$

where $i:\mathbb{S}^2\setminus A\to\mathbb{S}^2\setminus\{a,b\}$ and $j:\mathbb{S}^2\setminus B\to\mathbb{S}^2\setminus\{a,b\}$ are inclusion maps, and i^*, j^* are maps induced by these inclusions. We prove that i^* is the trivial map (the one which sends all morphisms in $\pi_{x_0}(\mathbb{S}^2 \setminus A)$ to the identity).

First, parameterize A by a map $\gamma : \mathbb{I} \to A$ which sends 0 to b and 1 to a. Now for any loop $f : [0, r] \to \mathbb{S}^2 \setminus \{a, b\}$, based at x_0, we show $i^*([f]) = [i \circ f] = id$. We do so by showing that the map $g : \mathbb{S}^1 \to \mathbb{S}^2 \setminus \{a, b\}$ defined by $g(e^{it}) = i\left(f\left(\frac{r}{2\pi}t\right)\right)$ is homotopic to a constant function, and conclude $i \circ f \sim c_{x_0}$ by Lemma 6.15. Let $R : \mathbb{S}^2 \setminus \{a\} \to \mathbb{S}^2 \setminus \{N\}$ be the restriction of the homeomorphism in Lemma 8.1 and $h : \mathbb{S}^2 \setminus \{N\} \to \mathbb{R}^2$ the homeomorphism in Lemma 8.2. Define $T : \mathbb{S}^2 \setminus \{a\} \to \mathbb{R}^2$ by $T(x) = h(R(x)) - h(R(b))$, so $T(b)$ is the origin and $\lim_{x \to a, x \in \mathbb{S}^2} |T(x)| \geq \lim_{x \to a, x \in \mathbb{S}^2} |h(R(x))| - |h(R(b))| = \infty - h(R(b)) = \infty$. Since g does not have b in its range, the composition $T \circ g$ maps $\mathbb{S}^1$ to $\mathbb{R}^2 \setminus \{0\}$, and also $T(\gamma(0)) = T(b) = (0, 0)$.

Now since g is continuous and $\mathbb{S}^1$ is compact, the image of $T \circ g$ is compact, and so by Heine-Borel it (Lemma 3.9) is bounded. Thus there is some $r > 0$ such that $|T(g(x))| < r$ for all $x \in \mathbb{S}^1$. Since $\lim_{t \to 1, t \in \mathbb{I}} |T(\gamma(t))| = \lim_{x \to a, x \in \mathbb{S}^2} |T(x)| = \infty > r$, there is some $t_0 \in [0, 1)$ such that $|T(\gamma(t_0))| > r$. Let $y = T(\gamma(t_0))$. We can't have $t_0 = 0$ since then $|T(\gamma(t_0))| = |T(b)| = |0| = 0$. Let $\lambda : \mathbb{I} \to \mathbb{R}^2$ be the path from 0 to y given by $\lambda(t) = T(\gamma(t/t_0))$. Define the homotopy $H : \mathbb{S}^1 \times \mathbb{I} \to \mathbb{R}^2$ from $T \circ g$ to the constant map $x \mapsto -y$ by

$$H(x, t) = \begin{cases} T(g(x)) - \lambda(2t) & \text{if } 0 \leq t \leq 1/2 \\ (2 - 2t)T(g(x)) - y & \text{if } 1/2 \leq t \leq 1 \end{cases}.$$

This is well defined because $T(g(x)) - \lambda(2 \cdot 1/2) = T(g(x)) - y = (2 - 2 \cdot 1/2)T(g(x)) - y$, and is continuous by the Gluing Lemma. We prove it is never 0. If $H(x, t) = 0$ for $t \in [0, 1/2]$, then $T(g(x)) = \lambda(2t) = T(\gamma(2t/t_0))$, so $g(x) = \gamma(2t/t_0)$. But $\operatorname{im} g = \operatorname{im} f \subseteq \mathbb{S}^2 \setminus A$ and $\operatorname{im} \gamma = A$. Thus $H(x, t) \neq 0$ for $0 \leq t \leq 1/2$. If $1/2 \leq t \leq 1$, then $H(x, t) = 0$ iff $(2 - 2t)T(g(x)) = y$. But $1/2 \leq t \leq 1$ implies $0 \leq 2 - 2t \leq 1$, and thus $|(2 - 2t)T(g(x))| \leq |T(g(x))| < r < |y|$, so this is impossible. Now finally note that $H(x, 0) = T(g(x)) - \lambda(2 \cdot 0) = T(g(x))$ and $H(x, 1) = (2 - 2 \cdot 1)T(g(x)) - y = -y$. Thus H is a homotopy between $T \circ g$ and $x \mapsto -y$ in $\mathbb{R}^2 \setminus$. Then $T^{-1} \circ H$ is a homotopy between g and $x \mapsto T^{-1}(-y)$ in $\mathbb{S}^2 \setminus \{a\}$. But we also checked that H misses 0, so this is a homotopy in $\mathbb{S}^2 \setminus \{a, b\}$. Thus i^* sends all loops to the trivial loop.

