Yet another book on algebraic topology I: covering spaces and the fundamental group

Andrew Putman

Department of Mathematics, University of Notre Dame, 255 Hurley Hall, Notre Dame, IN 46556

Email address: andyp@nd.edu

Table of contents

Part 1

Basic topics

CHAPTER 1

Definition and basic properties of covering spaces

Our first main topic is the theory of covering spaces. This chapter contains some basic definitions and a large number of examples. A first-time reader might be tempted to skip the examples and focus on the theory. This would be a mistake. The richness of the examples is what gives this subject its flavor, and it is impossible to understand the theoretical aspects of covering spaces without having absorbed a large store of these examples.

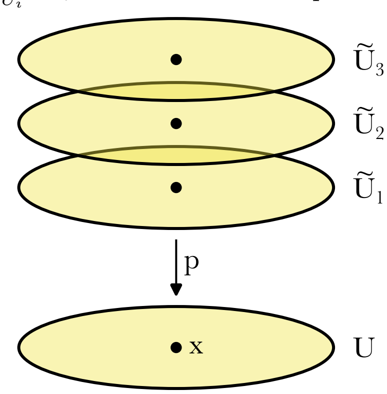
1.1. Definition and examples

Recall that a *local homeomorphism* is a map $p\colon Z \to X$ such that all $z \in Z$ have open neighborhoods V with $U = p(V)$ open and $p|_V\colon V \to U$ a homeomorphism. Roughly speaking, in a covering space this condition is strengthened by adding a uniformity condition to these V.

1.1.1. Definition of covering space. The definition is as follows:

Definition 1.1.1. A *covering space* or simply a *cover* of a space X is a space $\widetilde{X}$ equipped with a map $p\colon \widetilde{X} \to X$ such that for all $x \in X$, there is an open neighborhood U of x satisfying:

- the preimage $p^{-1}(U)$ is the disjoint union of open subsets $\{\widetilde{U}_i\}_{i \in \mathcal{I}}$ of $\widetilde{X}$ such that for all $i \in \mathcal{I}$, the restriction $p|_{\widetilde{U}_i}\colon \widetilde{U}_i \to U$ is a homeomorphism.



We call U a *trivialized neighborhood* of x (or just a *trivialized open set* if we do not want to emphasize x) and each $\widetilde{U}_i$ a *sheet* of $\widetilde{X}$ over U. We will also often call the map $p\colon \widetilde{X} \to X$ a covering space, and refer to X as the *base* of the cover. □

Remark 1.1.2. We allow $p^{-1}(U) = \emptyset$. In particular, for any space X the map $p\colon \emptyset \to X$ is a covering space. This convention is controversial, and some authors require the maps in covering spaces to be surjective. □

Remark 1.1.3. Covering spaces are local homeomorphisms, but the converse does not hold. However, if $\widetilde{X}$ is compact Hausdorff then all local homeomorphisms $p\colon \widetilde{X} \to X$ are covering spaces. We will say more about this in §1.2 below. □

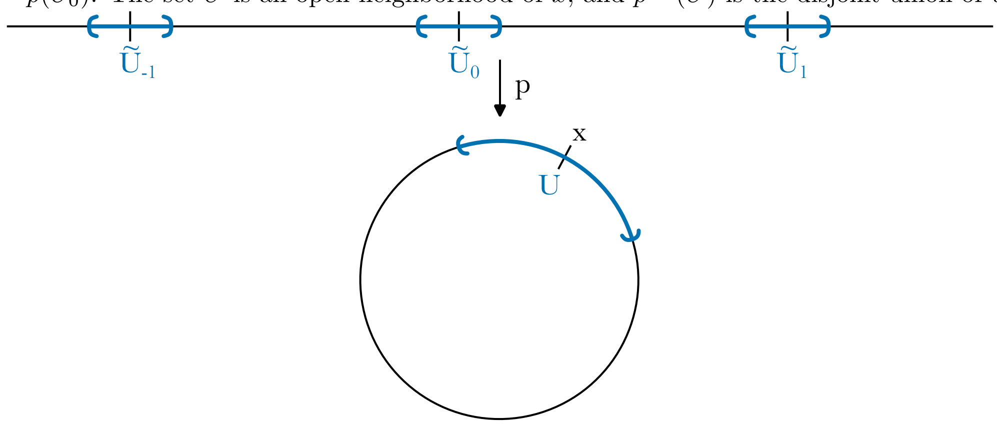
1.1.2. Two examples of covering spaces. Here are two basic examples:

Example 1.1.4 (Trivial cover). For a space X, the identity map $\mathbb{1}_X\colon X \to X$ is a covering space. More generally, for any discrete set $\mathcal{I}$ the projection map $p\colon X \times \mathcal{I} \to X$ is a covering space. We will call these the *trivial covers* of X. □

EXAMPLE 1.1.5 (Universal cover of circle). Regard $\mathbb{S}^1$ as the unit circle in the complex plane $\mathbb{C}$. Let $p\colon \mathbb{R} \to \mathbb{S}^1$ be the map $p(\theta) = e^{2\pi i\theta}$. This is a covering space. Indeed, consider $x \in \mathbb{S}^1$. Write $x = e^{2\pi i\theta_0}$. Pick $\epsilon > 0$ with $\epsilon < 1$. For $n \in \mathbb{Z}$, set

$$\widetilde{U}_n = (\theta_0 + n - \epsilon, \theta_0 + n + \epsilon) \subset \mathbb{R}.$$

Let $U = p(\widetilde{U}_0)$. The set U is an open neighborhood of x, and $p^{-1}(U)$ is the disjoint union of the $\widetilde{U}_n$:



Each $\widetilde{U}_n$ projects homeomorphically to U, so U is a trivialized neighborhood of x and the $\widetilde{U}_n$ are the sheets over U. The covering space $p\colon \mathbb{R} \to \mathbb{S}^1$ is the *universal cover* of $\mathbb{S}^1$. The reason for this name will become clear later when we classify covering spaces in Chapter YYY. □

1.1.3. Degree of cover. Let $p\colon \widetilde{X} \to X$ be a covering space. The preimages $p^{-1}(x) \subset \widetilde{X}$ of points $x \in X$ are called the *fibers* of $p\colon \widetilde{X} \to X$. For $x \in X$, the fiber $p^{-1}(x)$ is called the *fiber over* x. The first main property of covering spaces is that if X is connected, then the cardinalities of its fibers are all equal. More generally:

LEMMA 1.1.6. *Let $p\colon \widetilde{X} \to X$ be a covering space. Let $f\colon X \to \mathbb{Z} \cup \{\infty\}$ be the function*

$$f(x) = |p^{-1}(x)| \quad \text{for } x \in X.$$

Then f is locally constant. In particular, if X is connected then f is constant.

PROOF. Consider $x \in X$. Let U be a trivialized neighborhood of x and let $\{\widetilde{U}_i\}_{i\in\mathcal{I}}$ be the sheets of $\widetilde{X}$ over U. For $y \in U$, the preimage $p^{-1}(y)$ consists of one point in each $\widetilde{U}_i$, and thus $f(y) = |\mathcal{I}|$. The lemma follows. □

This suggests the following definition:

DEFINITION 1.1.7. Let $p\colon \widetilde{X} \to X$ be a covering space. We say that $p\colon \widetilde{X} \to X$ has *degree* n if all of its fibers have cardinality n. This degree might be infinity. We will also say that $p\colon \widetilde{X} \to X$ is an n*-sheeted* or an n*-fold* cover. □

Lemma 1.1.6 implies that if X is connected, then every covering space $p\colon \widetilde{X} \to X$ has a degree.[1] For instance, the degree of the universal cover $p\colon \mathbb{R} \to \mathbb{S}^1$ is infinity.

1.1.4. More examples. Here are some more examples of covering spaces:

EXAMPLE 1.1.8 (Degree n cover of circle). Regard $\mathbb{S}^1$ as the unit circle in $\mathbb{C}$. Fix some $n \geq 1$, and define $p_n\colon \mathbb{S}^1 \to \mathbb{S}^1$ via the formula $p_n(z) = z^n$. This is a degree n covering space. Indeed, consider $x \in \mathbb{S}^1$. The preimage $p_n^{-1}(x)$ consists of n distinct points: writing $x = e^{2\pi i\theta_0}$, we have

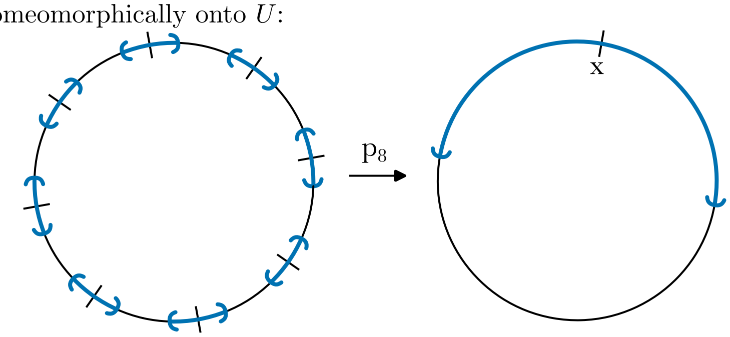
$$p_n^{-1}(x) = \left\{ e^{2\pi i(\theta_0+m)/n} \mid m \text{ is an integer with } 0 \leq m < n \right\}.$$

Fix some $\epsilon > 0$ with $\epsilon < 1$, and let

$$U = \left\{ e^{2\pi i\theta} \mid \theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon) \right\}.$$

[1] This is false for non-connected spaces. See Exercise 1.3.

The set U is an open neighborhood of x, and $p_n^{-1}(U)$ is the disjoint union of n subsets of $\mathbb{S}^1$ each of which projects homeomorphically onto U:



Thus U is a trivialized neighborhood of x and the components of $p_n^{-1}(U)$ are the sheets over U. $\square$

Example 1.1.9 (Cosets of discrete subgroups). Let $\mathbf{G}$ be a topological group, i.e., a group that is a topological space such that the product map $\mathbf{G} \times \mathbf{G} \to \mathbf{G}$ and inversion map $\mathbf{G} \to \mathbf{G}$ are continuous. Let $\mathbf{H}$ be a discrete subgroup of $\mathbf{G}$. Here are two examples to keep in mind:

- $\mathbf{G}$ the additive group $\mathbb{R}^n$, and $\mathbf{H} = \mathbb{Z}^n$; and
- $\mathbf{G} = \mathrm{SL}_n(\mathbb{R})$ and $\mathbf{H} = \mathrm{SL}_n(\mathbb{Z})$.

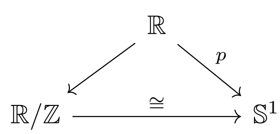
Endow the set $\mathbf{G}/\mathbf{H} = \{g\mathbf{H} \mid g \in \mathbf{G}\}$ of left cosets with the quotient topology. Then the quotient map $p\colon \mathbf{G} \to \mathbf{G}/\mathbf{H}$ is a cover of degree $|\mathbf{H}|$. Indeed, consider a point $g_0\mathbf{H}$ of $\mathbf{G}/\mathbf{H}$. Since $\mathbf{H}$ is a discrete subgroup of $\mathbf{G}$, we can find an open neighborhood V of $1 \in \mathbf{G}$ whose translates $\{Vh \mid h \in \mathbf{H}\}$ are all disjoint.[2] Set $U = p(g_0V)$, so

$$p^{-1}(U) = \bigsqcup_{h \in \mathbf{H}} g_0Vh.$$

These are all disjoint sets that project homeomorphically to U, so U is a trivialized neighborhood and the sets g_0Vh with $h \in \mathbf{H}$ are the sheets above U. $\square$

Example 1.1.10. Two of our previous examples are special cases of Example 1.1.9:

- The universal cover $p\colon \mathbb{R} \to \mathbb{S}^1$. Indeed, the additive topological group $\mathbb{R}$ contains the discrete subgroup $\mathbb{Z}$. The satisfies $\mathbb{R}/\mathbb{Z} \cong \mathbb{S}^1$, and this homeomorphism fits into a commutative diagram



  Using this, we can identify the covers $\mathbb{R} \to \mathbb{R}/\mathbb{Z}$ and $p\colon \mathbb{R} \to \mathbb{S}^1$.
- The covers $p_n\colon \mathbb{S}^1 \to \mathbb{S}^1$ defined by $p_n(z) = z^n$. Indeed, $\mathbb{S}^1 \subset \mathbb{C}$ is a topological group under multiplication, and it contains the discrete group μ_n of n^{th} roots of unity. The quotient $\mathbb{S}^1/\mu_n$ is homeomorphic to $\mathbb{S}^1$, and just like above we can identify the covers $\mathbb{S}^1 \to \mathbb{S}^1/\mu_n$ and $p_n\colon \mathbb{S}^1 \to \mathbb{S}^1$.

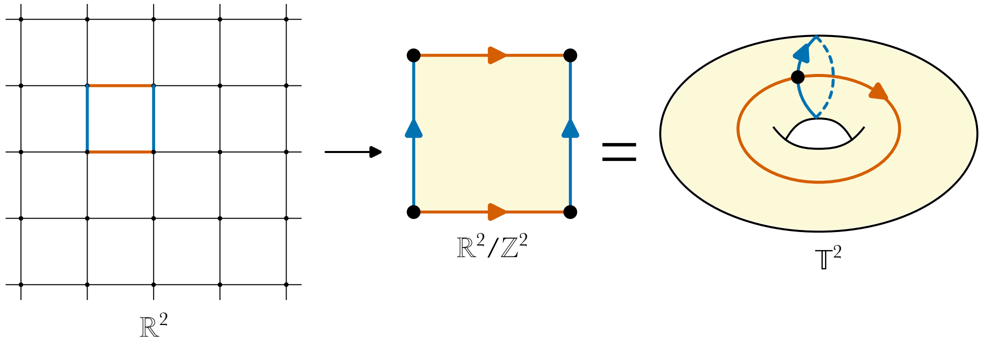
As another example, as we noted in Example 1.1.9 the additive group $\mathbb{R}^n$ contains the discrete subgroup $\mathbb{Z}^n$. The quotient $\mathbb{R}^n/\mathbb{Z}^n$ is homeomorphic to an n-dimensional torus $\mathbb{T}^n = (\mathbb{S}^1)^{\times n}$:

---

[2]Here are some more details. Since $\mathbf{H}$ is discrete, we can find an open neighborhood W of $1 \in \mathbf{G}$ such that $W \cap \mathbf{H} = \{1\}$. Let $f\colon \mathbf{G} \times \mathbf{G} \to \mathbf{G}$ be the map $f(xy) = xy^{-1}$. Since f is continuous, the set $f^{-1}(W)$ is an open neighborhood of $(1,1)$ and thus we can find open neighborhoods V_1 and V_2 of 1 such that $V_1 \times V_2 \subset f^{-1}(W)$. Letting $V = V_1 \cap V_2$, we then have $f(V \times V) \subset W$. We now claim that the sets $\{Vh \mid h \in \mathbf{H}\}$ are all disjoint. Indeed, if $h_1, h_2 \in \mathbf{H}$ are such that $(Vh_1) \cap (Vh_2) \neq \emptyset$, then we can find $v_1, v_2 \in V$ with $v_1h_1 = v_2h_2$, and hence

$$h_2h_1^{-1} = v_1v_2^{-1} \in f(V \times V) \cap \mathbf{H} \subset W \cap \mathbf{H} = \{1\}.$$

In other words, $h_1 = h_2$.

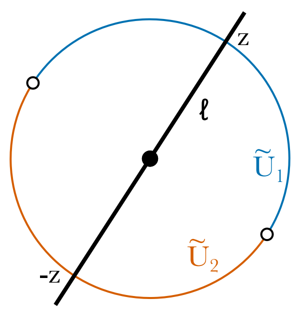


This figure shows the case $n = 2$. The action of $\mathbb{Z}^2$ on $\mathbb{R}^2$ identifies all the horizonal (resp. vertical) lines together, and the indicated region is a fundamental domain which in the quotient becomes a square with sides identified as indicated. Identifying $\mathbb{R}^n/\mathbb{Z}^n$ with $\mathbb{T}^n$ via this isomorphism, we get an infinite-degree cover $p\colon \mathbb{R}^n \to \mathbb{T}^n$. □

Example 1.1.11 (Real projective space). Let $\mathbb{RP}^n$ be n-dimensional real projective space, that is, the set of lines through the origin in $\mathbb{R}^{n+1}$. Topologize $\mathbb{RP}^n$ as follows:

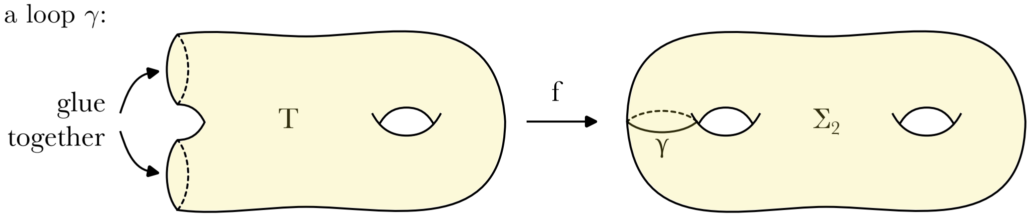
- Let $\pi\colon \mathbb{R}^{n+1} \setminus 0 \to \mathbb{RP}^n$ be the map taking a nonzero point $z \in \mathbb{R}^{n+1}$ to the line determined by 0 and z. Give $\mathbb{RP}^n$ the quotient topology determined by π, so a set $U \subset \mathbb{RP}^n$ is open if and only if $\pi^{-1}(U)$ is open.

We have $\mathbb{S}^n \subset \mathbb{R}^{n+1}$. Let $p\colon \mathbb{S}^n \to \mathbb{RP}^n$ be the restriction of π to p. This is a degree 2 covering space. Indeed, consider some point $\ell \in \mathbb{RP}^n$. The line ℓ intersects $\mathbb{S}^n$ in two antipodal points $z, -z \in \mathbb{S}^n$. Let $U \subset \mathbb{RP}^n$ be the set of all lines ℓ' that are *not* orthogonal to ℓ. This is an open set, and the preimage $p^{-1}(U)$ is the disjoint union of two open hemispheres $\widetilde{U}_1$ and $\widetilde{U}_2$ centered at z and $-z$, respectively:

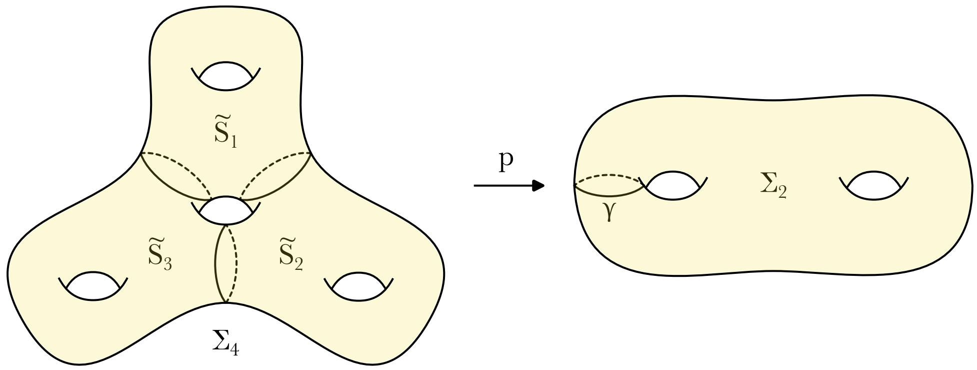


Each $\widetilde{U}_i$ projects homeomorphically to U, so U is a trivialized neighborhood and the $\widetilde{U}_i$ are the sheets over U. □

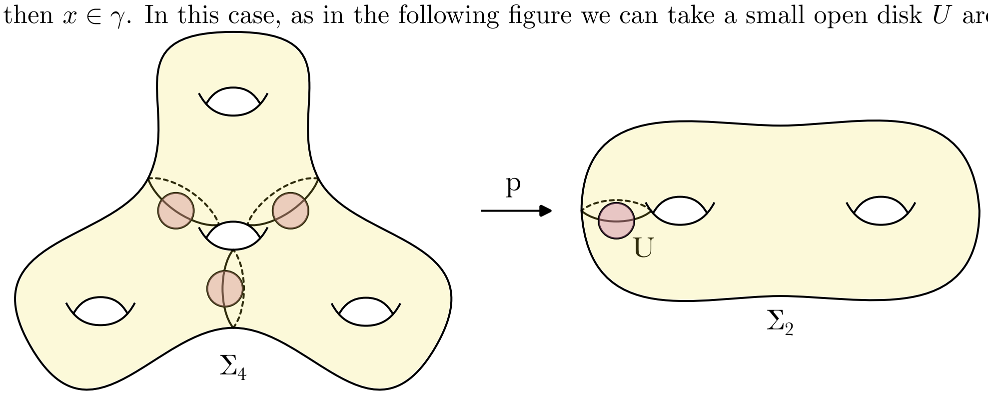
Example 1.1.12 (Cover of surface). Let Σ_2 be a genus 2 surface and let T be a genus 1 surface with two boundary components. Let $f\colon T \to \Sigma_2$ be the map that glues the boundary components to form a loop γ:



For $1 \leq i \leq 3$, let $\widetilde{S}_i$ be a copy of T. As in the following figure, we can glue the $\widetilde{S}_i$ together to form a genus 4 surface Σ_4 and use f to map each $\widetilde{S}_i$ to Σ_2, yielding a map $p\colon \Sigma_4 \to \Sigma_2$:



Each of the three black loops in Σ_4 maps homeomorphically onto the black loop γ in Σ_2. The map $p\colon \Sigma_4 \to \Sigma_2$ is a degree 3 covering space. Indeed, consider a point $x \in \Sigma_2$. If $x \notin \gamma$, then for our trivialized neighborhood we can take $U = \Sigma_2 \setminus \gamma$. The sheets above U are the $\mathrm{Int}(\widetilde{S}_i)$. If instead $x \notin S$, then $x \in \gamma$. In this case, as in the following figure we can take a small open disk U around x:



The three disks shown in Σ_4 are each mapped homeomorphically to U, so U is a trivialized neighborhood. □

1.2. Covers versus local homeomorphisms, and the fundamental theorem of algebra

We now give conditions that ensure a local homeomorphism is a covering space, and as an application give a simple proof of the fundamental theorem of algebra.

1.2.1. Criterion. Recall that a space Z is locally compact Hausdorff[3] if it is Hausdorff and every point $z \in Z$ has an open neighborhood W whose closure $\overline{W}$ is compact. Also, a map $f\colon Y \to Z$ is proper if for all compact subsets $K \subset Z$, the preimage $f^{-1}(K)$ is compact. With these definitions, we have:

LEMMA 1.2.1. *Let $p\colon \widetilde{X} \to X$ be a proper local homeomorphism between locally compact Hausdorff spaces. Then $p\colon \widetilde{X} \to X$ is a covering space.*

PROOF. Consider $x \in X$. Since p is proper, the set $p^{-1}(x)$ is compact. Since p is a local homeomorphism at each point of $p^{-1}(x)$, the set $p^{-1}(x)$ is also discrete. We deduce that $p^{-1}(x)$ is finite. Enumerate it as $p^{-1}(x) = \{\widetilde{x}_1, \ldots, \widetilde{x}_n\}$. For each $1 \le i \le n$, there exists a neighborhood $\widetilde{V}_i$ of $\widetilde{x}_i$ such that $p|_{\widetilde{V}_i}$ is a homeomorphism onto its image $V_i \subset X$. Since $\widetilde{X}$ is Hausdorff, we can shrink the $\widetilde{V}_i$ and assume they are all disjoint. Set $U = V_1 \cap \cdots \cap V_n$ and $\widetilde{U}_i = \widetilde{V}_i \cap p^{-1}(U)$. The set $\widetilde{U}_i$ is an open neighborhood of $\widetilde{x}_i$, and $p|_{\widetilde{U}_i}$ is a homeomorphism onto U.

By construction, $p^{-1}(U)$ contains $\widetilde{U}_1 \sqcup \cdots \sqcup \widetilde{U}_n$. However, we are not done since $p^{-1}(U)$ might contain points that do not lie in some $\widetilde{U}_i$. We want to shrink U to ensure that this does not happen. Since we need U to be open, we need to delete a closed set C of "bad points" from U.

The first step is to shrink U to ensure that $\overline{U}$ is compact. Since X is locally compact, we can find an open neighborhood W of x such that $\overline{W}$ is compact. Replacing U with $W \cap U$ and each $\widetilde{U}_i$

[3] Local compactness is poorly behaved on non-Hausdorff spaces.

with $p^{-1}(W) \cap \widetilde{U}_i$, we can assume that $U \subset W$. Since $\overline{U}$ is a closed subset of the compact set $\overline{W}$, it follows that $\overline{U}$ is compact.

Since p is proper $p^{-1}(\overline{U})$ is compact, so since $\widetilde{X}$ is Hausdorff $p^{-1}(\overline{U})$ is closed. Let

$$\widetilde{C} = p^{-1}(\overline{U}) \setminus \bigcup_{i=1}^{n} \widetilde{U}_i.$$

Since $\widetilde{C}$ is a closed subset of the compact set $p^{-1}(\overline{U})$, it follows that $\widetilde{C}$ is compact. This implies that $C = p(\widetilde{C})$ is compact, and hence closed. Replacing U with $U \setminus C$ and each $\widetilde{U}_i$ with $\widetilde{U}_i \setminus p^{-1}(C)$, we now have $p^{-1}(U) = \widetilde{U}_1 \sqcup \cdots \sqcup \widetilde{U}_n$, as desired. □

1.2.2. Fundamental theorem of algebra. As a first application of Lemma 1.2.1, we give a simple proof of the fundamental theorem of algebra. One way of stating this theorem is that every degree-n complex polynomial has n roots, at least if you count these roots with multiplicity. The nontrivial part of this is that every nonconstant polynomial has a root, so this is what we will prove:

THEOREM 1.2.2 (Fundamental theorem of algebra). *Let $f(z) \in \mathbb{C}[z]$ be a nonconstant polynomial. Then there exists some $z_0 \in \mathbb{C}$ such that $f(z_0) = 0$.*

PROOF. We will prove more generally that regarded as a map $f\colon \mathbb{C} \to \mathbb{C}$, the polynomial $f(z)$ is surjective. Let $f'(z)$ be the derivative of $f(z)$ and $C = \{z \in \mathbb{C} \mid f'(z) = 0\}$. Define $X = f(C)$ and $Y = f^{-1}(X)$. Both X and Y are finite sets, and f restricts to a map $F\colon \mathbb{C} \setminus Y \to \mathbb{C} \setminus X$. It is enough to prove that F is surjective. Since $C \setminus X$ is path-connected, it is enough to prove that $F\colon \mathbb{C} \setminus Y \to \mathbb{C} \setminus X$ is a covering space; indeed, since the image of F contains *some* point of $\mathbb{C} \setminus X$ the degree of this covering space must be postive. To do this, we check the hypotheses of Lemma 1.2.1:

- That $\mathbb{C} \setminus Y$ and $\mathbb{C} \setminus X$ are locally compact Hausdorff spaces is clear.
- Since $f(z)$ is a nonconstant polynomial, $f\colon \mathbb{C} \to \mathbb{C}$ is proper. Indeed, this follows from the fact that $\lim_{z \mapsto \infty} f(z) = \infty$. This implies that $F\colon \mathbb{C} \setminus Y \to \mathbb{C} \setminus X$ is proper.
- For $z \in \mathbb{C} \setminus Y$, since $f'(z) \neq 0$ the inverse function theorem implies that f and hence F is a local homeomorphism at z. □

1.2.3. More with polynomials. Here is another example of how Lemma 1.2.1 can be used.

EXAMPLE 1.2.3 (Roots of square-free polynomials). For some $n \geq 1$, let Poly_n be the space of degree-n monic polynomials over $\mathbb{C}$. Such an $f \in \text{Poly}_n$ can be written as

$$f(z) = z^n + a_1 z^{n-1} + a_2 z^{n-2} + \cdots + a_n \quad \text{with } a_1, \ldots, a_n \in \mathbb{C}.$$

The topology comes from the coefficients, so $\text{Poly}_n \cong \mathbb{C}^n$. By the fundamental theorem of algebra, such a polynomial has n roots (counted with multiplicity). Define

$$\text{RPoly}_n = \{(f, x) \in \text{Poly}_n \times \mathbb{C} \mid f(x) = 0\}.$$

In other words, RPoly_n is the space of polynomials equipped with a root. Let $p\colon \text{RPoly}_n \to \text{Poly}_n$ be the map $p(f, x) = f$. For $n \geq 2$ this is not a covering space since the fibers of p have different cardinalities. For example,

$$|p^{-1}(z^n)| = |\{(z^n, 0)\}| = 1 \quad \text{but} \quad |p^{-1}(z^n - 1)| = |\left\{(z^n - 1, \mu) \mid \mu \text{ an } n^{\text{th}} \text{ root of unity}\right\}| = n.$$

As suggested by this, the issue arises because of polynomials with repeated roots. Define

$$\text{Poly}_n^{\text{sf}} = \{f \in \text{Poly}_n \mid f \text{ has } n \text{ distinct roots}\}$$

and

$$\text{RPoly}_n^{\text{sf}} = \left\{(f, x) \in \text{Poly}_n^{\text{sf}} \times \mathbb{C} \mid f(x) = 0\right\}.$$

The "sf" stands for "square-free". The spaces $\text{Poly}_n^{\text{sf}}$ and $\text{RPoly}_n^{\text{sf}}$ are open subsets of Poly_n and RPoly_n, respectively.[4] The projection $p\colon \text{RPoly}_n^{\text{sf}} \to \text{Poly}_n^{\text{sf}}$ is a degree-n covering space. Indeed,

[4]This is an elementary exercise. A sophisticated way to see it is to use the fact that having a multiple root is equivalent to the vanishing of the discriminant, which is a polynomial in the coefficients of the polynomial.

since p is a proper map[5] whose fibers all have cardinality n, by Lemma 1.2.1 it is enough to prove that $p\colon \mathrm{RPoly}_n^{\mathrm{sf}} \to \mathrm{Poly}_n^{\mathrm{sf}}$ is a local homeomorphism. But this is easy: for $(f,x) \in \mathrm{RPoly}_n^{\mathrm{sf}}$, since $f(z)$ has no repeated roots we have $f'(x) \neq 0$, so by the implicit function theorem there is a neighborhood $U \subset \mathrm{Poly}_n$ of f such that around (f,x) the subspace

$$\mathrm{RPoly}_n^{\mathrm{sf}} \subset \mathrm{Poly}_n \times \mathbb{C} \subset \mathbb{C}^n \times \mathbb{C}$$

is the graph of a function $U \to \mathbb{C}$. □

1.3. Isomorphisms between covering spaces

We would like to classify the covers of a space X. To do this, we must first define what it means for two covers to be be the same, i.e., we must say what it means to have an isomorphism between two covers of X.

1.3.1. Isomorphisms.

The definition is as follows:

Definition 1.3.1. Let X be a space and let $p_1\colon \widetilde{X}_1 \to X$ and $p_2\colon \widetilde{X}_2 \to X$ be two covers of X. A *covering space isomorphism* from $\widetilde{X}_1$ to $\widetilde{X}_2$ is a homeomorphism $f\colon \widetilde{X}_1 \to \widetilde{X}_2$ such that the diagram

$$\begin{array}{ccc} \widetilde{X}_1 & \xrightarrow{\;f\;} & \widetilde{X}_2 \\ & {\scriptstyle p_1}\searrow \quad \swarrow{\scriptstyle p_2} & \\ & X & \end{array}$$

commutes, i.e., such that $p_2 \circ f = p_1$. If a covering space isomorphism from $\widetilde{X}_1$ to $\widetilde{X}_2$ exists, we say that $\widetilde{X}_1$ and $\widetilde{X}_2$ are *isomorphic* covers of X. This is clearly an equivalence relation. □

Remark 1.3.2. This can be rephrased using categorical language as follows. Recall that Top is the category of topological spaces and continuous maps. For a space X, let $\mathrm{Top}_{/X}$ be the category whose objects are spaces Y equipped with maps $\phi\colon Y \to X$ and whose morphisms from $\phi_1\colon Y_1 \to X$ to $\phi_2\colon Y_2 \to X$ are maps $f\colon Y_1 \to Y_2$ such that the diagram

$$\begin{array}{ccc} Y_1 & \xrightarrow{\;f\;} & Y_2 \\ & {\scriptstyle \phi_1}\searrow \quad \swarrow{\scriptstyle \phi_2} & \\ & X & \end{array}$$

commutes. A covering space $p\colon \widetilde{X} \to X$ is an object of $\mathrm{Top}_{/X}$, and a covering space isomorphism is an isomorphism in $\mathrm{Top}_{/X}$ between two covering spaces. □

1.3.2. Basic examples.

Here are two basic examples.

Example 1.3.3. For $\lambda \neq 0$, define $p^\lambda\colon \mathbb{R} \to \mathbb{S}^1$ via the formula $p_\lambda(\theta) = e^{2\pi i \lambda\theta}$. The universal cover of $\mathbb{S}^1$ is thus $p^1\colon \mathbb{R} \to \mathbb{S}^1$. Each $p^\lambda\colon \mathbb{R} \to \mathbb{S}^1$ is also a covering space, but is isomorphic to the universal cover. Indeed, letting $f\colon \mathbb{R} \to \mathbb{R}$ be the homeomorphism $f(\theta) = \lambda\theta$, the diagram

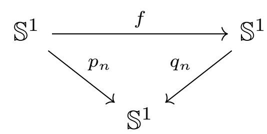
$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\;f\;} & \mathbb{R} \\ & {\scriptstyle p^\lambda}\searrow \quad \swarrow{\scriptstyle p^1} & \\ & \mathbb{S}^1 & \end{array}$$

commutes, so f is a covering space isomorphism from $p^\lambda\colon \mathbb{R} \to \mathbb{S}^1$ to $p^1\colon \mathbb{R} \to \mathbb{S}^1$. □

Example 1.3.4. For $n \geq 1$, let $p_n\colon \mathbb{S}^1 \to \mathbb{S}^1$ be the covering space defined by the formula $p_n(z) = z^n$ and let $q_n\colon \mathbb{S}^1 \to \mathbb{S}^1$ be the covering space defined by the formula $q_n(z) = z^{-n}$. The covers $p_n\colon \mathbb{S}^1 \to \mathbb{S}^1$ and $q_n\colon \mathbb{S}^1 \to \mathbb{S}^1$ are isomorphic. Indeed, letting $f\colon \mathbb{S}^1 \to \mathbb{S}^1$ be the homeomorphism $f(z) = z^{-1}$, the diagram

---

[5] To see that $p\colon \mathrm{RPoly}_n^{\mathrm{sf}} \to \mathrm{Poly}_n^{\mathrm{sf}}$ is proper, note first that the projection $q\colon \mathrm{Poly}_n \times \mathbb{C} \to \mathrm{Poly}_n$ is proper. This implies that the map $p\colon \mathrm{RPoly}_n \to \mathrm{Poly}_n$ is proper; indeed, for a compact $K \subset \mathrm{Poly}_n$ we have $p^{-1}(K) = q^{-1}(K) \cap \mathrm{RPoly}_n$, which is compact since it is a closed subset of the compact set $q^{-1}(K)$. This immediately implies that $p\colon \mathrm{RPoly}_n^{\mathrm{sf}} \to \mathrm{Poly}_n^{\mathrm{sf}}$ is proper.



commutes, so f is a covering space homomorphism from $p_n \colon \mathbb{S}^1 \to \mathbb{S}^1$ to $q_n \colon \mathbb{S}^1 \to \mathbb{S}^1$. □

1.3.3. Goal. Our of our main goals is to classify all the covers of a space up to isomorphism. Remarkably, for a reasonable space X there is a simple *algebraic* classification of covers of X. We will describe this classification later after we define the fundamental group of X (see Chapter YYY). As an example of what it says, here is one example:

EXAMPLE 1.3.5. For the circle $\mathbb{S}^1$, we have already seen the universal cover $p \colon \mathbb{R} \to \mathbb{S}^1$ and the covers $p_n \colon \mathbb{S}^1 \to \mathbb{S}^1$ defined by $p_n(z) = z^n$. Once we have the classification, we will see that every covering space $q \colon \widetilde{X} \to \mathbb{S}^1$ with $\widetilde{X}$ connected is isomorphic to either $p \colon \mathbb{R} \to \mathbb{S}^1$ or to some $p_n \colon \mathbb{S}^1 \to \mathbb{S}^1$. □

1.4. Deck transformations

We now study the automorphisms of a cover, i.e., the isomorphisms from the cover to itself.

1.4.1. Deck group. These automorphisms are called deck transformations:

DEFINITION 1.4.1. Let $p \colon \widetilde{X} \to X$ be a covering space. A *deck transformation* of $p \colon \widetilde{X} \to X$. is a covering space isomorphism $f \colon \widetilde{X} \to \widetilde{X}$. These form a group under composition called the *deck group* of $p \colon \widetilde{X} \to X$, denoted $\mathrm{Deck}(p \colon \widetilde{X} \to X)$ or simply $\mathrm{Deck}(\widetilde{X})$. □

Here is an example:

EXAMPLE 1.4.2. Let $p \colon \mathbb{R} \to \mathbb{S}^1$ be the universal cover, so $p(\theta) = e^{2\pi i \theta}$ for all $\theta \in R$. For each $n \in \mathbb{Z}$, we can define a deck transformation $f_n \colon \mathbb{R} \to \mathbb{R}$ via the formula $f_n(\theta) = \theta + n$. □

1.4.2. Determined by one point. The following lemma says that in favorable situations deck transformations are completely determined by what they do to a single point.

LEMMA 1.4.3. *Let $p \colon \widetilde{X} \to X$ be a covering space with $\widetilde{X}$ connected. Let $f, g \colon \widetilde{X} \to \widetilde{X}$ be two deck transformations such that there exists some $z_0 \in \widetilde{X}$ with $f(z_0) = g(z_0)$. Then $f = g$.*

PROOF. Let $E = \left\{ z \in \widetilde{X} \mid f(z) = g(z) \right\}$. Our goal is to prove that $E = \widetilde{X}$. By assumption $z_0 \in E$, so since $\widetilde{X}$ is connected it is enough to prove that E is both open and closed.[6] Consider $z \in \widetilde{X}$. We must prove that if $z \in E$ (resp. $z \notin E$) then there is an open neighborhood of z contained in E (resp. disjoint from E). Let U be a trivialized neighborhood of $p(z)$.

Assume first that $z \in E$. Let $\widetilde{U}$ be the sheet above U containing $f(z) = g(z)$. Set $V = f^{-1}(\widetilde{U}) \cap g^{-1}(\widetilde{U})$, so V is an open neighborhood of z with $f(V), g(V) \subset \widetilde{U}$. For $z' \in V$, both $f(z')$ and $g(z')$ are the unique point of $\widetilde{U}$ projecting to $p(z') \in U$, so in particular $f(z') = g(z')$. This implies that $V \subset E$, as desired.

Assume now that $z \notin E$, so $f(z) \neq g(z)$. Let $\widetilde{U}_1$ and $\widetilde{U}_2$ be the sheets above U with $f(z) \in \widetilde{U}_1$ and $g(z) \in \widetilde{U}_2$. Since $f(z) \neq g(z)$, the sheets $\widetilde{U}_1$ and $\widetilde{U}_2$ are distinct and hence disjoint. Set $W = f^{-1}(\widetilde{U}_1) \cap g^{-1}(\widetilde{U}_2)$, so W is an open neighborhood of z with $f(W) \subset \widetilde{U}_1$ and $g(W) \subset \widetilde{U}_2$. Since $\widetilde{U}_1 \cap \widetilde{U}_2 = \emptyset$, this implies that $f(z') \neq g(z')$ for all $z' \in W$, so W is disjoint from E, as desired. □

[6] Note that if X is Hausdorff (like most spaces in this book) it is automatic that E is closed; indeed, if X is Hausdorff then for any continuous maps $f, g \colon Y \to X$ the set of $y \in Y$ with $f(y) = g(y)$ is closed.

1.4.3. Determining the deck group. The following example shows a typical way to determine the deck group of a covering space:

EXAMPLE 1.4.4 (Universal cover of circle). Let $p\colon \mathbb{R} \to \mathbb{S}^1$ be the universal cover of $\mathbb{S}^1$. For $n \in \mathbb{Z}$, let $f_n\colon \mathbb{R} \to \mathbb{R}$ be the deck transformation defined by the formula $f_n(z) = z + n$. We claim that

$$\text{Deck}(p\colon \mathbb{R} \to \mathbb{S}^1) = \{f_n \mid n \in \mathbb{Z}\} \cong \mathbb{Z}.$$

To see this, consider an arbitrary deck transformation $f\colon \mathbb{R} \to \mathbb{R}$. Since $p(f(0)) = p(0)$, we must have $f(0) = n$ for some $n \in \mathbb{Z}$. It follows that $f_{-n} \circ f(0) = 0$, so by Lemma 1.4.3 we have $f_{-n} \circ f = \mathbb{1}_{\mathbb{R}}$ and hence $f = f_n$. □

1.5. Regular covers

Roughly speaking, a regular cover[7] is a cover with a deck group that is as large as possible.