By the same argument, j^* is trivial. Then by Lemma 4.9[†], $\pi(\mathbb{S}^2 \setminus \{a, b\}, x_0)$

[†]Techincally we need to show that the pushout of one object groupoids above induces a pushout of the corresponding groups, but this is easy.

is trivial, contradicting Lemma 8.5. Thus $\{x_0\}$ cannot be representative, so $\mathbb{S}^2$ isn't path connected. □

This gives us our first version of the Jordan Curve Theorem.

Theorem 9.6. Let C be a simple closed curve on the sphere. Then $\mathbb{S}^2 \setminus C$ has exactly two path components.

We now prove that the path components of $\mathbb{S}^2 \setminus C$ agree with the connected components, and then translate this result from te sphere to the plane. As we saw in the previous section, the sphere is just the plane plus a point "at infinity", so given a curve in the plane we can lift it to one on the sphere.

Lemma 9.7. Let C be a simple closed curve on the sphere. Then the path components of $\mathbb{S}^2 \setminus C$ coincide with the connected components. In particular, there are exactly two connected components.

Proof. By Thoerem 9.6, $\mathbb{S}^2 \setminus C$ has exactly two path components, say P and Q. Since all points lie in some path component, $P \cup Q = \mathbb{S}^2 \setminus C$. Also P and Q are disjoint, since if $x \in P$ and $x \in Q$, then for any $y \in P$ and $z \in Q$, we can connect y to x and x to z by paths, so y and z are connected by a path, from which we see $P = Q$ (a contradiction). We show that P is open using Lemma 8.6, and conclude by symmetry that Q is as well. Let x be a point in P. Then since C is closed, $\mathbb{S}^2 \setminus C$ is open, and thus a neighborhood of x. Then by Lemma 8.6 there is some path connected neighborhood M of x on $\mathbb{S}^2$ contained in $\mathbb{S}^2 \setminus C$. This is also neighborhood of x on $\mathbb{S}^2 \setminus C$, and since all points in M can be joined to x by a path, $M \subseteq P$. Thus P contains a neighborhood of each of its points, so P is open, and by symmetry Q is as well.

We now show P (and thus Q) is connected—in fact, this argument shows any path connected set is connected. We do this by showing that any continuous function $f : P \to \{0, 1\}$ is not a surjection. Let f be such a continuous function and pick a point $x_0 \in P$. Then for any $y \in P$, there is a path $p : \mathbb{I} \to P$ connecting x_0 and y. Then $q = f \circ p : \mathbb{I} \to \{0, 1\}$ is a continuous function such that $q(0) = f(p(0)) = f(x_0)$ and $q(1) = f(p(1)) = f(y)$. Since $\mathbb{I}$ is connected, q must be constant, so $f(x_0) = f(y)$. Thus f is constant, so P is connected.

Thus P and Q are nonempty connected open sets which partition $\mathbb{S}^2 \setminus C$. We show that if $R \subseteq \mathbb{S}^2 \setminus C$ is connected, then $R = P$ or $R = Q$. This will imply that P and Q are the largest connected sets of $\mathbb{S}^2 \setminus C$, and thus by definition the connected components. Suppose R is not contained in P or in Q. Then there are points $x, y \in R$ such that $x \notin P$ and $y \notin Q$. Since $P \cup Q = \mathbb{S}^2 \setminus C$, this implies $y \in P$ and $x \in Q$. Let $A = R \cap P$ and $B = R \cap Q$. Then A and B are open in R by definition of the subspace topology, since P and Q are open in $\mathbb{S}^2 \setminus C$. Also, $A \cap B = (P \cap R) \cap (Q \cap R) = (P \cap Q) \cap R = \emptyset$ since P and Q are disjoint. Thus (A, B) is a disconnection of R in the sense of Lemma 3.6, so R is not connected. By the contrapositive, this implies every connected set is contained entirely in P or Q, and so P and Q are the connected components of $\mathbb{S}^2 \setminus C$. □

We've almost established the Jordan Curve Theorem, and really all the hard work is past. Given a curve on the plane, we can reverse-stereographically-project that curve onto the punctured sphere, then re-project the components of the complement into the plane.