1.5.1. Definition of regular cover. Let $p\colon \widetilde{X} \to X$. The group $\text{Deck}(\widetilde{X})$ acts on $\widetilde{X}$. For $x \in X$, the action of $\text{Deck}(\widetilde{X})$ on $\widetilde{X}$ preserves the fiber $f^{-1}(x)$, so $\text{Deck}(\widetilde{X})$ acts on $f^{-1}(x)$. For $z_1, z_2 \in f^{-1}(x)$, Lemma 1.4.3 implies that if $\widetilde{X}$ is connected then there exists at most one $f \in \text{Deck}(\widetilde{X})$ with $f(z_1) = z_2$. A regular cover is a cover where such an f always exists:

DEFINITION 1.5.1. A *regular cover* is a cover $p\colon \widetilde{X} \to X$ such that for all $x \in X$, the group $\text{Deck}(\widetilde{X})$ acts transitively on $f^{-1}(x)$. A cover that is not regular is *irregular*. □

1.5.2. Examples. The calculation in Example 1.4.4 shows that the universal cover $p\colon \mathbb{R} \to \mathbb{S}^1$ is regular. In fact, most of the covers we have seen so far are regular:

EXAMPLE 1.5.2 (Trivial cover). Let X be a space and $\mathcal{I}$ be a discrete set. Consider the trivial cover $p\colon X \times \mathcal{I} \to X$. Set $G = \text{Deck}(p\colon X \times \mathcal{I} \to X)$. For each bijection $\sigma\colon \mathcal{I} \to \mathcal{I}$, we can define an element $f_\sigma \in G$ via the formula $f_\sigma(x, i) = (x, \sigma(i))$. These elements act transitively on the fibers, so $p\colon X \times \mathcal{I} \to X$ is regular. One can check that all elements of G are of the form f_σ if X is connected (see Exercise 1.10). □

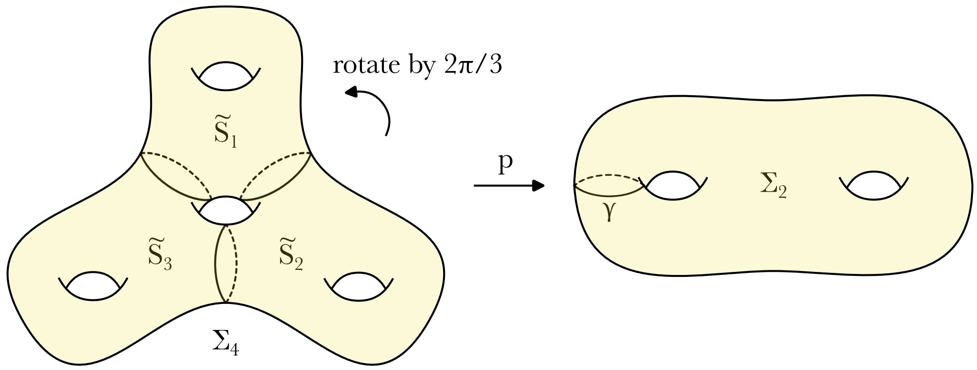
EXAMPLE 1.5.3 (Degree n cover of circle). Let $p_n\colon \mathbb{S}^1 \to \mathbb{S}^1$ be the cover defined by the formula $p_n(z) = z^n$. We claim that $p_n\colon \mathbb{S}^1 \to \mathbb{S}^1$ is a regular cover with deck group isomorphic to the cyclic group C_n of order n. Indeed, let $G = \text{Deck}(p_n\colon \mathbb{S}^1 \to \mathbb{S}^1)$. Let $f \in G$ be the map $f\colon \mathbb{S}^1 \to \mathbb{S}^1$ defined by the formula $f(z) = e^{2\pi i/n} z$. The element f has order n and its powers act transitively on the fiber $p_n^{-1}(1)$, which equals the n^{th} roots of unity. This implies that the cover is regular, and also by Lemma 1.4.3 that G is the cyclic group of order n generated by f. □

EXAMPLE 1.5.4 (Cosets of discrete subgroups). Let $\mathbf{G}$ be a topological group and let $\mathbf{H} < \mathbf{G}$ be a discrete subgroup. Then the projection $p\colon \mathbf{G} \to \mathbf{G}/\mathbf{H}$ is a regular cover. Indeed, for $h \in \mathbf{H}$ define $f_h\colon \mathbf{G} \to \mathbf{G}$ via the formula $f_h(g) = gh$. Then $f_h \in \text{Deck}(p\colon \mathbf{G} \to \mathbf{G}/\mathbf{H})$, and the f_h act transitively on the fibers of $p\colon \mathbf{G} \to \mathbf{G}/\mathbf{H}$. If $\mathbf{G}$ is connected, then by Lemma 1.4.3 this is the entire deck group, so $\text{Deck}(p\colon \mathbf{G} \to \mathbf{G}/\mathbf{H}) \cong \mathbf{H}$. As a special case, the deck group of the cover $p\colon \mathbb{R}^n \to \mathbb{T}^n$ is the group $\mathbb{Z}^n$, which acts on $\mathbb{R}^n$ by translations. □

EXAMPLE 1.5.5 (Real projective space). The cover $p\colon \mathbb{S}^n \to \mathbb{RP}^n$ is regular. Indeed, the map $f\colon \mathbb{S}^n \to \mathbb{S}^n$ defined by $f(z) = -z$ is an element of the deck group that swaps the two elements in the fiber over any point of $\mathbb{RP}^n$. By Lemma 1.4.3, the deck group of $p\colon \mathbb{S}^n \to \mathbb{RP}^n$ is the cyclic group C_2 of order 2 generated by f. □

EXAMPLE 1.5.6 (Cover of surface). Consider the covering space $p\colon \Sigma_4 \to \Sigma_2$ from Example 1.1.12. Set $G = \text{Deck}(p\colon \Sigma_4 \to \Sigma_2)$. There is a deck transformation $f \in G$ that rotates Σ_4 by $2\pi/3$ as follows:

[7]These are also often called *normal covers*.



The element f has order 3 and its powers act transitively on all the fibers. This implies that the cover is regular, and also by Lemma 1.4.3 that G is the cyclic group of order 3 generated by f. □

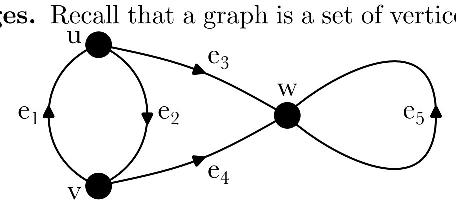
EXAMPLE 1.5.7 (Roots of square-free polynomials). The degree-n covering space $p\colon \text{RPoly}_n^{\text{sf}} \to \text{Poly}_n^{\text{sf}}$ discussed in Example 1.2.3 is regular for $n = 2$ (see Exercise 1.9), but is irregular for $n \geq 3$. We do not have the technology to prove this yet (see §YYY for the proof), but it should not be surprising. Indeed, if it was regular then the deck group G would act simply transitively on the roots of every degree-n polynomial with distinct roots, and if such a canonical group action existed then we would surely teach about it in elementary abstract algebra classes.[8] □

We will meet more examples of irregular covers in the next section when we discuss covers of graphs.

1.6. Graphs

Graphs[9] provide a rich source of examples of covering spaces.

1.6.1. Vertices and edges. Recall that a graph is a set of vertices connected by oriented edges:



We formalize this as follows.[10] A *graph* is a topological space X constructed as follows:

- Start with a discrete set $V = X^{(0)}$, called the *vertices* or the 0-*simplices*.
- Let $\{\mathbb{D}^1_e\}_{e \in E}$ be a set of copies of the interval $\mathbb{D}^1 = [-1, 1]$ indexed by a set E. We call these the *edges* or the 1-*simplices*. For $e \in E$, we have a map f_e from $\partial \mathbb{D}^1_e = \{-1, 1\}$ to $X^{(0)}$ called the *attaching map*. We call $f_e(-1)$ the *initial vertex* of e and $f_e(1)$ the *terminal vertex* of e, and we say that e *connects* $f_e(-1)$ to $f_e(1)$.
- The space $X = X^{(1)}$ is formed by attaching the $\mathbb{D}^1_e$ to $X^{(0)}$ using the f_e. Formally,

$$X = X^{(1)} = \left(X^{(0)} \sqcup \bigsqcup_{e \in E} \mathbb{D}^1_e \right) / \sim,$$

  where for $e \in E$ the equivalence relation $\sim$ identifies $x \in \partial \mathbb{D}^1_e$ with $f_e(x) \in X^{(0)}$.

For $e \in E$, there is a map $F_e\colon \mathbb{D}^1_e \to X$ taking $\mathbb{D}^1_e$ to $\mathbb{D}^1_e / \sim$ called the *characteristic map*. We will identify e with the image of the characteristic map, so depending on whether the attaching map is injective or not e is homeomorphic to either $\mathbb{D}^1 = [-1, 1]$ or to

$$\mathbb{D}^1 / (-1 \sim 1) \cong \mathbb{S}^1.$$

[8] For $n = 2$, this group action just exchanges the two roots.

[9] Our conventions about graphs are that unless otherwise specified they are oriented and we allow multiple edges and loops.

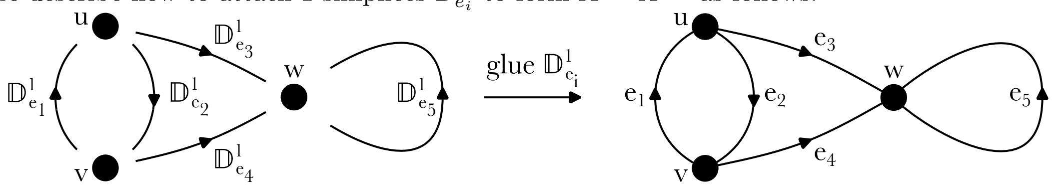
[10] When the reader learns about CW complexes (see Chapter YYY), they will see that these are exactly the 1-dimensional CW complexes.

The space X is given the weak topology with respect to the characteristic maps, so a set $U \subset X$ is open if and only if $F_e^{-1}(U) \subset \mathbb{D}_e^1$ is open for all $e \in E$. The orientations on the $\mathbb{D}_e^1 = [-1, 1]$ give orientations on our edges. To help the reader absorb this, we spell it out in the above example:

EXAMPLE 1.6.1. Let $V = X^{(0)}$ be the discrete set $\{u, v, w\}$ and let $E = \{e_1, \ldots, e_5\}$. Define the attaching maps via the formulas

$$\begin{array}{ccccc} f_{e_1}(-1) = v & f_{e_2}(-1) = u & f_{e_3}(-1) = u & f_{e_4}(-1) = v & f_{e_5}(-1) = w \\ f_{e_1}(1) = u & f_{e_2}(1) = v & f_{e_3}(1) = w & f_{e_4}(1) = w & f_{e_5}(1) = w \end{array}$$

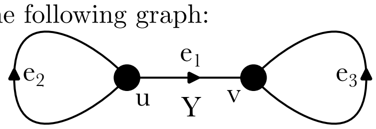
These describe how to attach 1-simplices $\mathbb{D}_{e_i}^1$ to form $X = X^{(1)}$ as follows:



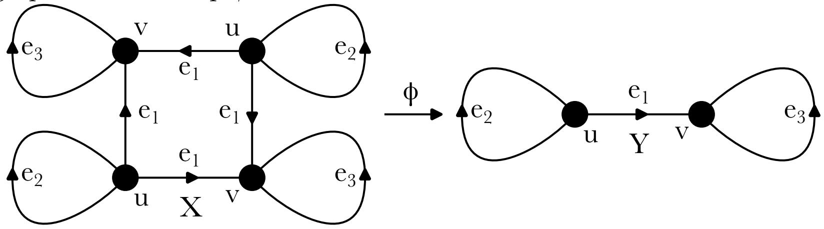
Note that the characteristic maps $F_{e_i} : \mathbb{D}_{e_i}^1 \to X$ are injective for $1 \le i \le 4$, but F_{e_5} is not injective since it identifies the two endpoints of $\mathbb{D}_{e_5}^1$ to the single point w. □

1.6.2. Maps of graphs. Let X and Y be graphs. To define a continuous map $\phi\colon X \to Y$, we must specify where ϕ sends each vertex and edge. This is particularly easy to do if we require our map to take vertices to vertices and oriented edges to oriented edges, which will suffice for the examples in this section. This is best explained by an example:

EXAMPLE 1.6.2. Let Y be the following graph:



We specify a graph X and a map $\phi\colon X \to Y$ as follows:



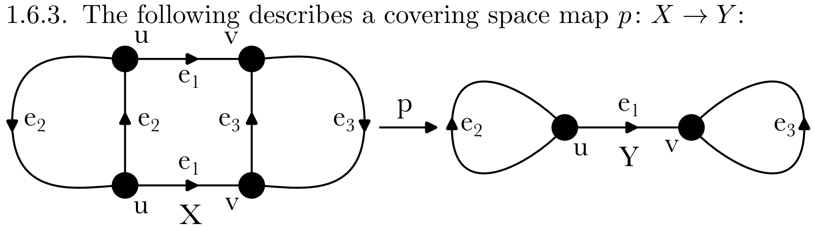
Here we label the vertices and oriented edges of X by the vertices and oriented edges they map to. What this map does is map each edge of the central square of X to the single non-loop edge of Y and map each loop in X to the appropriate loop in Y. On the interiors of edges, the map ϕ respects the evident linear structure coming from the construction of the graph. In other words, if e is an edge of X and $\phi(e)$ is the corresponding edge of Y, then letting $F_e\colon \mathbb{D}_e^1 \to X$ and $G_{\phi(e)}\colon \mathbb{D}_{f(e)}^1 \to Y$ be the characteristic maps the diagram

$$\begin{array}{ccccc} \mathbb{D}_e^1 & = & [-1,1] & = & \mathbb{D}_{f(e)}^1 \\ \downarrow{\scriptstyle F_e} & & & & \downarrow{\scriptstyle G_{f(e)}} \\ X & & \xrightarrow{\phi} & & Y \end{array}$$

commutes. □

1.6.3. Covers of graphs. The above example is not a covering map. The problem is that it is not a local homeomorphism at the vertices. What is needed for a covering map is informally that for each vertex "the same edges enter and exist as in the target". Here is an example of a covering map with the same Y as above but a different X:

EXAMPLE 1.6.3. The following describes a covering space map $p\colon X \to Y$:

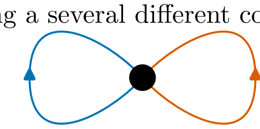


This is a covering space map since:

- for both vertices of X mapping to u, one edge exits mapping to e_1, one edge exits mapping to e_2, and one edge enters mapping to e_2; and
- for both vertices of X mapping to v, one edge enters mapping to e_1, one edge exits mapping to e_3, and one edge enters mapping to e_3; and

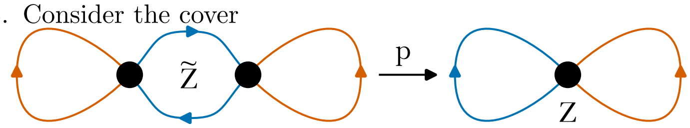
This is a regular cover with deck group isomorphic to C_2. The generator of C_2 acts on X by the involution that swaps the two vertices labeled u, the two vertices labeled v, and for $i = 1, 2, 3$ the two oriented edges labeled e_i. □

We conclude this section by giving a several different covers of the following graph Z:



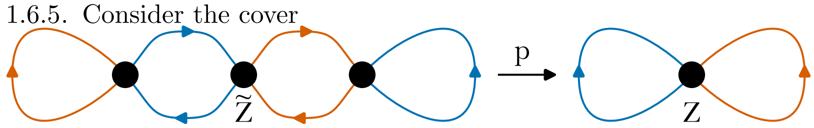
Since Z has only one vertex, there is no need to give it a name since all vertices of a cover map to that one vertex. We also use colors rather than letters to distinguish the two edges of Z, and label the edges in the domain of our covering space maps by coloring them with the appropriate colors.

EXAMPLE 1.6.4. Consider the cover



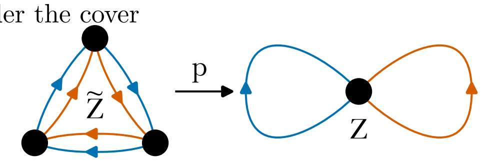
This is a degree 2 regular cover. The deck group is isomorphic to C_2, and acts on $\widetilde{Z}$ by the involution that swaps the two vertices, the two orange loops, and the two blue edges. □

EXAMPLE 1.6.5. Consider the cover



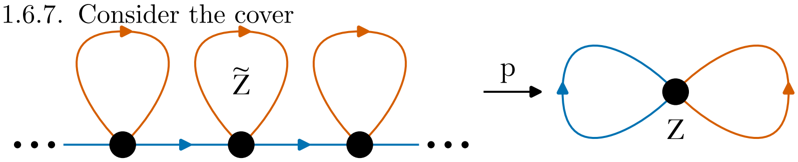
This is an irregular cover. To see this, note that a homeomorphism $f\colon \widetilde{Z} \to \widetilde{Z}$ must take vertices to vertices since the vertices are the only $z \in \widetilde{Z}$ such that $\widetilde{Z} \setminus z$ is disconnected.[11] One can then check that other than the identity there is no such homeomorphism that also preserves the orientations and colors of the edges, so the deck group is trivial. □

EXAMPLE 1.6.6. Consider the cover



This ia a degree 3 regular cover. The deck group is C_3, which acts on $\widetilde{Z}$ by rotations. □

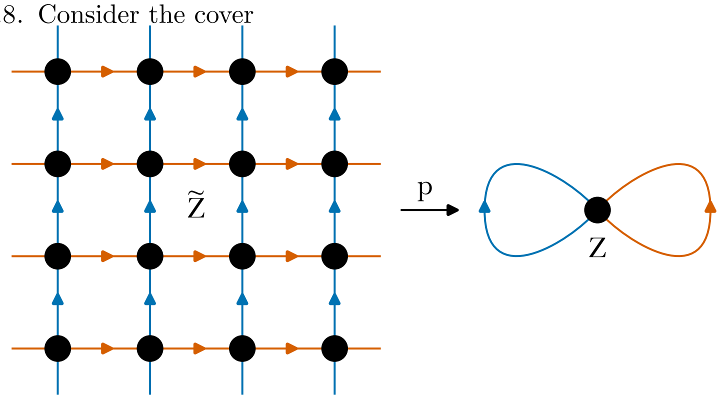
EXAMPLE 1.6.7. Consider the cover



[11] In fact, if X is a graph with no valence 2 vertices, then any homeomorphism $X \to X$ must take vertices to vertices and edges to edges. This can be proved directly with a certain amount of pain, but we will wait to prove it until we have introduced homology, which makes the proof easy.

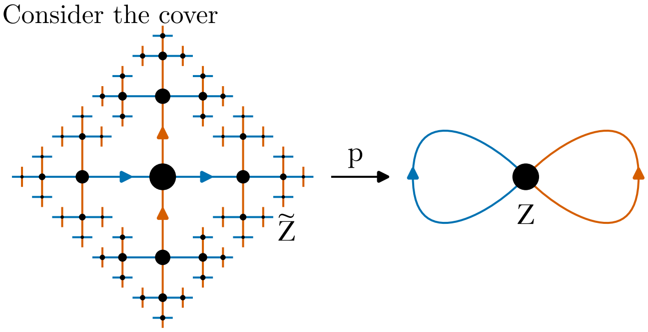
This is an infinite degree regular cover. The deck group is isomorphic to $\mathbb{Z}$, which acts on $\widetilde{Z}$ as translations. □

EXAMPLE 1.6.8. Consider the cover



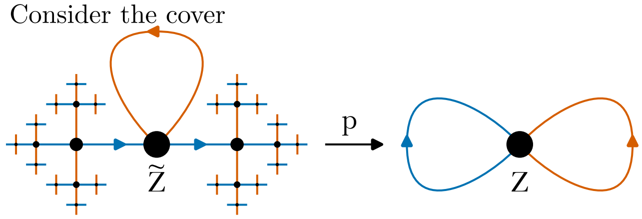
Here $\widetilde{Z}$ is the graph embedded in $\mathbb{R}^2$ whose vertices are at $\mathbb{Z}^2$ and whose edges are horizontal and vertical lines. This is an infinite degree regular cover. The deck group is isomorphic to $\mathbb{Z}^2$, which acts on $\widetilde{Z} \subset \mathbb{R}^2$ via the action of $\mathbb{Z}^2$ on $\mathbb{R}^2$ by integer translations. □

EXAMPLE 1.6.9. Consider the cover



The indicated pattern in the domain repeats infinitely often, making it an infinite 4-valent[12] tree.[13] The horizontal edges are oriented going right, and the vertical edges are oriented going up. This is an infinite degree regular cover (we leave this as Exercise 1.11). □

EXAMPLE 1.6.10. Consider the cover



This is an infinite degree irregular cover. To see this, note that any homeomorphism $f\colon \widetilde{Z} \to \widetilde{Z}$ must take the interior of the orange loop to itself since the interior of the orange loop equals the set of all $p \in \widetilde{Z}$ such that $\widetilde{Z} \setminus p$ is connected. This implies that it must fix the vertex at which the orange loop is based. Since $\widetilde{Z}$ is connected, the identity is the only element of the deck group that fixes a vertex, so we conclude that the deck group is trivial. □

1.7. Covering space actions

Let $p\colon \widetilde{X} \to X$ be a cover with deck group G. The group G acts on $\widetilde{X}$. If $\widetilde{X}$ is connected, then Lemma 1.4.3 says that action is *free*, i.e., that for all $z \in \widetilde{X}$ the stabilizer subgroup G_z is trivial. In fact, even more is true, and this section is devoted to studying this action.

[12]This means that the valence of each vertex is 4. The valence of a vertex of a graph is the number of edges entering/exiting it. If there is a loop based at a vertex, then it counts for two edges in the valence, one going in and one going out.

[13]A tree is a nonempty graph with no cycles, that is, no embedded circles.

Remark 1.7.1. In this book, a group action is always assumed to preserve any structure a set has. In particular, an action of a group G on a topological space Z is assumed to be continuous. In other words, for all $g \in G$ the map $Z \to Z$ that multiplies points by g is assumed to be a homeomorphism. □

1.7.1. Covering space actions. The following lemma isolates the key property of the action of the deck group of a connected cover:

Lemma 1.7.2. *Let $p\colon \widetilde{X} \to X$ be a covering space with $\widetilde{X}$ connected, let $G = \mathrm{Deck}(p\colon \widetilde{X} \to X)$, and let $z \in \widetilde{X}$. Then there is an open neighborhood V of z whose translates $\{g{\cdot}V \mid g \in G\}$ are all disjoint.*

Proof. Let U be a trivialized neighborhood of $p(z)$ and let $\widetilde{U}$ be the sheet lying above U with $z \in \widetilde{U}$. We claim that $V = \widetilde{U}$ has the indicated property. Indeed, let $g_1, g_2 \in G$ satisfy $(g_1{\cdot}\widetilde{U}) \cap (g_2{\cdot}\widetilde{U}) \neq \emptyset$. We must prove that $g_1 = g_2$. Pick $z_1, z_2 \in \widetilde{U}$ with $g_1{\cdot}z_1 = g_2{\cdot}z_2$. Since the action of G preserves the fibers of $p\colon \widetilde{X} \to X$, the points $z_1, z_2 \in \widetilde{U}$ must lie in the same fiber. Since the restriction of $p\colon \widetilde{X} \to X$ to $\widetilde{U}$ is injective, this implies that $z_1 = z_2$. Letting $w = z_1 = z_2$ be this common value, we have $g_1{\cdot}w = g_2{\cdot}w$. Lemma 1.4.3 now implies that $g_1 = g_2$, as desired. □

Actions satisfying the conclusions of this lemma are important, so we give them a special name:

Definition 1.7.3. A *covering space action* is an action of a group G on a space Z such that for all $z \in Z$, there exists an open neighborhood V of z such that the translates $\{g{\cdot}V \mid g \in G\}$ are all disjoint. □

Remark 1.7.4. All covering space actions are free. If G is finite and Z is Hausdorff, then the converse is true: all free action of G on Z are covering space actions (see Exercise 1.8). □

1.7.2. Covers from quotients. Let G be a group and let Z be a space equipped with a left action of G. Endow the quotient[14] X/G with the quotient topology. In other words, if $q\colon Z \to Z/G$ is the projection then a set $U \subset Z/G$ is open if and only if $q^{-1}(U)$ is open. If the action of G on Z is a covering space action, then the quotient map $q\colon Z \to Z/G$ is a regular covering space:

Lemma 1.7.5. *Let G be a group acting a space Z by a covering space action. Then quotient map $q\colon Z \to Z/G$ is a regular covering space. Moreover, if Z is connected then $G = \mathrm{Deck}(q\colon Z \to Z/G)$.*

Proof. Consider $x \in Z/G$. Write $x = q(z)$ with $z \in Z$. Let V be an open neighborhood of z such that the sets in the G-orbit of V are disjoint. Set $U = q(V)$. We have

$$q^{-1}(U) = \bigcup_{g \in G} g{\cdot}V.$$

Since each $g{\cdot}V$ is open, it follows that $q^{-1}(U)$ is open and thus by definition U is open. The $g{\cdot}V$ are disjoint open subsets of Z and each projects homeomorphically onto U. We conclude that U is a trivialized neighborhood of x and the $g{\cdot}V$ are the sheets lying above U. This implies that $q\colon Z \to Z/G$ is a covering space. By construction, the action of G on Z is by deck transformations of $q\colon Z \to Z/G$, so $G < \mathrm{Deck}(q\colon Z \to Z/G)$. This action is transitive on fibers, so if Z is connected then Lemma 1.4.3 implies that $G = \mathrm{Deck}(q\colon Z \to Z/G)$. □

1.7.3. Examples. All of our examples of regular covering spaces could have been constructed using Lemma 1.7.2. For instance, C_2 acts on $\mathbb{S}^n$ via the antipodal map $z \mapsto -z$. This is a free action, so since C_2 is finite it is a covering space action (c.f. Remark 1.7.4). We could have defined $\mathbb{RP}^n = \mathbb{S}^n/C_2$ and identified the covering space $p\colon \mathbb{S}^n \to \mathbb{RP}^n$ with the quotient projection. This would be a little artificial, but here is an example where this point of view is essential:

[14]This is potentially confusing notation since G is acting on the left. A purist would insist that Z/G is the quotient of Z by an action of G on the right, and denote the quotient of Z by an action of G on the left by $G\backslash Z$. However, our notation is common and traditional, and we will follow it. There will be a few situations where we will have both left and right actions, and we will work hard to be clear about what our notation means in those cases.

EXAMPLE 1.7.6 (Configuration space). Let X be any space. The *ordered configuration space* of n points on X is the space[15]

$$\mathrm{PConf}_n(X) = \left\{(x_1, \ldots, x_n) \in X^{\times n} \mid x_i \neq x_j \text{ for all distinct } 1 \leq i, j \leq n\right\}.$$

Topologize this as a subset of $X^{\times n}$. The symmetric group $\mathfrak{S}_n$ on n letters acts on $\mathrm{Conf}_n(X)$ via the formula

$$\sigma \cdot (x_1, \ldots, x_n) = (x_{\sigma^{-1}(1)}, \ldots, x_{\sigma^{-1}(n)}) \quad \text{for } \sigma \in \mathfrak{S}_n \text{ and } (x_1, \ldots, x_n) \in \mathrm{PConf}_n(X).$$

The inverses are there to make this a left action.[16] This is a free action since the x_i are all distinct, and thus since $\mathfrak{S}_n$ is finite it is a covering space action. The *configuration space* of n points on X is the quotient $\mathrm{Conf}_n(X) = \mathrm{PConf}_n(X)/\mathfrak{S}_n$. Points of $\mathrm{Conf}_n(X)$ can be viewed as unordered sets $\{x_1, \ldots, x_n\}$ of n distinct points in X. The projection $p\colon \mathrm{PConf}_n(X) \to \mathrm{Conf}_n(X)$ is a regular covering space. □

1.8. Exercises

EXERCISE 1.1. Carefully prove that the following are covering spaces. Let $\mathbb{C}^\times = \mathbb{C} \setminus \{0\}$.

(a) The map $p\colon \mathbb{C} \to \mathbb{C}^\times$ defined by $p(z) = e^z$.
(b) For $n \in \mathbb{Z} \setminus \{0\}$, the map $p\colon \mathbb{C}^\times \to \mathbb{C}^\times$ defined by $p(z) = z^n$. □

EXERCISE 1.2. Prove that the map $p\colon \mathbb{C} \to \mathbb{C}$ defined by $p(z) = z^2$ is not a covering space. □

EXERCISE 1.3. Let $p_1\colon \widetilde{X}_1 \to X_1$ and $p_2\colon \widetilde{X}_2 \to X_2$ be covering spaces. Define $q\colon \widetilde{X}_1 \sqcup \widetilde{X}_2 \to X_1 \sqcup X_2$ via the formula

$$q(z) = \begin{cases} p_1(z) & \text{if } z \in \widetilde{X}_1, \\ p_2(z) \text{if } z \in \widetilde{X}_2. \end{cases}$$

Prove that $q\colon \widetilde{X}_1 \sqcup \widetilde{X}_2 \to X_1 \sqcup X_2$ is a cover space. Use this construction to find a covering space over a non-connected base that does not have a degree. □

EXERCISE 1.4. Let $p_1\colon \widetilde{X}_1 \to X_1$ and $p_2\colon \widetilde{X}_2 \to X_2$ be covering spaces. Define $q\colon \widetilde{X}_1 \times \widetilde{X}_2 \to X_1 \times X_2$ via the formula $q(z_1, z_2) = (p_1(z_1), p_2(z_2))$. Prove that $q\colon \widetilde{X}_1 \times \widetilde{X}_2 \to X_1 \times X_2$ is a covering space. □

EXERCISE 1.5. Prove the following:

(a) Let $p\colon \widetilde{X} \to X$ be a cover and let $X' \subset X$ be a subspace. Define $\widetilde{X}' = f^{-1}(X')$ and $p' = p|_{\widetilde{X}'}$. Prove that $p'\colon \widetilde{X}' \to X'$ is a covering space. We will call this the *restriction* of p to X'.
(b) Let X be a locally connected space with connected components $\{X_j\}_{j \in J}$. For each $j \in J$, let $q_j\colon Y_j \to X_j$ be a covering space. Define

$$Y = \bigsqcup_{j \in J} Y_j,$$

and let $q\colon Y \to X$ be the map that for $j \in J$ and $y \in Y_j$ satisfies $q(y) = q_j(y) \in Y_j \subset Y$. Prove that $q\colon Y \to X$ is a covering space. □
(c) Construct a counterexample to part (b) in the case where X is not locally connected.

EXERCISE 1.6. Let $p\colon \widetilde{X} \to X$ be a covering space such that $p^{-1}(x)$ is finite and nonempty for all $x \in X$. Prove that X is compact Hausdorff if and only if $\widetilde{X}$ is compact Hausdorff. □

EXERCISE 1.7. Let $\alpha \in \mathbb{R}$ be an irrational number. Let $G \cong \mathbb{Z}$ be an infinite cyclic group generated by $s \in G$. Let G act on $\mathbb{S}^1$ via the formula

$$t \cdot z = e^{2\pi i \alpha} z \quad \text{for } z \in \mathbb{S}^1.$$

Let $p\colon \mathbb{S}^1 \to \mathbb{S}^1/G$ be the quotient map.

(a) Prove directly that p is not a covering space.

[15]This is sometimes also called the *pure configuration space*, which is why it is written $\mathrm{PConf}_n(X)$.
[16]This is the same reason that inverses appear in the action of $\mathrm{GL}(V)$ on the dual of a vector space V.

(b) Prove that $\mathbb{S}^1/G$ is not Hausdorff. □

EXERCISE 1.8. Let X be a Hausdorff space and let G be a finite group acting freely on X. Prove that the action of G on X is a covering space action. □

EXERCISE 1.9. Let $p\colon \widetilde{X} \to X$ be a degree 2 cover. Prove that $\widetilde{X}$ is a regular cover. □

EXERCISE 1.10. Let X be a space and $\mathcal{I}$ be a discrete set, and let $p\colon X \times \mathcal{I} \to X$ be the trivial cover.

(a) If X is connected, prove that all elements of the deck group of $p\colon X \times \mathcal{I} \to X$ are of the form $f_\sigma(x, i) = (x, \sigma(i))$ for some bijection $\sigma\colon \mathcal{I} \to \mathcal{I}$.
(b) If X is not connected, construct elements of the deck group that are not of this form. □

EXERCISE 1.11. Verify that the cover in Example 1.6.9 is regular. □

CHAPTER 2

Lifting paths and homotopies

This chapter studies lifting problems, which play a key role in both the classification of covering spaces and their applications. As an application, we develop the theory of winding numbers and degrees of maps from $\mathbb{S}^1$ to iteself. More applications are in Chapter 3.

2.1. Lifting problems in general

Let $p\colon \widetilde{X} \to X$ be a covering space and let $f\colon Y \to X$ be a map. A *lift* of f through p is a map $\widetilde{f}\colon Y \to \widetilde{X}$ such that the diagram

$$\begin{array}{ccc} & & \widetilde{X} \\ & \overset{\widetilde{f}}{\nearrow} & \big\downarrow p \\ Y & \xrightarrow{f} & X \end{array}$$

commutes, i.e., such that $f = p \circ \widetilde{f}$.

Example 2.1.1. A deck transformation $\widetilde{f}\colon \widetilde{X} \to \widetilde{X}$ is a lift of the covering space map $p\colon \widetilde{X} \to X$ itself:

$$\begin{array}{ccc} & & \widetilde{X} \\ & \overset{\widetilde{f}}{\nearrow} & \big\downarrow p \\ \widetilde{X} & \xrightarrow{p} & X \end{array}$$

Of course, it is possible that a lift of $p\colon \widetilde{X} \to X$ to a map $\widetilde{f}\colon \widetilde{X} \to \widetilde{X}$ exists such that $\widetilde{f}$ is not a homeomorphism, so not all such lifts are deck transformations. □

A lift might or might not exist. However, just like a deck transformation if a lift exists then under favorable hypotheses it is determined by what it does to a single point:

Lemma 2.1.2. *Let $p\colon \widetilde{X} \to X$ be a covering space and let $f\colon Y \to X$ be a map. Assume that Y is connected. Let $\widetilde{f}_1, \widetilde{f}_2\colon Y \to \widetilde{X}$ be two lifts of f through p such that there is some $y_0 \in Y$ with $\widetilde{f}_1(y_0) = \widetilde{f}_2(y_0)$. Then $\widetilde{f}_1 = \widetilde{f}_2$.*

Proof. The proof is identical to that of Lemma 1.4.3, but for the reader's convenience we repeat the argument. Let $E = \left\{ y \in y \mid \widetilde{f}_1(y) = \widetilde{f}_2(y) \right\}$. Our goal is to prove that $E = Y$. By assumption $y_0 \in E$, so since Y is connected it is enough to prove that E is both open and closed. Consider $y \in Y$. We must prove that if $y \in E$ (resp. $y \notin E$) then there is an open neighborhood of y contained in E (resp. disjoint from E). Let U be a trivialized neighborhood of $f(y)$.

Assume first that $y \in E$. Let $\widetilde{U}$ be the sheet above U containing $\widetilde{f}_1(y) = \widetilde{f}_2(y)$. Set $V = \widetilde{f}_1^{-1}(\widetilde{U}) \cap \widetilde{f}_2^{-1}(\widetilde{U})$, so V is an open neighborhood of y with $\widetilde{f}_1(V), \widetilde{f}_2(V) \subset \widetilde{U}$. For $y' \in V$, both $\widetilde{f}_1(y')$ and $\widetilde{f}_2(y')$ are the unique point of $\widetilde{U}$ projecting to $f(y') \in U$, so in particular $\widetilde{f}_1(y') = \widetilde{f}_2(y')$. This implies that $V \subset E$, as desired.

Assume now that $y \notin E$, so $\widetilde{f}_1(y) \neq \widetilde{f}_2(y)$. Let $\widetilde{U}_1$ and $\widetilde{U}_2$ be the sheets above U with $\widetilde{f}_1(y) \in \widetilde{U}_1$ and $\widetilde{f}_2(y) \in \widetilde{U}_2$. Since $\widetilde{f}_1(y) \neq \widetilde{f}_2(y)$, the sheets $\widetilde{U}_1$ and $\widetilde{U}_2$ are distinct and hence disjoint. Set $W = \widetilde{f}_1^{-1}(\widetilde{U}_1) \cap \widetilde{f}_2^{-1}(\widetilde{U}_2)$, so W is an open neighborhood of y with $\widetilde{f}_1(W) \subset \widetilde{U}_1$ and $\widetilde{f}_2(W) \subset \widetilde{U}_2$. Since $\widetilde{U}_1 \cap \widetilde{U}_2 = \emptyset$, this implies that $\widetilde{f}_1(y') \neq \widetilde{f}_2(y')$ for all $y' \in W$, so W is disjoint from E, as desired. □

2.2. Sections

We now discuss a special kind of lifting problem.

2.2.1. Definition of section. Let $p\colon \widetilde{X} \to X$ be a covering space. A *section* of p is a lift $\sigma\colon X \to \widetilde{X}$ of the identity map $\mathbb{1}_X\colon X \to X$ through p. In other words, $\sigma\colon X \to \widetilde{X}$ is a map such that $p(\sigma(x)) = x$ for all $x \in X$.

EXAMPLE 2.2.1. Let X be a space and $\mathcal{I}$ be a discrete set. Consider the trivial cover $p\colon X \times \mathcal{I} \to X$. Let $i_0 \in \mathcal{I}$, and define $\sigma\colon X \to X \times \mathcal{I}$ via the formula $\sigma(x) = (x, i_0)$. Then σ is a section. □

2.2.2. Most covers have no sections. The above example might be unsatisfying; however, covers typically have no sections:

LEMMA 2.2.2. *Let $p\colon \widetilde{X} \to X$ be a covering space with $\widetilde{X}$ connected. Assume that there exists a section $\sigma\colon X \to \widetilde{X}$. Then $p\colon \widetilde{X} \to X$ is a homeomorphism, and in particular has degree* 1.

PROOF. It is enough to prove that $p\colon \widetilde{X} \to X$ has degree 1; indeed, this will imply that p is a bijection, and since covering space maps are open maps we will be able to conclude that p is a homeomorphism.[1] To prove that p has degree 1, it is enough to prove that σ is surjective. Since $\widetilde{X}$ is connected, this will follow if we prove that the image of σ is both open and closed.

Consider $x \in X$. It is enough to prove that $\sigma(x)$ lies in the interior of $\sigma(X)$ and that all points of $p^{-1}(x)$ other than $\sigma(x)$ lie in the interior of $\widetilde{X} \setminus \sigma(X)$. Let U be a trivialized neighborhood of x and let $\{\widetilde{U}_i\}_{i \in \mathcal{I}}$ be the sheets lying above U. Let $i_0 \in \mathcal{I}$ be such that $\sigma(x) \in \widetilde{U}_{i_0}$. Naively, one might expect that $\sigma(U) = \widetilde{U}_{i_0}$; however, without further assumptions (like U being connected, which could only be ensured if X is locally connected) this need not hold.