Theorem 9.8 (The Jordan Curve Theorem)**.** Let C be a simple closed curve in the plane. Then $\mathbb{R}^2 \setminus C$ has two connected components, A and B. One of these components is bounded and the other is unbounded. Considering A and B as subsets of $\mathbb{R}^2$, $\overline{A} \setminus A = C$ and $\overline{B} \setminus B = C$, i.e. the boundary of both A and B is C.

Proof. Let C be such a simple closed curve in the plane. Let $h : \mathbb{S}^2 \setminus \{N\} \to \mathbb{R}^2$ be the stereographic projection homeomorphism from Lemma 8.2. Then $h^{-1}(C)$ is homeomorphic to C is homeomorphic to a circle, so $h^{-1}(C)$ is a simple closed curve $\mathbb{S}^2 \setminus \{N\}$, and thus is one on the sphere. By Lemma 9.7, there are two connected components A' and B' of $\mathbb{S}^2 \setminus h^{-1}(C)$, and by an argument in the proof of Lemma 9.7, A' and B' partition $\mathbb{S}^2 \setminus h^{-1}(C)$. Thus N is in exactly one of A', B'; w.l.o.g. suppose it's A'. Since A' is open (see the proof of the previous lemma), there's a ball U around N contained in A', and so points sufficiently close to N are in A'. If we let $A = h(A' \setminus \{A\})$, this implies A is unbounded, since $\lim_{x \to N, x \in \mathbb{S}^2} |h(x)| = \infty$. Then also if $B = h(B')$, we have $B \subseteq h(\mathbb{S}^2 \setminus U)$. $\mathbb{S}^2 \setminus U$ is bounded since it's contained in $\mathbb{S}^2$, and is closed as a subset of $\mathbb{S}^2$ because U was taken to be open. Thus there's a closed set W of $\mathbb{R}^3$ such that $\mathbb{S}^2 \setminus U = \mathbb{S}^2 \cap W$, and since $\mathbb{S}^2$ is closed, this shows $\mathbb{S}^2 \setminus U$ is closed as well. Thus $\mathbb{S}^2 \setminus U$ is compact by Heine-Borel (Lemma 3.9), so $h(\mathbb{S}^2 \setminus U)$ is also compact, and so by the reverse direction of Heine-Borel it is bounded. Thus B is contained in a bounded set, and so is bounded.

We now show A and B are the connected components of $\mathbb{R}^2 \setminus C$. It's immediate that B is connected, since it is homeomorphic to a connected space (namely B'). To show A is connected, it suffices to show $A' \setminus \{N\}$ is. Because U is open, there must be some $\varepsilon > 0$ such that $U' = \{(a, b, c) \in \mathbb{S}^2 : c > 1-\varepsilon\}$ is contained in U. $U' \setminus \{N\}$ is homeomorphic to the disk minus its center, which is path connected by paths which go around the inner circles and along radial lines. Thus $U' \setminus \{N\}$ is path connected, and so is connected. Now suppose we had a continuous map $f : A' \setminus \{N\} \to \{0, 1\}$. By continuity, f must be connected on $U' \setminus \{N\}$, say with value n. Thus the function $g : U' \to \{0, 1\}$ defined by $g(x) = n$ is continuous, and so by the Gluing Lemma the function $h : A' \to \{0, 1\}$ which is g on U' and f on $A' \setminus \{N\}$ is continuous. Since A' is connected, h must be constant on A', and so f must be constant on $A' \setminus \{N\}$. Thus $A' \setminus \{N\}$ is connected. Thus A and B are connected subsets of $\mathbb{R}^2 \setminus C$. We prove they are maximal (and thus connected components) as follows. Suppose we had a connected set D in $\mathbb{R}^2 \setminus C$. If $A \subseteq D$, then $h^{-1}(A) \subseteq h^{-1}(D) \subseteq \mathbb{R}^2 \setminus h^{-1}(C)$, so $A' = h^{-1}(A) \cup \{N\} \subseteq h^{-1}(D) \cup \{N\}$. The set $h^{-1}(D) \cup \{N\}$ is connected because $h^{-1}(D) \cup \{N\} = h^{-1}(D \cup A) \cup \{N\} = h^{-1}(D) \cup A'$ is the union of two connected sets with connected intersection (namely $h^{-1}(A) = A' \setminus \{N\}$). Because A' is a connected component of $\mathbb{R}^2 \setminus h^{-1}(C)$, this implies $h^{-1}(A) \cup \{N\} = h^{-1}(D) \cup \{N\}$. Since N is not in the image of h, we get from this that $h^{-1}(A) = h^{-1}(D)$, so $A = D$. The same argument (without juggling around N) proves B is a connected component.