However, let $V = \sigma^{-1}(\widetilde{U}_{i_0})$. The set V is also a trivialized neighborhood of x. Let $\{\widetilde{V}_i\}_{i \in \mathcal{I}}$ be the sheets lying above V, enumerated such that $\widetilde{V}_i \subset \widetilde{U}_i$ for all $i \in \mathcal{I}$. Then $\widetilde{V}_{i_0} = \sigma(V)$ is an open neighborhood of $\sigma(x)$ lying in $\sigma(X)$. Also, the union of the $\widetilde{V}_i$ with $i \neq i_0$ is an open neighborhood of $p^{-1}(x) \setminus \sigma(x)$ lying in $\widetilde{X} \setminus \sigma(X)$. The lemma follows. □

2.2.3. Square-free polynomials. We explain an interesting application of Lemma 2.2.2. As discussed in Example 1.2.3, let $\mathrm{Poly}_n^{\mathrm{sf}}$ be the space of monic degree-n polynomials without repeated roots, let $\mathrm{RPoly}_n^{\mathrm{sf}}$ be the space of pairs (f, x) with $f \in \mathrm{Poly}_n^{\mathrm{sf}}$ and $f(x) = 0$, and let $p\colon \mathrm{RPoly}_n^{\mathrm{sf}} \to \mathrm{Poly}_n^{\mathrm{sf}}$ be the map $p(f, x) = f$, so p is a degree n covering space. We start by proving that $\mathrm{RPoly}_n^{\mathrm{sf}}$ is path-connected:

LEMMA 2.2.3. *For $n \geq 1$, the space $\mathrm{RPoly}_n^{\mathrm{sf}}$ is path-connected.*

PROOF. Let (f_1, x_1) and (f_2, x_2) be two points of $\mathrm{RPoly}_n^{\mathrm{sf}}$. We want to find a path from (f_1, x_1) to (f_2, x_2). Since the polynomial $f_i(z)$ has no repeated roots, we can factor it as

$$f_i(z) = (z - x_i)(z - \lambda_{i,1}) \cdots (z - \lambda_{i,n-1}).$$

Here the $\lambda_{i,j}$ are distinct complex numbers that are different from x_i. We remark that the ordering on $\{\lambda_{i,1}, \ldots, \lambda_{i,n-1}\}$ is not canonical. We can move (f_i, x_i) in $\mathrm{RPoly}_n^{\mathrm{sf}}$ by moving x_i and the $\lambda_{i,j}$ while keeping them distinct. Moving x_1 and the $\lambda_{1,j}$ slightly, we can ensure that the numbers

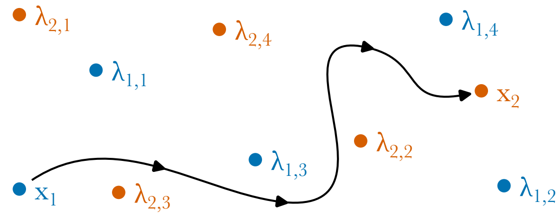
$$Z = \{x_1, \lambda_{1,1}, \ldots, \lambda_{1,n-1}, x_2, \lambda_{2,1}, \ldots, \lambda_{2,n-1}\}$$

are all distinct. We will now move the points $x_1, \lambda_{1,1}, \ldots, \lambda_{1,n-1}$ to $x_2, \lambda_{2,1}, \ldots, \lambda_{2,n-1}$ one at a time, starting with x_1.

Since removing finitely many points from $\mathbb{C}$ does not disconnect it, the space $(\mathbb{C} \setminus Z) \cup \{x_1, x_2\}$ is path-connected. We can therefore find a path in $(\mathbb{C} \setminus Z) \cup \{x_1, x_2\}$ from x_1 to x_2:

---

[1] We remark that there is a simpler proof that p has degree 1 if $\widetilde{X}$ is path connected. Consider $x \in X$. Set $z_1 = \sigma(x)$, so $z_1 \in p^{-1}(x)$. Consider $z_2 \in p^{-1}(x)$. We must prove that $z_1 = z_2$. Let $\widetilde{\gamma}\colon [0,1] \to \widetilde{X}$ be a path with $\widetilde{\gamma}(0) = z_1$ and $\widetilde{\gamma}(1) = z_2$. Set $\gamma = p \circ \gamma$, so $\gamma\colon [0,1] \to X$ is a path in X from $x = p(z_1)$ to $x = p(z_2)$. Define $\widetilde{\gamma}' = \sigma \circ \gamma$. Both $\widetilde{\gamma}$ and $\widetilde{\gamma}'$ are lifts of $\gamma\colon [0,1] \to X$ to $\widetilde{X}$ with $\widetilde{\gamma}(0) = \widetilde{\gamma}'(0) = z_1$, so by Lemma 2.1.2 we have $\widetilde{\gamma} = \widetilde{\gamma}'$. We conclude that $z_1 = \widetilde{\gamma}(1) = \widetilde{\gamma}'(1) = z_2$, as desired.



By moving x_1 along this path, we move (f_1, x_1) and reduce ourselves to the case where $x_1 = x_2$. Next, the space $(\mathbb{C} \setminus Z) \cup \{\lambda_{1,1}, \lambda_{2,1}\}$ is path-connected, so we can find a path in it from $\lambda_{1,1}$ to $\lambda_{2,1}$. By moving $\lambda_{1,1}$ along this path, we move (f_1, x_1) and reduce ourselves to the case where $x_1 = x_2$ and $\lambda_{1,1} = \lambda_{2,1}$. Repeating this process, we move (f_1, x_1) to (f_2, x_2). □

Combining this with Lemma 2.2.2, we deduce the following:

Corollary 2.2.4. *For $n \geq 2$, the covering space $p\colon \mathrm{RPoly}_n^{\mathrm{sf}} \to \mathrm{Poly}_n^{\mathrm{sf}}$ has no section.*

Why is this interesting? Recall that $\mathrm{Poly}_n^{\mathrm{sf}} \subset \mathbb{C}^n$, where

$$f(z) = z^n + a_1 z^{n-1} + a_2 z^{n-2} + \cdots + a_n \in \mathrm{Poly}_n^{\mathrm{sf}}$$

is identified with $(a_1, \ldots, a_n) \in \mathbb{C}^n$. A section of $p\colon \mathrm{RPoly}_n^{\mathrm{sf}} \to \mathrm{Poly}_n^{\mathrm{sf}}$ is thus a function that takes as input the coefficients of a polynomial $f(z)$ with no repeated roots and returns (f, x), where $x \in \mathbb{C}$ is a root of $f(z)$. In other words, it is a continuous "formula" for the roots of a degree-n polynomial.

The fact that these do not exist for $n \geq 2$ might seem to contradict that we do in fact have such formulas in low degrees. For instance, we have the quadratic formula: for a quadratic polynomial $f(z) = z^2 + bz^1 + c$, its roots are

$$\frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

The point here is that this is not really a well-defined function because of the $\pm$, and indeed there is no way to choose a canonical square root of a complex number in a continuous way. In Chapter YYY, we will see that this forms the germ of a beautiful proof of Arnold of the classical fact (usually proved with Galois theory) that there is no elementary formula for the roots of a degree-n polynomial for $n \geq 5$, even if you allow multivalued k^{th} roots like in the quadratic formula.

2.3. Lifting paths

Once the basic theory of the fundamental group is in place, we will be able to give a satisfying necessary and sufficient condition for a lift to exist, at least for reasonable spaces (see §YYY). Before we can do this, we need to solve some important special cases. As notation, let $I = [0, 1]$. A *path* in a space X is a map $\gamma\colon I \to X$. The *initial point* of γ is $\gamma(0)$ and the *terminal point* is $\gamma(1)$, and we say that γ goes from $\gamma(0)$ to $\gamma(1)$. Paths can always be lifted:

Lemma 2.3.1. *Let $p\colon \widetilde{X} \to X$ be a covering space and let $\gamma\colon I \to X$ be a path. For all $\widetilde{x}_0 \in \widetilde{X}$ with $p(\widetilde{x}_0) = \gamma(0)$, there exists a unique lift $\widetilde{\gamma}\colon I \to \widetilde{X}$ of γ through p with $\widetilde{\gamma}(0) = \widetilde{x}_0$.*

Proof. Uniqueness follows from Lemma 2.1.2, so we must only prove existence. Using the Lebesgue number lemma,[2] we can partition I into subintervals

$$0 = \epsilon_1 < \epsilon_2 < \cdots < \epsilon_n = 1$$

[2]Recall that the Lebesgue number lemma says that if Z is a compact metric space and $\{W_j\}_{j \in J}$ is an open cover of Z, then we can find some $\epsilon > 0$ such that for all $z \in Z$ the ϵ-ball $B_\epsilon(z)$ is contained in some W_j. To find the indicated partition of I, apply this to the cover of I by preimages of trivialized open subsets of X and choose the partition such that each segment $[\epsilon_k, \epsilon_{k+1}]$ has diameter at most the $\epsilon > 0$ given by the lemma.

The Lebesgue number lemma can be proved as follows. Since Z is compact, we can write Z as a union of open balls

$$Z = B_{\epsilon_1}(z_1) \cup \cdots \cup B_{\epsilon_m}(z_m) \quad \text{for some } z_1, \ldots, z_m \in Z \text{ and } \epsilon_1, \ldots, \epsilon_m > 0$$

such that for each $1 \leq k \leq m$ the open ball $B_{2\epsilon_k}(z_k)$ is contained in some W_j. Set $\epsilon = \min(\epsilon_1, \ldots, \epsilon_m)$, and consider $z \in Z$. We can find some $1 \leq k \leq m$ such that $z \in B_{\epsilon_k}(z_k)$. By assumption the set $B_{2\epsilon_k}(z_k)$ is contained in some W_j, and by the triangle inequality we have $B_\epsilon(z) \subset B_{2\epsilon_k}(z_k)$ and thus $B_\epsilon(z) \subset W_j$.

such that for all $1 \leq k < n$ the image $\gamma([\epsilon_k, \epsilon_{k+1}])$ is contained in a trivialized open set in X. We construct our lift $\widetilde{\gamma}$ inductively as follows.

First, define $\widetilde{\gamma}(0) = \widetilde{x}_0$. Next, assume that for some $1 \leq k < n$ we have constructed a lift $\widetilde{\gamma}\colon [0, \epsilon_k] \to \widetilde{X}$ of $\gamma|_{[0,\epsilon_k]}\colon [0, \epsilon_k] \to X$. We extend $\widetilde{\gamma}$ to $[0, \epsilon_{k+1}]$ as follows. Let U be a trivialized open set in X such that $\gamma([\epsilon_k, \epsilon_{k+1}]) \subset U$. Let $\widetilde{U}$ be the sheet lying above U with $\widetilde{\gamma}(\epsilon_k) \in \widetilde{U}$. The restriction $p|_{\widetilde{U}}\colon \widetilde{U} \to U$ is a homeomorphism, and on the interval $[\epsilon_k, \epsilon_{k+1}]$ we define $\widetilde{\gamma}$ to be the composition

$$[\epsilon_k, \epsilon_{k+1}] \xrightarrow{\gamma} U \xrightarrow{(p|_{\widetilde{U}})^{-1}} \widetilde{U} \hookrightarrow \widetilde{X}.$$

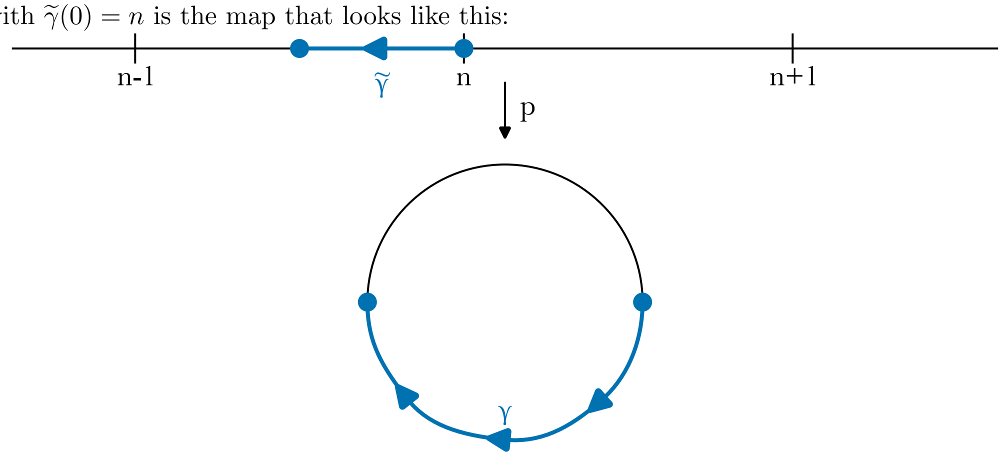
By construction, this agrees with our already-constructed partial lift $\widetilde{\gamma}\colon [0, \epsilon_k] \to \widetilde{X}$ at ϵ_k. □

To help the reader understand the content of this lemma, we give several examples.

Example 2.3.2 (Circle). Let $p\colon \mathbb{R} \to \mathbb{S}^1$ be the universal cover of $\mathbb{S}^1$, so $p(\theta) = e^{2\pi i\theta}$. Let $\gamma\colon [0, 1] \to \mathbb{S}^1$ be the path that starts at $1 \in \mathbb{S}^1 \subset \mathbb{C}$ and travels clockwise half-way around the circle:

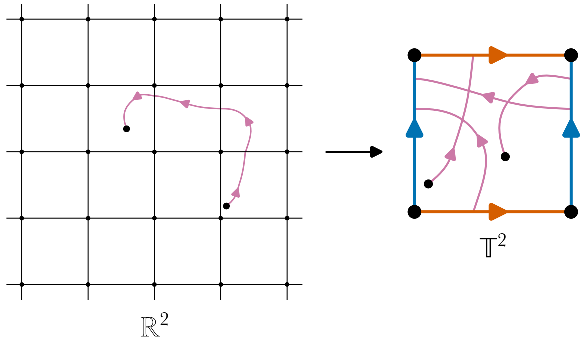
$$\gamma(t) = e^{-\pi i t} \quad \text{for } 0 \leq t \leq 1.$$

The points of $\mathbb{R}$ that project to $\gamma(0) = 1$ are precisely the integers. For $n \in \mathbb{Z}$, the lift $\widetilde{\gamma}\colon [0, 1] \to \mathbb{R}$ of γ with $\widetilde{\gamma}(0) = n$ is the map that looks like this:



In coordinates, $\widetilde{\gamma}(t) = n - t/2$ for $0 \leq t \leq 1$. □

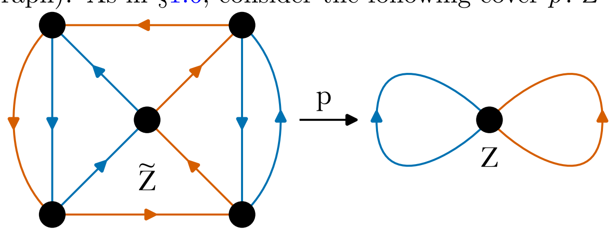
Example 2.3.3 (Torus). As in Example 1.1.10, identify $\mathbb{R}^2/\mathbb{Z}^2$ with the 2-torus $\mathbb{T}^2$ and let $p\colon \mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}^2 = \mathbb{T}^2$ be the associated cover. Here is an example of a path $\gamma\colon [0, 1] \to \mathbb{T}^2$ and one choice of lift $\widetilde{\gamma}\colon [0, 1] \to \mathbb{T}^2$:[3]



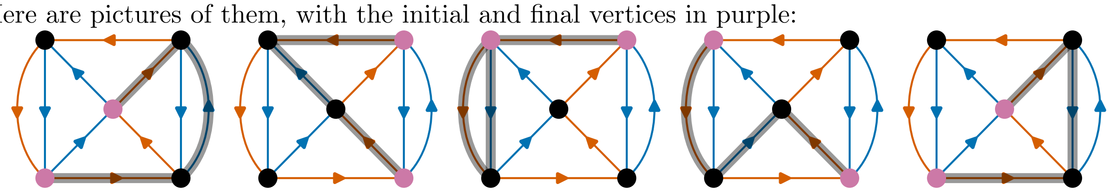
The other possible lifts are obtained by varying the initial point, which results in translating the entire lift by some element of $\mathbb{Z}^2$. □

[3] The torus on the right is obtained by gluing the sides of the square together as indicated. Because of this gluing, a path can e.g. pass through the top edge of the square and come out of the bottom edge.

Example 2.3.4 (Graph). As in §1.6, consider the following cover $p\colon \widetilde{Z} \to Z$:



Let $\gamma\colon [0,1] \to Z$ be the path that starts at the vertex, goes around the orange circle in the positive direction, then goes around the blue circle in the positive direction, and finally goes around the orange circle in the negative direction. There are five possible lifts, one starting at each vertex of $\widetilde{Z}$. Here are pictures of them, with the initial and final vertices in purple:



Constructing these illustrates the necessity that each vertex of $\widetilde{Z}$ has one incoming edge of each color and one outgoing edge of each color. □

Example 2.3.5 (Polar coordinates). Points of $\mathbb{R}^2 \setminus 0$ can be expressed using polar coordinates (r, θ) with $r > 0$ and $\theta \in \mathbb{R}$:

$$(x, y) = (r\cos(\theta), r\sin(\theta)).$$

While $r = \sqrt{x^2 + y^2}$ is unambiguous, the θ-coordinate is ambiguous since $(r, \theta) = (r, \theta + 2\pi n)$ for all $n \in \mathbb{Z}$. Letting $p\colon \mathbb{R} \to \mathbb{S}^1$ be the universal cover, a choice of polar coordinates for $(x, y) \in \mathbb{R}^2 \setminus 0$ is the same as a choice of lift $\theta \in \mathbb{R}$ for the point

$$\left(\frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}}\right) \in \mathbb{S}^1.$$

Lemma 2.3.1 explains why maps $f\colon I \to \mathbb{R}^2 \setminus 0$ can always be continuously expressed using polar coordinates. This is depends on the topology of I, and I cannot be replaced by an arbitrary space.

For instance, the inclusion $\iota\colon \mathbb{S}^1 \hookrightarrow \mathbb{R}^2 \setminus 0$ cannot be continuously described using polar coordinates as $\iota(x) = (r(x), \theta(x))$ for some $r\colon \mathbb{S}^1 \to \mathbb{R}_{>0}$ and $\theta\colon \mathbb{S}^1 \to \mathbb{R}$. Indeed, in such an expression the function r would be identically 1, but the function θ would be a section of the cover $p\colon \mathbb{R} \to \mathbb{S}^1$, and Lemma 2.2.2 implies that such a section does not exist. □

2.4. Homotopies

In algebraic topology, spaces are modeled by algebra. Spaces can vary continuously, while algebraic objects are typically discrete. In this section, we introduce a formalism called homotopy for studying deformations of maps. The algebraic invariants we later study will be insensitive to these deformations.

Remark 2.4.1. It might not be immediately obvious why the reader should care about homotopies. In the next chapter (Chapter 3), we will give a number of applications of our work that use homotopies in an essential way. □

2.4.1. Homotopies of maps. Consider two maps $f, g\colon X \to Y$. We say that f and g are *homotopic* if there exists a continuous map $H\colon X \times I \to Y$ such that

$$f(x) = H(x, 0) \quad \text{and} \quad g(x) = H(x, 1)$$

for all $x \in X$. For $t \in I$, let $h_t\colon X \to Y$ be the map $h_t(x) = H(x, t)$. We thus have $f = h_0$ and $g = h_1$, and we view the h_t as a continuous family of maps witnessing f being deformed to g. Typically we will demonstrate that f and g are homotopic by describing the h_t rather than H, and will call h_t a *homotopy from f to g*. This is an equivalence relations on the set of maps from X to Y (see Exercise 2.1).

EXAMPLE 2.4.2. Let X be a space. Any two maps $f, g\colon X \to \mathbb{R}^n$ are homotopic via the straight-line homotopy $h_t\colon X \to \mathbb{R}^n$ defined via the formula

$$h_t(x) = (1-t)f(x) + tg(x) \quad \text{for all } x \in X \text{ and } t \in I.$$

In this, we have $h_0 = f$ and $h_1 = g$. □

2.4.2. Null-homotopic. We say that a map $f\colon X \to Y$ is *null-homotopic* if f is homotopic to a constant map. As Example 2.4.2 shows, any map $f\colon X \to \mathbb{R}^n$ is null-homotopic. Here is another example:

EXAMPLE 2.4.3. Let X be a space. Then any map $f\colon \mathbb{R}^n \to X$ is null-homotopic via the homotopy $h_t\colon \mathbb{R}^n \to X$ defined via the formula

$$h_t(x) = f((1-t)x) \quad \text{for all } x \in X \text{ and } t \in I.$$

In this, we have $h_0 = f$ and h_1 is the constant map taking all points of $\mathbb{R}^n$ to $f(0)$. □

To show that two maps are homotopic, one typically exhibits an explicit homotopy. It is harder to show that two maps are *not* homotopic. This requires invariants of maps. For instance, the identity map $\mathbb{S}^1 \to \mathbb{S}^1$ is not null-homotopic, but this is not so easy to prove directly. In §2.6 below we will prove this by developing the theory of degrees and winding numbers. In fact, using this we will completely describe *all* homotopy classes of maps $\mathbb{S}^1 \to \mathbb{S}^1$. Doing this requires studying the interaction between homotopies and lifting problems.