Thus it suffices to prove that the boundaries of A and B in $\mathbb{R}^2$ are C. Because A' and B' are open in $\mathbb{S}^2 \setminus h^{-1}(C)$ (as we showed in Lemma 9.7) and h is a homeomorphism, A and B must also be open in $\mathbb{R}^2 \setminus C$. Since C is homeomorphic to the circle, it is compact, and so closed in $\mathbb{R}^2$. Thus $\mathbb{R}^2 \setminus C$ is open, and so A and B are open in $\mathbb{R}^2$. This implies $\overline{A} \cap B = \emptyset$, because if there were a point $x \in \overline{A} \cap B$, then since B is open B would contain a neighborhood of x, and then by the fact that $x \in \overline{A}$ this neighborhood would contain a point of A. But since A' and B' are disjoint, so are A and B, as

$$A \cap B = h(A' \setminus \{N\}) \cup h(B') \subseteq h(A') \cup h(B') = h(A' \cup B') = h(\emptyset) = \emptyset.$$

Then since $\mathbb{R}^2 = A \cup B \cup C$, we must have $\overline{A} \subseteq A \cup C$, and so $\overline{A} \setminus A \subseteq C$. By the same argument, $\overline{B} \setminus B \subseteq C$. Then since A and B are disjoint from C, it suffices to show $C \subseteq \overline{A}$ and $C \subseteq \overline{B}$. Let x be some point of C and let M be any neighborhood of x, and w.l.o.g. suppose it is open. We show M meets both A and B. Since h is continuous, $M' = h^{-1}(M)$ is a neighborhood of x in $\mathbb{S}^2 \setminus \{N\}$. By the definition of the subspace topology, M' is the intersection of $\mathbb{S}^2 \setminus \{N\}$ with some neighborhood of $h^{-1}(x)$ in $\mathbb{S}^2$. But since $\mathbb{S}^2 \setminus \{N\}$ is open it is a neighborhood of $h^{-1}(x)$ in $\mathbb{S}^2$, and so M' is as well.

Now choose points $a \in A'$ and $b \in B'$; since a and b are in different path components of $\mathbb{S}^2 \setminus C$, they cannot be connected by a path. By Lemma 9.4, $\mathbb{S}^2 \setminus D$ is path connected, and so there is a path $p : \mathbb{I} \to \mathbb{S}^2 \setminus D$ from a to b. Then there must be some point both E and the image of p, since otherwise p would be a path from x to y in $\mathbb{S}^2 \setminus C$. Then the set $T = \{t \in \mathbb{I} : p(t) \in E\}$ is nonempty. Let $t_0 = \inf T$. Then T is closed (it is the preimage of the closed set E), we must have $p(t_0) \in E$. Since $p(t_0) \notin D$, $p(t_0)$ is neither of e_0 or e_1. Thus it is in the interior of M', so points sufficiently close to $p(t_0)$ are in M'. Let $t_1 < t_0$ be such that $p(t_1)$ is close to $p(t_0)$. Since t_0 is the inf of the set T, we must have $p(t) \notin E$ for $t \leq t_0$, and also $p(t_1)$ must not be p in E. Thus there is a path $q : [0, t_1] \to \mathbb{S}^2 \setminus C$ given by $q(t) = p(t)$ from a to $p(t_1)$. Then because $a \in A'$ and A' is a path component, this implies $p(t_1) \in A'$. Thus A' meets M', and since $M' \subseteq \mathbb{S}^2 \setminus \{N\}$, this implies $A' \setminus \{N\}$ meets M'. Pushing through h, we see A meets M. By applying the same argument to the supremum of T, this shows B meets M. Thus every neighborhood of x meets A and B, so x is in in $\overline{A}$ and $\overline{B}$. □

References

[1] Ronald Brown. *Topology and Groupoids.* BookSurge, 2006.

[2] Ronald Brown. *Corrigendum to "Groupoids, the Phragmen-Brouwer Property, and the Jordan Curve Theorem".* J. Homotopy and Related Structures, 2006.

[3] Paolo Aluffi. *Algebra: Chapter 0.* American Mathematical Society, 2009.