2.5. Lifting homotopies

Roughly speaking, our goal in this section is to prove that lifting problems are insensitive to homotopies. To make this precise, consider a covering space $p\colon \widetilde{X} \to X$. Let $f, g\colon Y \to X$ be two homotopic maps. One thing we would like to prove is that a lift $\widetilde{f}\colon Y \to \widetilde{X}$ of f exists if and only if a lift $\widetilde{g}\colon Y \to \widetilde{X}$ exists. We would also like to prove that if these lifts exist then we can choose lifts $\widetilde{f}\colon Y \to \widetilde{X}$ and $\widetilde{g}\colon Y \to \widetilde{X}$ such that $\widetilde{f}$ and $\widetilde{g}$ are themselves homotopic. The following result implies both of these claims.

LEMMA 2.5.1. *Let $p\colon \widetilde{X} \to X$ be a covering space. Let $f\colon Y \to X$ be a map and let $\widetilde{f}\colon Y \to \widetilde{X}$ be a lift of f through p. Let $h_t\colon Y \to X$ be a homotopy with $h_0 = f$. There is then a unique lift of h_t through p to a homotopy $\widetilde{h}_t\colon Y \to \widetilde{X}$ such that $\widetilde{h}_0 = \widetilde{f}$.*

PROOF. Let $H\colon Y \times I \to X$ be the map with $H(y,t) = h_t(y)$ for all $y \in Y$ and $t \in I$. Our goal is to prove that there is a unique lift $\widetilde{H}\colon Y \times I \to \widetilde{X}$ of H through p such that $\widetilde{H}(y,0) = \widetilde{f}(y)$ for all $y \in Y$. Uniqueness follows from Lemma 2.1.2, so we must only prove existence. In fact, even more is true. For $y \in Y$, let $\gamma_y\colon I \to X$ be the path $\gamma_y(t) = H(y,t)$. By path-lifting (Lemma 2.3.1), we can lift γ_y to a path $\widetilde{\gamma}_y\colon I \to \widetilde{X}$ such that $\widetilde{\gamma}_y(0) = \widetilde{f}(0)$. Define $\widetilde{H}\colon Y \times I \to \widetilde{X}$ via the formula

$$\widetilde{H}(y,t) = \widetilde{\gamma}_y(t) \quad \text{for all } y \in Y \text{ and } t \in I.$$

By construction, $\widetilde{H}$ is a lift of H with $\widetilde{H}(y,0) = \widetilde{f}(y)$ for all $y \in Y$.

There is only one problem: it is not obvious that this H is continuous. To see that it is, fix some $y_0 \in Y$. We will prove that $\widetilde{H}$ is continuous at all points of the form (y_0, t) by imitating our proof of the path-lifting lemma (Lemma 2.3.1). Just like in that proof, using the Lebesgue number lemma we can partition I into subintervals

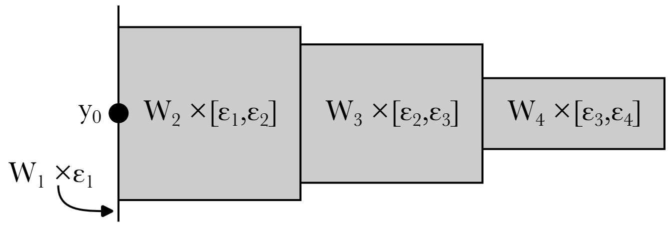
$$0 = \epsilon_1 < \epsilon_2 < \cdots < \epsilon_n = 1$$

such that for all $1 \leq k < n$ the image $H(y_0 \times [\epsilon_k, \epsilon_{k+1}])$ is contained in a trivialized open set in X. In fact, we can even find some open neighborhood V_k of y_0 such that the image $H(V_k \times [\epsilon_k, \epsilon_{k+1}])$ is contained in a trivialized open set in X.

Define $W_1 = V_1 \cap \cdots \cap V_{n-1}$, so W_1 is an open neighborhood of y_0. We will find a nested sequence

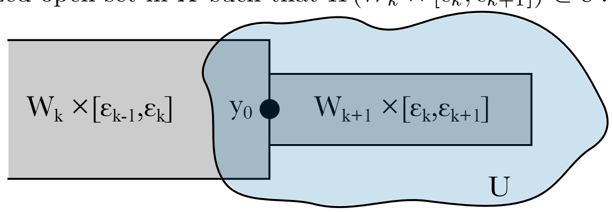
$$W_1 \supset W_2 \supset \cdots \supset W_n$$

of open neighborhoods of y_0 such that $\widetilde{H}$ is continuous on each $W_k \times [0, \epsilon_k]$ by constructing $\widetilde{H}$ on this set in such a way that it is clearly continuous. The picture is:



The construction will be inductive. First, define $\widetilde{H}\colon W_1 \times 0 \to \widetilde{X}$ via the formula $\widetilde{H}(y,0) = \widetilde{f}(y)$. Next, assume that for some $1 \le k < n$ we have constructed a continuous lift $\widetilde{H}\colon W_k \times [0,\epsilon_k] \to \widetilde{X}$ of $H|_{W_k \times [0,\epsilon_k]}\colon W_k \times [0,\epsilon_k] \to X$. We find an open neighborhood W_{k+1} of y_0 with $W_{k+1} \subset W_k$ and an extension of $\widetilde{H}$ to $W_{k+1} \times [0,\epsilon_{k+1}]$ as follows.

Let U be a trivialized open set in X such that $H(W_k \times [\epsilon_k, \epsilon_{k+1}]) \subset U$:



Let $\widetilde{U}$ be the sheet lying above U with $H(y_0, \epsilon_k) \in \widetilde{U}$. Let W_{k+1} be the preimage of $\widetilde{U}$ under the map[4]

$$W_k = W_k \times \epsilon_k \xrightarrow{\widetilde{H}(-,\epsilon_k)} \widetilde{X}.$$

The set W_{k+1} is an open neighborhood of y_0 and $\widetilde{H}$ takes $W_{k+1} \times \epsilon_k$ to $\widetilde{U}$. The restriction $p|_{\widetilde{U}}\colon \widetilde{U} \to U$ is a homeomorphism, and on $W_{k+1} \times [\epsilon_k, \epsilon_{k+1}]$ we define $\widetilde{H}$ to be the composition

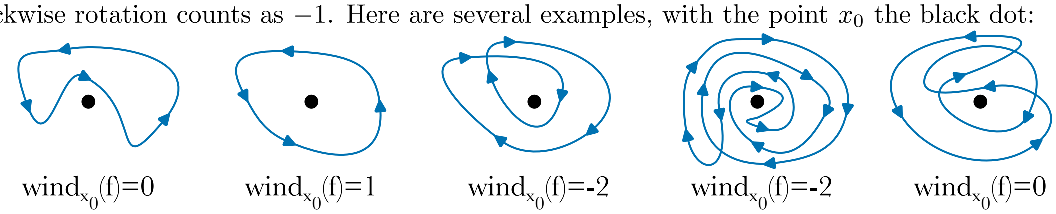
$$W_{k+1} \times [\epsilon_k, \epsilon_{k+1}] \xrightarrow{\ H\ } U \xrightarrow{(p|_{\widetilde{U}})^{-1}} \widetilde{U} \hookrightarrow \widetilde{X}.$$

By construction, this agrees with our already-constructed partial lift $\widetilde{H}\colon W_{k+1} \times [0,\epsilon_k] \to \widetilde{X}$ on $W_{k+1} \times \epsilon_k$. □

2.6. Winding numbers and degrees

To illustrate the meaning of all this machinery, we study the classical subject of winding numbers and degrees of maps from $\mathbb{S}^1$ to itself. Fix a point $x_0 \in \mathbb{C}$.

2.6.1. Intuition. Consider a map $f\colon \mathbb{S}^1 \to \mathbb{C} \setminus x_0$. Roughly speaking, the winding number $\text{wind}_{x_0}(f)$ of f around x_0 measures the number of times the vector $f(z) - x_0$ rotates as z moves around $\mathbb{S}^1$. The "number" here includes a sign: a counterclockwise rotation counts as $+1$, while a clockwise rotation counts as -1. Here are several examples, with the point x_0 the black dot:



One feature of the winding number is that it is invariant under homotopies of f through maps that avoid x_0. In fact, we will prove that it is a *complete* invariant of such maps. It is enlightening to verify that the different f above with the same winding number are homotopic.

[4]If we could ensure that $\widetilde{H}(W_k \times \epsilon_k) \subset \widetilde{U}$ (which would hold, for instance, if W_k and $\widetilde{U}$ were connected), then there would be no need to pass to the nested sequence $W_1 \supset W_2 \supset \cdots$. Since we are not assuming that our spaces are locally connected, this is unfortunately necessary.

2.6.2. Formal definition. Let $p\colon \mathbb{R} \to \mathbb{S}^1$ be the universal cover. Consider some $f\colon \mathbb{S}^1 \to \mathbb{C} \setminus x_0$. Define a map $F\colon I \to \mathbb{S}^1$ via the formula

$$F(t) = \frac{f(e^{2\pi i t}) - x_0}{\|f(e^{2\pi i t}) - x_0\|} \quad \text{for } t \in I.$$

This definition makes sense since $f(z) \neq x_0$ for all $z \in \mathbb{S}^1$. Pick some $\theta_0 \in \mathbb{R}$ such that $p(\theta_0) = F(0)$, and use path lifting (Lemma 2.3.1) to lift F through p to $\widetilde{F}\colon I \to \mathbb{R}$ with $\widetilde{F}(0) = \theta_0$. Since $F(0) = F(1)$, the lifts $\widetilde{F}(0) = \theta_0$ and $\widetilde{F}(1)$ differ by an integer. We define

$$\mathrm{wind}_{x_0}(f) = \widetilde{F}(1) - \widetilde{F}(0) \in \mathbb{Z}.$$

The only arbitrary choice we made was the lift θ_0. Any other choice of θ_0 is of the form $\theta_0 + m$ for some $m \in \mathbb{Z}$, and using $\theta_0 + m$ as our initial lift would change $\widetilde{F}$ to $\widetilde{F} + m$. Since

$$(\widetilde{F}(1) + m) - (\widetilde{F}(0) + m) = \widetilde{F}(1) - \widetilde{F}(0),$$

this would not change $\mathrm{wind}_{x_0}(f)$. In other words, $\mathrm{wind}_{x_0}(f) \in \mathbb{Z}$ is well-defined.

EXAMPLE 2.6.1. Fix $k \in \mathbb{Z}$, and define $f\colon \mathbb{S}^1 \to \mathbb{C} \setminus x_0$ via the formula

$$f(z) = x_0 + z^k \quad \text{for } z \in \mathbb{S}^1 \subset \mathbb{C}.$$

In the above recipe, we then have

$$F(t) = \frac{f(e^{2\pi i t}) - x_0}{\|f(e^{2\pi i t}) - x_0\|} = e^{2\pi i k t} \quad \text{for } t \in I.$$

We can take $\theta_0 = 0$, and then

$$\widetilde{F}(\theta) = k\theta \quad \text{for } \theta \in \mathbb{R}.$$

It follows that $\mathrm{wind}_{x_0}(f) = k$. We thus see that all integers can be winding numbers. □

2.6.3. Homotopy invariance. One of the main properties of the winding number is that it is unchanged under homotopies:

LEMMA 2.6.2. *Let $x_0 \in \mathbb{C}$, and let $f, g\colon \mathbb{S}^1 \to \mathbb{C} \setminus x_0$ be homotopic maps. Then* $\mathrm{wind}_{x_0}(f) = \mathrm{wind}_{x_0}(g)$.

PROOF. Let $p\colon \mathbb{R} \to \mathbb{S}^1$ be the universal cover. As in the definition of the winding number, define $F\colon I \to \mathbb{S}^1$ and $G\colon I \to \mathbb{S}^1$ via the formulas[5]

$$F(s) = \frac{f(e^{2\pi i s}) - x_0}{\|f(e^{2\pi i s}) - x_0\|} \quad \text{and} \quad G(s) = \frac{g(e^{2\pi i s}) - x_0}{\|g(e^{2\pi i s}) - x_0\|} \quad \text{for } s \in I.$$

Pick some $\theta_0 \in \mathbb{R}$ such that $p(\theta_0) = F(0)$, and use path lifting (Lemma 2.3.1) to lift F through $p\colon \mathbb{R} \to \mathbb{S}^1$ to $\widetilde{F}\colon I \to \mathbb{R}$ with $\widetilde{F}(0) = \theta_0$. We then have $\mathrm{wind}_{x_0}(f) = \widetilde{F}(1) - \widetilde{F}(0)$. We will hold off on constructing the lift of G that would determine $\mathrm{wind}_{x_0}(g)$.

Now let $h_t\colon \mathbb{S}^1 \to \mathbb{C} \setminus x_0$ be a homotopy from f to g. Define $H_t\colon I \to \mathbb{S}^1$ via the formula

$$H_t(s) = \frac{h_t(e^{2\pi i s}) - x_0}{\|h_t(e^{2\pi i s}) - x_0\|} \quad \text{for } s \in I,$$

so H_t is a homotopy from $H_0 = F$ to $H_1 = G$. Use the homotopy lifting lemma (Lemma 2.5.1) to lift H_t to a homotopy $\widetilde{H}_t\colon I \to \mathbb{R}$ with $\widetilde{H}_0 = \widetilde{F}$. It follows that $\widetilde{H}_1$ is a lift of G, so $\mathrm{wind}_{x_0}(g) = \widetilde{H}_1(1) - \widetilde{H}_1(0)$. More generally, we have $\mathrm{wind}_{x_0}(h_t) = \widetilde{H}_t(1) - \widetilde{H}_t(0)$ for all $t \in I$. This implies that the map

$$t \mapsto \widetilde{H}_t(1) - \widetilde{H}_t(0) \quad \text{for } t \in I$$

is a continuous integer-valued function. It is thus constant, so

$$\mathrm{wind}_{x_0}(f) = \widetilde{H}_0(1) - \widetilde{H}_0(0) = \widetilde{H}_1(1) - \widetilde{H}_1(0) = \mathrm{wind}_{x_0}(g).$$ □

[5]We use s instead of t since we will later use t when we talk about homotopies.

2.6.4. Degree of map of circle. Consider a map $f\colon \mathbb{S}^1 \to \mathbb{S}^1$. We can regard f as a map to $\mathbb{C} \setminus 0$, giving an integer $\text{wind}_0(f)$ that we will call the *degree*[6] of f. Denote this by $\deg(f)$. Lemma 2.6.2 implies that $\deg(f)$ is invariant under homotopy, and just like in Example 2.6.1 we have $\deg(z^n) = n$ for all $n \in \mathbb{Z}$. In particular, the degree of the identity is 1 and the degree of a constant map is 0. Since these are different, we deduce the following, which was promised at the end of §2.4:

LEMMA 2.6.3. *The identity map* $\mathbb{1}\colon \mathbb{S}^1 \to \mathbb{S}^1$ *is not nullhomotopic.*

2.6.5. Completeness of degree. The following basic result says that the degree is a *complete* invariant of homotopy classes of maps from $\mathbb{S}^1$ to itself:

LEMMA 2.6.4. *Let* $f, g\colon \mathbb{S}^1 \to \mathbb{S}^1$ *be maps with* $\deg(f) = \deg(g)$. *Then* f *is homotopic to* g.

PROOF. By postcomposing f and g with paths of rotations of $\mathbb{S}^1$, we can homotope them to maps with $f(1) = g(1) = 1$. Define $F\colon I \to \mathbb{S}^1$ and $G\colon I \to \mathbb{S}^1$ via the formulas

$$F(s) = f(e^{2\pi i s}) \quad \text{and} \quad G(s) = g(e^{2\pi i s}) \quad \text{for } s \in I.$$

We thus have $F(0) = G(0) = 1$. Letting $p\colon \mathbb{R} \to \mathbb{S}^1$ be the universal cover, by the path lifting lemma (Lemma 2.3.1) we can lift F and G through p to maps $\widetilde{F}, \widetilde{G}\colon I \to \mathbb{R}$ with $\widetilde{F}(0) = \widetilde{G}(0) = 0$. We have $\widetilde{F}(1) = \deg(f)$ and $\widetilde{G}(1) = \deg(g)$, which are equal by assumption. Define $H_t\colon I \to \mathbb{S}^1$ via the formula

$$H_t(s) = p((1-t)\widetilde{F}(s) + t\widetilde{G}(s)) \quad \text{for } s \in I.$$

The maps H_t are a homotopy from F to G, and since $\widetilde{F}(0) = \widetilde{G}(0) = 0$ and $\widetilde{F}(1) = \widetilde{G}(1) \in \mathbb{Z}$ we have $H_t(0) = 1$ and $H_t(1) = 1$ for all $t \in I$. This implies that there exists some $h_t\colon \mathbb{S}^1 \to \mathbb{S}^1$ with

$$H_t(s) = h_t(e^{2\pi i s}) \quad \text{for } s \in I.$$

This h_t is a homotopy from f to g. □

REMARK 2.6.5. Once we have developed some basic results about homology, we will generalize the notion of degree to an integer-valued invariant of maps $f\colon M^n \to N^n$ with M^n and N^n compact oriented n-manifolds. Lemma 2.6.4 generalizes to a deep theorem of Hopf saying that this degree is a complete invariant for maps $f\colon M^n \to \mathbb{S}^n$. □

2.6.6. Completeness of winding number. The following is the analogue for the winding number of Lemma 2.6.4:

LEMMA 2.6.6. *Let* $x_0 \in \mathbb{C}$ *and let* $f, g\colon \mathbb{S}^1 \to \mathbb{C} \setminus x_0$ *be maps with* $\text{wind}_{x_0}(f) = \text{wind}_{x_0}(g)$. *Then* f *is homotopic to* g.

PROOF. Translating everything, we can assume that $x_0 = 0$. Define homotopies $f_t\colon \mathbb{S}^1 \to \mathbb{C} \setminus 0$ and $g_t\colon \mathbb{S}^1 \to \mathbb{C} \setminus 0$ as follows:

$$f_t(z) = \left((1-t) + \frac{t}{\|f(z)\|}\right) f(z) \quad \text{and} \quad g_t(z) = \left((1-t) + \frac{t}{\|g(z)\|}\right) g(z) \quad \text{for } t \in I \text{ and } z \in \mathbb{S}^1.$$

This makes sense since $f_t(z) \neq 0$ and $g_t(z) \neq 0$ for all $z \in \mathbb{S}^1$. We have $f_0 = f$ and $g_0 = g$, and the images of f_1 and g_1 lie in $\mathbb{S}^1$. We can thus talk about the degrees of f_1 and g_1. Since winding numbers are invariant under homotopy (Lemma 2.6.2), we have

$$\deg(f_1) = \text{wind}_0(f_1) = \text{wind}_0(f_0) = \text{wind}_0(g_0) = \text{wind}_0(g_1) = \deg(g_1).$$

Lemma 2.6.4 thus implies that f_1 is homotopic to g_1. The lemma follows. □

2.7. Exercises

EXERCISE 2.1. Let X and Y be spaces. Prove that the relation of being homotopic is an equivalence relation on maps from X to Y. □

[6] If f is a covering space, this is different from the degree of f as a covering space. For instance, it can be negative.

CHAPTER 3

Applications of winding numbers and degrees

This section contains four applications of the material from Chapter 2: the fundamental theorem of algebra (§3.1), the two-dimensional Brouwer fixed point theorem (§3.2), the two-dimensional Borsuk–Ulam theorem (§3.3), and the ham Sandwich theorem (§3.4).

REMARK 3.0.1. The statements of all the theorems we prove do not involve homotopies, but as the reader will see their proofs use homotopies in an important way. □

3.1. Fundamental theorem of algebra

We gave a proof of the fundamental theorem of algebra using covering spaces in §1.2. Here is a proof using winding numbers:

THEOREM 1.2.2 (Fundamental theorem of algebra). *Let $f(z) \in \mathbb{C}[z]$ be a nonconstant polynomial. Then there exists some $z_0 \in \mathbb{C}$ such that $f(z_0) = 0$.*

PROOF. Let $n \geq 1$ be the degree of $f(z)$. Multiplying $f(z)$ by an appropriate nonzero constant, we can assume that $f(z)$ is monic, i.e.,

$$f(z) = z^n + a_{n-1}z^{n-1} + a_{n-2}z^{n-2} + \cdots + a_0 \quad \text{for some } a_0, \ldots, a_{n-1} \in \mathbb{C}.$$

Assume for the sake of contradiction that $f(z)$ has no roots, so $f(z) \neq 0$ for all $z \in \mathbb{C}$. For $R \geq 0$, let $h_R \colon \mathbb{S}^1 \to \mathbb{C} \setminus 0$ be the map

$$h_R(z) = f(Rz) \quad \text{for } z \in \mathbb{S}^1.$$

The map h_0 is the constant map $f(0)$, so each h_R is homotopic to a constant map. Since the winding number is invariant under homotopies (Lemma 2.6.2), we thus have

$$\text{wind}_0(h_R) = 0 \quad \text{for all } R \geq 0.$$

On the other hand, pick R large enough such that

$$R > |a_{n-1}| + |a_{n-2}| + \cdots + |a_0| + 1.$$

For $z \in \mathbb{S}^1$, we then have

$$\begin{aligned}|a_{n-1}(Rz)^{n-1} + a_{n-2}(Rz)^{n-2} + \cdots + a_0| &\leq |a_{n-1}R^{n-1}| + |a_{n-2}R^{n-2}| + \cdots + |a_0| \\ &\leq R^{n-1}(|a_{n-1}| + |a_{n-2}| + \cdots + |a_0|) \\ &< R^n.\end{aligned}$$

For $t \in I$, let $g_t \colon \mathbb{S}^1 \to \mathbb{C}$ be the map

$$g_t(z) = (Rz)^n + t\left(a_{n-1}(Rz)^{n-1} + a_{n-2}(Rz)^{n-2} + \cdots + a_0\right) \quad \text{for } z \in \mathbb{S}^1.$$

Since the above inequality is strict, we have $g_t(z) \neq 0$ for all $z \in \mathbb{S}^1$ and $t \in I$. We can therefore regard g_t as a map $g_t \colon \mathbb{S}^1 \to \mathbb{C} \setminus 0$. We have $g_1 = h_R$, so again using the invariance of the winding number under homotopies (Lemma 2.6.2) we have

$$0 = \text{wind}_0(h_R) = \text{wind}_0(g_1) = \text{wind}_0(g_0).$$

We have $g_0(z) = (Rz)^n = R^n z^n$. For $t \in I$, let $h_t \colon \mathbb{S}^1 \to \mathbb{C} \setminus 0$ be the map

$$h_t(z) = ((1-t) + \frac{t}{R^n})R^n z^n \quad \text{for } z \in \mathbb{S}^1.$$

We have $h_0 = g_0$ and $h_1(z) = z^n$. As we noted in Example 2.6.1, the winding number around 0 of $h_1(z) = z^n$ is n. We conclude that

$$\mathrm{wind}_0(g_0) = \mathrm{wind}_0(h_1) = n \neq 0,$$

a contradiction. □

3.2. Two-dimensional Brouwer fixed point theorem

Our next goal is the two-dimensional Brouwer fixed point theorem, which says that every map $f\colon \mathbb{D}^2 \to \mathbb{D}^2$ fixes a point of $\mathbb{D}^2$.

3.2.1. Retractions. A *retraction* of a space X to a subspace $Y \subset X$ is a map $r\colon X \to Y$ such that $r(y) = y$ for all $y \in Y$. If such a retraction exists, we say that Y is *retract* of X.

EXAMPLE 3.2.1. Let ℓ be a line in $\mathbb{R}^n$. We can then define a retraction $r\colon \mathbb{R}^n \to \ell$ by letting $r(x)$ be the point of ℓ that is closest to x for all $x \in \mathbb{R}^n$. □

EXAMPLE 3.2.2. Let $\mathbb{D}^1 = [-1, 1]$, so $\mathbb{S}^0 = \{-1, 1\}$ is $\partial\mathbb{D}^1$. There does not exist a retraction $r\colon \mathbb{D}^1 \to \mathbb{S}^0$. Indeed, since $\mathbb{D}^1$ is connected the image of any function $\mathbb{D}^1 \to \mathbb{D}^0$ must be either $\{-1\}$ or $\{1\}$. □

3.2.2. Retractions of disks. For $n \geq 1$, let $\mathbb{D}^n$ be the closed unit disk in $\mathbb{R}^n$, so $\mathbb{S}^{n-1} = \partial\mathbb{D}^n$. Generalizing Example 3.2.2, it seems reasonable to expect that there does not exist a retraction $r\colon \mathbb{D}^n \to \mathbb{S}^{n-1}$ since intuitively such a retraction must "tear" $\mathbb{D}^n$ at some point. This turns out to be true, but is not easy to prove directly. We have the technology to prove it for $n = 2$:

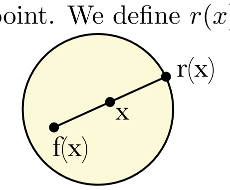
PROPOSITION 3.2.3. *There does not exist a retraction* $r\colon \mathbb{D}^2 \to \mathbb{S}^1$.

PROOF. Assume that a retraction $r\colon \mathbb{D}^2 \to \mathbb{S}^1$ exists. Define a homotopy $f_t\colon \mathbb{S}^1 \to \mathbb{S}^1$ via the formula $f_t(z) = r((1-t)z)$ for $z \in \mathbb{S}^1$. We have $f_0(z) = r(z) = z$, so $f_0 = \mathbb{1}$. On the other hand, $f_1(z) = r(0)$, so f_1 is a constant map. Since the identity map $\mathbb{1}\colon \mathbb{S}^1 \to \mathbb{S}^1$ is not null-homotopic (see Lemma 2.6.3), this is a contradiction. □

3.2.3. Brouwer fixed point theorem. We can now prove our main result:

THEOREM 3.2.4 (Two-dimensional Brouwer fixed point theorem). *Let* $f\colon \mathbb{D}^2 \to \mathbb{D}^2$ *be a continuous map. Then* f *has a fixed point, i.e., there exists some* $x \in \mathbb{D}^2$ *with* $f(x) = x$.

PROOF. Assume that f has no fixed points. Define a function $r\colon \mathbb{D}^2 \to \mathbb{S}^1$ as follows. For $x \in \mathbb{D}^2$, consider the ray starting at $f(x)$ and passing through x. This is well-defined since $f(x) \neq x$, and it intersects the boundary $\mathbb{S}^1$ in a single point. We define $r(x)$ to be that intersection point:



For $x \in \mathbb{S}^1$, we have $r(x) = x$. In other words, r is a retraction from $\mathbb{D}^2$ to its boundary $\mathbb{S}^1$, contradicting Proposition 3.2.3. □

REMARK 3.2.5. Once we have developed the basic theory of homology in a later volume, we will extend Proposition 3.2.3 and Theorem 3.2.4 to all $\mathbb{D}^n$. The case $n = 1$ can be proved by substituting Example 3.2.2 for Proposition 3.2.3 in the proof of Theorem 3.2.4, or more directly using the intermediate value theorem (see Exercise 3.1). □

3.3. Two-dimensional Borsuk–Ulam theorem

We now turn to the Borsuk–Ulam theorem.

3.3.1. One-dimensional warmup. We start with the following:

LEMMA 3.3.1. *Let $f\colon \mathbb{S}^1 \to \mathbb{R}$ be a map. Then there exists some $z \in \mathbb{S}^1$ such that $f(z) = f(-z)$.*

PROOF. Define $F\colon I \to \mathbb{R}$ via the formula

$$F(t) = f(e^{\pi i t}) - f(e^{\pi i (t+1)}) \quad \text{for } t \in I.$$

We have

$$F(0) = f(1) - f(-1) = -(f(-1) - f(1)) = -F(1).$$

By the intermediate value theorem, there exists some $t_0 \in I$ such that $F(t_0) = 0$, which implies that

$$f(e^{\pi i t_0}) = f(e^{\pi i (t_0+1)}) = f(-e^{\pi i t_0}). \qquad \square$$

3.3.2. Two-dimensional theorem. Lemma 3.3.1 implies a similar result for maps $f\colon \mathbb{S}^2 \to \mathbb{R}$. In fact, even more is true: for *every* plane P through the origin in $\mathbb{R}^3$, we can apply Lemma 3.3.1 to the restriction of f to the circle $P \cap \mathbb{S}^2$ and find some $z \in P \cap \mathbb{S}^2$ with $f(z) = f(-z)$. Set

$$F = \left\{ z \in \mathbb{S}^2 \mid f(z) = f(-z) \right\}.$$

This set F is in general rather complicated, but by the above it cannot be too small. In fact, it is reasonable to hope that in some sense it is at least 1-dimensional. Given another $g\colon \mathbb{S}^2 \to \mathbb{R}$, this suggests that we might be able to apply something like Lemma 3.3.1 to the restriction of g to F and find some $z \in F$ with $g(z) = g(-z)$. For instance, we could directly apply the argument in the proof of Lemma 3.3.1 to do this if there existed a path $\gamma\colon I \to F$ with $\gamma(1) = -\gamma(0)$. We would therefore have found a single $z \in \mathbb{S}^2$ with

$$f(z) = f(-z) \quad \text{and} \quad g(z) = g(-z).$$

It is hard to turn the above into a rigorous proof, but the result it would give is true:

THEOREM 3.3.2 (Two-dimensional Borsuk–Ulam theorem). *Let $f, g\colon \mathbb{S}^2 \to \mathbb{R}$ be two maps. Then there exists some $z \in \mathbb{S}^2$ such that $f(z) = f(-z)$ and $g(z) = g(-z)$.*

As an illustrative example, regard $\mathbb{S}^2$ as the surface of the earth. Let $f\colon \mathbb{S}^2 \to \mathbb{R}$ and $g\colon \mathbb{S}^2 \to \mathbb{R}$ be the functions

$$f(z) = \text{temperature at } z,$$
$$g(z) = \text{elevation at } z.$$

Applying Theorem 3.3.2, we conclude that there exists a point z on the surface of the earth such that the temperature and elevation at z are the same as at $-z$.

PROOF OF THEOREM 3.3.2. Assume that no such z exists. Define maps $\phi\colon \mathbb{S}^2 \to \mathbb{R}^2 \setminus 0$ and $\psi\colon \mathbb{S}^2 \to \mathbb{S}^1$ via the formulas

$$\phi(z) = (f(z) - f(-z), g(z) - g(-z)) \quad \text{and} \quad \psi(z) = \frac{\phi(z)}{\|\phi(z)\|} \quad \text{for } z \in \mathbb{S}^2.$$

This makes sense since the fact that there is no z satisfying the conclusion of the theorem implies that $\phi(z) \neq 0$ for all $z \in \mathbb{S}^2$. The functions ϕ and ψ are *odd functions*, i.e., $\phi(-z) = -\phi(z)$ and $\psi(-z) = -\psi(z)$ for all $z \in \mathbb{S}^2$.

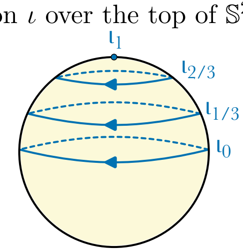
Regard $\mathbb{S}^1$ as a subspace of $\mathbb{R}^2$, and define $h\colon \mathbb{S}^1 \to \mathbb{S}^1$ as the composition

$$\mathbb{S}^1 \xhookrightarrow{\iota} \mathbb{S}^2 \xrightarrow{\psi} \mathbb{S}^1,$$

where the inclusion ι takes $(x, y) \in \mathbb{S}^1$ to $(x, y, 0) \in \mathbb{S}^2$. The function h is also an odd function. Below in Lemma 3.3.3 we will prove that this implies that $\deg(h) \in \mathbb{Z}$ is an odd number. However, the map ι is null-homotopic via the homotopy $\iota_t\colon \mathbb{S}^1 \to \mathbb{S}^2$ defined via the formula

$$\iota_t(x, y) = (tx, ty, 1 - t\sqrt{x^2 + y^2}) \in \mathbb{S}^2 \quad \text{for } t \in I \text{ and } (x, y) \in \mathbb{S}^1.$$

Geometrically, this "pulls" the inclusion ι over the top of $\mathbb{S}^2$ as in:



Letting $h_t = \psi \circ \iota_t$, we have $h_0 = h$ and $h_1 = \psi(0,0,1)$. In other words, h_t is a homotopy from h to a constant map. By the discussion in §2.6.4, this implies that $\deg(h) = 0$. Summarizing, we have proved that $\deg(h)$ is both odd and zero, which is a contradiction. □

The above proof used:

LEMMA 3.3.3. *Let* $h\colon \mathbb{S}^1 \to \mathbb{S}^1$ *be an odd function. Then* $\deg(h)$ *is odd.*

PROOF. Let $p\colon \mathbb{R} \to \mathbb{S}^1$ be the universal cover. In §2.6.4, we defined the degree in terms of the winding number. As in the definition of the winding number from §2.6, let $H\colon I \to \mathbb{S}^1$ be the path $H(t) = h(e^{2\pi i t})$, let $\theta_0 \in \mathbb{R}$ be such that $p(\theta_0) = H(0)$, and let $\widetilde{H}\colon I \to \mathbb{R}$ be the lift of H through p with $\widetilde{H}(0) = \theta_0$ (see Lemma 2.3.1). By definition, we have

$$\deg(h) = \operatorname{wind}_0(h) = \widetilde{H}(1) - \widetilde{H}(0) \in \mathbb{Z}.$$

We have $H(0) = h(1)$. Since h is an odd function, we have $H(1/2) = h(-1) = -h(1)$. It follows that

$$\widetilde{H}(1/2) = \widetilde{H}(0) + n + \frac{1}{2} \quad \text{for some } n \in \mathbb{Z}.$$

The function $\widetilde{H}'\colon I \to \mathbb{R}$ defined by

$$\widetilde{H}'(t) = \begin{cases} \widetilde{H}(t) & \text{if } 0 \leq t \leq 1/2 \\ n + \frac{1}{2} + \widetilde{H}(t - 1/2) & \text{if } 1/2 \leq t \leq 1 \end{cases}$$

is continuous, and since $h(-z) = -h(z)$ it is a lift of $H\colon I \to \mathbb{S}^1$ through p. We thus have $\widetilde{H}' = \widetilde{H}$, so

$$\deg(h) = \widetilde{H}(1) - \widetilde{H}(0) = \widetilde{H}'(1) - \widetilde{H}(0) = n + \frac{1}{2} + \widetilde{H}(1/2) - \widetilde{H}(0) = 2n + 1.$$ □

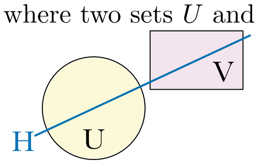
REMARK 3.3.4. We will generalize the results from this section to n dimensions in after we develop the theory of homology in a later volume. The general n-dimensional Borsuk–Ulam theorem says that for all maps $F\colon \mathbb{S}^n \to \mathbb{R}^n$, there exists some $z \in \mathbb{S}^n$ with $F(z) = F(-z)$. This is harder than the Brouwer Fixed Point Theorem, and will take more than just the basics of homology. The hardest part is generalizing Lemma 3.3.3, which requires a deeper structure called the cup product that exists on a variant of homology called cohomology. □

3.4. Ham sandwich theorem

We now give an interesting geometric application of the Borsuk–Ulam theorem called the "ham sandwich theorem".

3.4.1. Informal statement. Informally, it says that given any sandwich made up of bread and meat and cheese, there is a way to cut the sandwich such that each half has equal amounts of bread and meat and cheese. To make this precise, assume that the sandwich is in $\mathbb{R}^3$ and let U and V and W be the regions composed of the meat and cheese and bread. What the theorem says is that there is an affine hyperplane H that simultaneously divides U and V and W into sets of equal volume.

Since the author of this book is not good at 3-dimensional figures, here is an illustration of the corresponding fact in two dimensions, where two sets U and V are cut in half by a hyperplane:

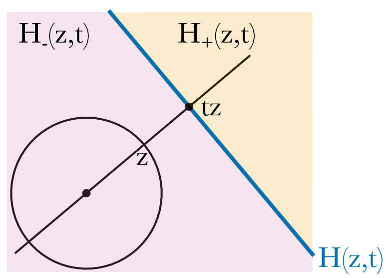


More generally, the sets U and V and W can be any bounded Lebesgue measurable sets (even ones that overlap). Proving this requires some preliminaries. Just like the Borsuk–Ulam theorem, there is a version of the ham sandwich theorem for subsets of $\mathbb{R}^n$. Since the preliminaries are no harder in $\mathbb{R}^n$, we discuss them in general.

3.4.2. Volume. For $X \subset \mathbb{R}^n$, denote by $\mu(X)$ the volume of X. One of the basic insights of measure theory is that this cannot be defined in a reasonable way for all X. For most of this section, we will restrict ourselves to X that are bounded and open. Such sets are simple enough that all definitions of $\mu(X)$ agree on them. For instance:

- For $d_1, \ldots, d_n > 0$, let X be the open n-dimensional rectangle
$$X = (0, d_1) \times \cdots \times (0, d_n) \subset \mathbb{R}^n.$$
Then $\mu(X) = d_1 \cdots d_n$.

3.4.3. Affine hyperplanes. An *affine hyperplane* in $\mathbb{R}^n$ is a subset of the form $p + L$ with $p \in \mathbb{R}^n$ and $L \subset \mathbb{R}^n$ an $(n-1)$-dimensional linear subspace. We parameterize these as follows. Consider $z \in \mathbb{S}^{n-1}$ and $t \in \mathbb{R}$. We then let $H(z,t)$ be the unique affine hyperplane passing through tz that is orthogonal to the line through 0 and z. The set $\mathbb{R}^n \setminus H(z,t)$ has two path components. Let $H_+(z,t)$ be the path component of $\mathbb{R}^n \setminus H(z,t)$ containing $(t+1)z$ and let $H_-(z,t)$ be the path component containing $(t-1)z$:



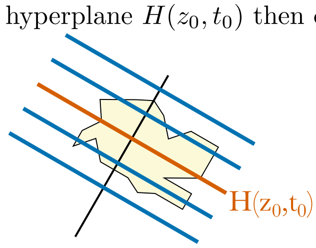
Note that

$$H(-z,-t) = H(z,t) \quad \text{and} \quad H_+(-z,-t) = H_-(z,t) \quad \text{and} \quad H_-(-z,-t) = H_+(z,t). \tag{3.4.1}$$

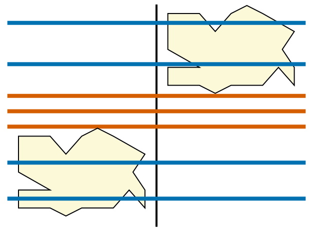
3.4.4. Cutting in half along affine hyperplanes. Let $U \subset \mathbb{R}^n$ be a bounded open subset. We can find an affine hyperplane dividing U in half as follows. Define $F\colon \mathbb{S}^{n-1} \times \mathbb{R} \to \mathbb{R}$ via the formula

$$F(z,t) = \mu(U \cap H_+(z,t)) \quad \text{for } z \in \mathbb{S}^{n-1} \text{ and } t \in \mathbb{R}. \tag{3.4.2}$$

This is a continuous function. For a fixed $z_0 \in \mathbb{S}^{n-1}$, the function $t \mapsto F(z_0,t)$ monotonically goes from 0 to $\mu(U)$ as t traverses $\mathbb{R}$, so by the intermediate value theorem there exists some $t_0 \in \mathbb{R}$ such that $F(z_0,t_0) = \mu(U)/2$. The affine hyperplane $H(z_0,t_0)$ then divides U in half:



The value t_0 need not be unique. However, this can only happen if the map $t \mapsto F(z_0,t)$ is not strictly monotonic at t_0. In other words, at least one of either increasing or decreasing t_0 does not cause the hyperplane to sweep out any more volume. This implies that U is *not* path-connected:



If U is path-connected, then the function $t \mapsto F(z_0, t)$ is *strictly* monotonic at all points where $F(z_0, t) \neq 0$ and $F(z_0, t) \neq \mu(U)$. It follows that in this case t_0 is unique, and we call the function $\mathcal{D}\colon \mathbb{S}^{n-1} \to \mathbb{R}$ taking $z_0 \in \mathbb{S}^{n-1}$ to the unique $t_0 \in \mathbb{R}$ with $F(z_0, t_0) = \mu(U)/2$ the *dividing function* of U.

3.4.5. Properties of dividing function. The dividing function satisfies:

Lemma 3.4.1. *Let $U \subset \mathbb{R}^n$ be a path-connected bounded open subset and let $\mathcal{D}\colon \mathbb{S}^{n-1} \to \mathbb{R}$ be the dividing function of U. Then $\mathcal{D}$ is continuous.*

Proof. Let $z_0 \in \mathbb{S}^{n-1}$ and let $\epsilon > 0$. Set $t_0 = \mathcal{D}(z_0)$. Our goal is to find an open neighborhood X of z_0 such that $\mathcal{D}(X) \subset (t_0 - \epsilon, t_0 + \epsilon)$. Let $F\colon \mathbb{S}^{n-1} \times \mathbb{R} \to \mathbb{R}$ be the continuous function (3.4.2) used to define $\mathcal{D}$. Since U is path-connected, the function $t \mapsto F(z_0, t)$ is strictly monotonic at t_0. In particular,

$$F(z_0, t_0 - \epsilon) < \mu(U)/2 \quad \text{and} \quad F(z_0, t_0 + \epsilon) > \mu(U)/2.$$

Since F is continuous, we can find some open neighborhood X of z_0 such that

$$F(z_1, t_0 - \epsilon) < \mu(U)/2 \quad \text{and} \quad F(z_1, t_0 + \epsilon) > \mu(U)/2 \quad \text{for all } z_1 \in X. \tag{3.4.3}$$

For $z_1 \in X$, the value $\mathcal{D}(z_1)$ is the unique number such that $F(z_1, \mathcal{D}(z_1)) = \mu(U)/2$. By (3.4.3), we must have

$$t_0 - \epsilon < \mathcal{D}(z_1) < t_0 + \epsilon.$$

In other words, $\mathcal{D}(X) \subset (t_0 - \epsilon, t_0 + \epsilon)$, as desired. □

It also satisfies:

Lemma 3.4.2. *Let $U \subset \mathbb{R}^n$ be a path-connected bounded open subset and let $\mathcal{D}\colon \mathbb{S}^{n-1} \to \mathbb{R}$ be the dividing function of U. Then $\mathcal{D}$ is an odd function, i.e., $\mathcal{D}(-z) = -\mathcal{D}(z)$ for all $z \in \mathbb{S}^{n-1}$.*

Proof. Fix some $z \in \mathbb{S}^{n-1}$. By definition, the hyperplanes $H(z, \mathcal{D}(z))$ and $H(-z, \mathcal{D}(-z))$ divide U in half. This implies that $H(z, \mathcal{D}(z)) = H(-z, \mathcal{D}(-z))$. Since the line through the origin and z intersects this hyperplane at one point, we must have $\mathcal{D}(z)\cdot z = \mathcal{D}(-z)\cdot(-z)$. This implies that $\mathcal{D}(-z) = -\mathcal{D}(z)$, as desired. □

3.4.6. Ham sandwich theorem for open sets. We now specialize to $\mathbb{R}^3$, and prove:

Theorem 3.4.3 (Ham sandwich theorem). *Let $U, V, W \subset \mathbb{R}^3$ be bounded open sets. There exists $z \in \mathbb{S}^2$ and $t \in \mathbb{R}$ such that the affine hyperplane $H(z, t)$ divides U and V and W in half:*

$$\mu(U \cap H_+(z, t)) = \mu(U)/2,$$
$$\mu(V \cap H_+(z, t)) = \mu(V)/2,$$
$$\mu(W \cap H_+(z, t)) = \mu(W)/2.$$

Proof. We divide the proof into two steps. The key geometric idea is in the first step.

Step 1. *The theorem is true when U is path-connected.*

Let $\mathcal{D}\colon \mathbb{S}^2 \to \mathbb{R}$ be the dividing function for U, so

$$\mu(U \cap H_+(z, \mathcal{D}(z))) = \mu(U)/2 \quad \text{for all } z \in \mathbb{S}^2.$$

By Lemma 3.4.1, the function $\mathcal{D}$ is continuous. Define $f\colon \mathbb{S}^2 \to \mathbb{R}$ and $g\colon \mathbb{S}^2 \to \mathbb{R}$ as follows:

$$f(z) = \mu(V \cap H_+(z, \mathcal{D}(z))) \quad \text{and} \quad g(z) = \mu(W \cap H_+(z, \mathcal{D}(z))) \quad \text{for all } z \in \mathbb{S}^2.$$

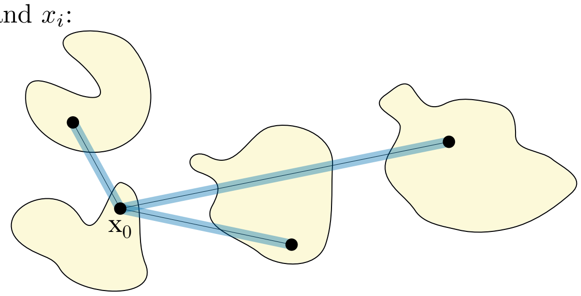
Since $\mathcal{D}$ is continuous, these are both continuous. The two-dimensional Borsuk–Ulam theorem (Theorem 3.3.2) implies that there exists some $z_0 \in \mathbb{S}^2$ such that $f(-z_0) = f(z_0)$ and $g(-z_0) = g(z_0)$. Since $f(-z_0) = f(z_0)$, we have

$$\mu(V \cap H_+(z_0, \mathcal{D}(z_0))) = \mu(V \cap H_+(-z_0, \mathcal{D}(-z_0))) = \mu(V \cap H_-(z_0, \mathcal{D}(z_0))). \tag{3.4.4}$$

The second equality uses the symmetries of our hyperplanes recorded in (3.4.1). The equality (3.4.4) says that the portions of V in the half-spaces on either side of the hyperplane $H(z_0, \mathcal{D}(z_0))$ contain equal volume. This implies that $\mu(V \cap H_+(z_0, \mathcal{D}(z_0))) = \mu(V)/2$. Similarly, we have $\mu(W \cap H_+(z_0, \mathcal{D}(z_0))) = \mu(W)/2$. By the definition of the dividing function $\mathcal{D}$, we also have $\mu(U \cap H_+(z_0, \mathcal{D}(z_0))) = \mu(U)/2$. The theorem follows.

Step 2. *The theorem is true in general.*

The open set U has countably many connected components. Let $x_0 \in U$ be a point in one of them, and let $\{x_i\}_{i \in \mathcal{I}}$ consist of one point in each of the other components. Let $n \geq 1$, and fix some collection $\{r_i\}_{i \in \mathcal{I}}$ of positive numbers. For $i \in \mathcal{I}$, let τ_i be an open tube of radius r_i about the line segment connecting x_0 and x_i:



Let U' be the union of U and the τ_i as i ranges over $\mathcal{I}$. Choosing the r_i small enough (and going to 0 fast enough if $\mathcal{I}$ is infinite), the following hold:

- U' is a path-connected bounded open set with $U \subset U'$; and
- $\mu(U) \leq \mu(U') \leq \mu(U) + 1/n$.

Applying Step 1 to U' and V and W, we get a $z_n \in \mathbb{S}^n$ and $t_n \in \mathbb{R}$ such that

$$\mu(U' \cap H_+(z_n, t_n)) = \mu(U')/2,$$
$$\mu(V \cap H_+(z_n, t_n)) = \mu(V)/2,$$
$$\mu(W \cap H_+(z_n, t_n)) = \mu(W)/2.$$

Since $U \subset U'$ and $\mu(U) \leq \mu(U') \leq \mu(U) + 1/n$, we have

$$\mu(U \cap H_+(z_n, t_n)) \leq \mu(U' \cap H_+(z_n, t_n)) = \mu(U')/2 \leq \mu(U)/2 + \frac{1}{2n}$$

and

$$\mu(U)/2 \leq \mu(U')/2 = \mu(U' \cap H_+(z_n, t_n)) \leq \mu(U \cap H_+(z_n, t_n)) + \frac{1}{n}.$$

Combining these two inequalities, we see that

$$\mu(U)/2 - \frac{1}{n} \leq \mu(U \cap H_+(z_n, t_n)) \leq \mu(U)/2 + \frac{1}{2n}. \tag{3.4.5}$$

The z_n are points of the compact space $\mathbb{S}^2$ and the $t_n \in \mathbb{R}$ are bounded. Passing to a subsequence, we can therefore assume that they converge to $z \in \mathbb{S}^2$ and $t \in \mathbb{R}$. By the continuity of volume, (3.4.5) implies that

$$\mu(U \cap H_+(z, t)) = \mu(U)/2.$$

Similarly, since $\mu(V \cap H_+(z_n, t_n)) = \mu(V)/2$ and $\mu(W \cap H_+(z_n, t_n)) = \mu(W)/2$ we have

$$\mu(V \cap H_+(z, t)) = \mu(V)/2,$$
$$\mu(W \cap H_+(z, t)) = \mu(W)/2.$$

The theorem follows. □

3.4.7. Measurable ham sandwich theorem. We now deduce the general form of the ham sandwich theorem. This requires a small amount of measure theory for the statement and proof, but no sophisticated measure theoretic results are needed. A reader who is not familiar with measure theory can replace "Lebesgue measurable set" with any condition they know that ensures that the set has a reasonable notion of volume.

THEOREM 3.4.4 (Measurable ham sandwich theorem). *Let $X, Y, Z \subset \mathbb{R}^3$ be bounded Lebesgue measurable sets. Then there exists $z \in \mathbb{S}^2$ and $t \in \mathbb{R}$ such that the affine hyperplane $H(z,t)$ divides X and Y and Z in half:*

$$\mu(X \cap H_+(z,t)) = \mu(X)/2,$$
$$\mu(Y \cap H_+(z,t)) = \mu(Y)/2,$$
$$\mu(Z \cap H_+(z,t)) = \mu(Z)/2.$$

PROOF. Fix some $n \geq 1$. Since X and Y and Z are bounded Lebesgue measurable sets, we can find bounded open sets U and V and W with $X \subset U$ and $Y \subset V$ and $Z \subset W$ such that

$$\mu(X) \leq \mu(U) \leq \mu(X) + \frac{1}{n},$$
$$\mu(Y) \leq \mu(V) \leq \mu(Y) + \frac{1}{n},$$
$$\mu(Z) \leq \mu(W) \leq \mu(Z) + \frac{1}{n}.$$

Applying Theorem 3.4.3, we obtain $z_n \in \mathbb{S}^2$ and $t_n \in \mathbb{R}$ such that

$$\mu(U \cap H_+(z_n,t_n)) = \mu(U)/2,$$
$$\mu(V \cap H_+(z_n,t_n)) = \mu(V)/2,$$
$$\mu(W \cap H_+(z_n,t_n)) = \mu(W)/2.$$

Just like in the proof of Theorem 3.4.3, we can pass to a subsequence of the z_n and t_n such that they converge to $z \in \mathbb{S}^2$ and $t \in \mathbb{R}$. Also like in that proof, the above inequalities and the continuity of measures implies that

$$\mu(X \cap H_+(z,t)) = \mu(X)/2,$$
$$\mu(Y \cap H_+(z,t)) = \mu(Y)/2,$$
$$\mu(Z \cap H_+(z,t)) = \mu(Z)/2,$$

as desired. □

REMARK 3.4.5. Once we have the general Borsuk–Ulam theorem discussed in Remark 3.3.4, the proof of Theorem 3.4.4 will generalize word-for-word to give what is sometimes called the "Stone–Tukey theorem". It says that if $X_1, \ldots, X_n \subset \mathbb{R}^n$ are bounded measurable sets, then there exists some $z \in \mathbb{S}^{n-1}$ and $t \in \mathbb{R}$ such that $\mu(X_i \cap H_+(z,t)) = \mu(X_i)/2$ for all $1 \leq i \leq n$. □

3.5. Exercises

EXERCISE 3.1. Use the intermediate value theorem to prove the 1-dimensional Brouwer fixed point theorem: all maps $f\colon \mathbb{D}^1 \to \mathbb{D}^1$ have fixed points. □

CHAPTER 4

Paths and the fundamental group

As is suggested by our results about lifting paths (Lemma 2.3.1) and homotopies (Lemma 2.5.1), the structure of covers of a space X is governed by the collection of homotopy classes of paths between points of X. In this section, we collect these paths in an algebraic object called the fundamental group.

4.1. Homotopies of paths

Let X be a space and let $p, q \in X$. Recall from §2.3 that a *path* in X from p to q is a map $\gamma\colon I \to X$ with $\gamma(0) = p$ and $\gamma(1) = q$.

4.1.1. Homotopies. We wish to study paths up to homotopy. This would be uninteresting if we allowed the endpoints of γ to move during the homotopy since then all paths would be homotopic if X is path-connected (see Exercise 4.1). We therefore make the following definition:

Definition 4.1.1. Let X be a space, let $p, q \in X$, and let $\gamma_0, \gamma_1\colon I \to X$ be two paths from p to q. We say that γ_0 and γ_1 are *homotopic* paths from p to q if there exists a homotopy $\gamma_t\colon I \to X$ from γ_0 to γ_1 that fixes the endpoints in the sense that

$$\gamma_t(0) = p \quad \text{and} \quad \gamma_t(1) = q \quad \text{for all } t \in I. \qquad \square$$

4.1.2. Examples. Here are two examples:

Example 4.1.2. For all $p, q \in \mathbb{R}^n$, there is a unique homotopy class of path from p to q. Indeed, let $\gamma_0\colon I \to \mathbb{R}^n$ be the straight line path

$$\gamma_0(s) = (1-s)p + sq \quad \text{for } s \in I.$$

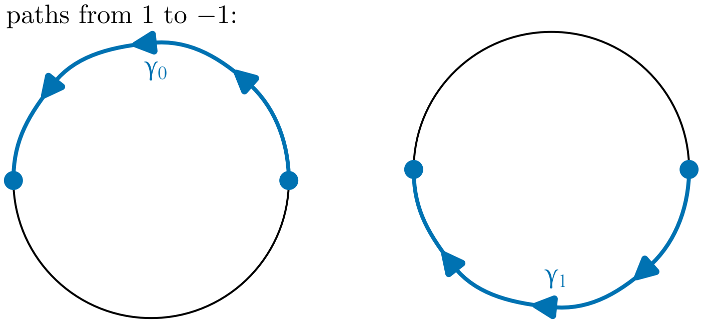
If $\gamma\colon I \to \mathbb{R}^n$ is any other path, then γ_0 is homotopic to γ via the homotopy $\gamma_t\colon I \to \mathbb{R}^n$ defined by

$$\gamma_t(s) = (1-t)\gamma_0(s) + t\gamma(s) \quad \text{for all } s \in I \text{ and } t \in I. \qquad \square$$

Example 4.1.3. View $\mathbb{S}^1$ as a subspace of $\mathbb{C}$, and let $\gamma_0\colon I \to \mathbb{S}^1$ and $\gamma_1\colon I \to \mathbb{S}^1$ be the paths defined by the formulas

$$\gamma_0(s) = e^{\pi i s} \quad \text{and} \quad \gamma_1(s) = e^{-\pi i s} \quad \text{for } s \in I. \tag{4.1.1}$$

Both γ_0 and γ_1 are paths from 1 to -1:



We claim that γ_0 and γ_1 are not homotopic. To see this, assume that they are homotopic and that $\gamma_t\colon I \to \mathbb{S}^1$ is a homotopy. Let $p\colon \mathbb{R} \to \mathbb{S}^1$ be the universal cover, so $p(\theta) = e^{2\pi i\theta}$ for all $\theta \in \mathbb{R}$. The lift of γ_0 to $\mathbb{R}$ starting at 0 is the map $\widetilde{\gamma}_0\colon I \to \mathbb{R}$ defined by

$$\widetilde{\gamma}_0(s) = s/2 \quad \text{for } s \in I. \tag{4.1.2}$$

By Lemma 2.5.1, we can lift the homotopy γ_t to a homotopy $\widetilde{\gamma}_t\colon I \to \mathbb{R}$ with $\widetilde{\gamma}_0$ the map (4.1.2). Since $\gamma_t(0) = 1$ and $\gamma_t(1) = -1$ for all $t \in I$, we have that

$$\widetilde{\gamma}_t(0) \in p^{-1}(1) = 2\pi\mathbb{Z} \quad \text{and} \quad \widetilde{\gamma}_t(1) \in p^{-1}(-1) = 2\pi\mathbb{Z} + \pi \quad \text{for } t \in I.$$

Since $2\pi\mathbb{Z}$ is discrete, it follows that both $\widetilde{\gamma}_t(0)$ and $\widetilde{\gamma}_t(1)$ are constant functions of t, i.e., that $\widetilde{\gamma}_t(0) = 0$ and $\widetilde{\gamma}_t(1) = 1/2$ for all $t \in I$. This implies in particular that $\widetilde{\gamma}_1$ is the lift of γ_1 to $\mathbb{R}$ starting at 0. From (4.1.1), we see that

$$\widetilde{\gamma}_1(s) = -s/2 \quad \text{for } s \in I.$$

In particular, $\widetilde{\gamma}_1(1) = -1/2$, contradicting the fact that $\widetilde{\gamma}_1(1) = 1/2$. □

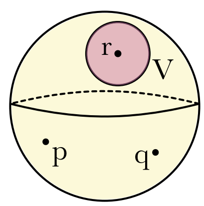
4.1.3. Spheres and general position. To illustrate an important tool for understanding paths, we prove the following, which the reader should contrast with the previous example:

Lemma 4.1.4. *Let $n \geq 2$ and let $p, q \in \mathbb{S}^n$. Then there is a unique homotopy class of paths from p to q.*

Proof. Let $r \in \mathbb{S}^n$ be a point with $r \neq p, q$. Since $\mathbb{S}^n \setminus r \cong \mathbb{R}^n$, it follows from Example 4.1.2 that there exists a unique homotopy class of path from p to q in $\mathbb{S}^n \setminus r$. Letting γ be a path from p to q in $\mathbb{S}^n$, to prove the claim it is enough to prove that γ can be homotoped into $\mathbb{S}^n \setminus r$. This is nontrivial since there do exist space-filling curves in $\mathbb{S}^n$.

One way to do this is to use smooth manifold techniques. Indeed, it follows from standard results that γ can be homotoped to a smooth map that is transverse to r. The point r is 0-dimensional, and thus is a codimension n submanifold of $\mathbb{S}^n$. It follows that $\gamma^{-1}(r)$ is a codimension $n \geq 2$ submanifold of I, and thus that $\gamma^{-1}(r) = \emptyset$.

Here is another approach that avoids using any technology. Let $V \cong \mathbb{R}^n$ be a small open neighborhood of r with $p, q \notin V$:



The subspace $V \cong \mathbb{R}^n$ is simply-connected (Example 4.1.2), and $V \setminus r$ is path-connected. Intuitively, we should be able to make a small homotopy to the portions of γ that pass through V to make them miss r. Indeed, this is what the smooth manifold approach in the previous paragraph did. Lemma 4.1.5 below shows that this is in fact possible even in more general settings where smooth manifold techniques are unavailable. □

The above proof used the following result, which for later use we prove in more generality than is needed for spheres:

Lemma 4.1.5 (General position). *Let X be a space, let $p, q \in X$, and let γ be a path in X from p to q. Let $\{r_i\}_{i \in I}$ be a discrete*[1] *set of points of X that does not contain p or q. Assume that for each $i \in I$ there is an open neighborhood V_i of r_i such that the following hold for all $i \in I$:*

- *V_i is simply-connected; and*
- *$V_i \setminus r_i$ is path-connected; and*
- *V_i does not contain $r_{i'}$ for any $i' \in I$ with $i' \neq i$.*

Then γ can be homotoped such that its image does not contain any of the r_i..

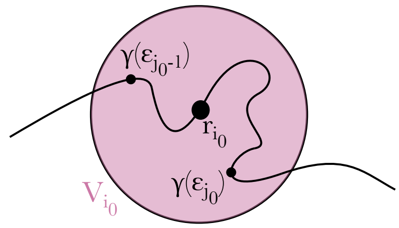
Proof. Set $U = X \setminus \{r_i \mid i \in I\}$. The set $\{U\} \cup \{V_i\}_{i \in I}$ is an open cover of X, so by the Lebesgue number lemma (cf. the proof of Lemma 2.3.1) we can find

$$0 = \epsilon_0 < \epsilon_1 < \cdots < \epsilon_k = 1$$

such that $\gamma([\epsilon_{j-1}, \epsilon_j])$ is contained in either U or some V_i for all $1 \leq j \leq k$. After possibly deleting some ϵ_j whose adjacent intervals are mapped to the same open set we can also assume that for all

[1] In other words, the subset $\{r_i \mid i \in I\}$ of X is closed and inherits the discrete topology.

$1 \leq j \leq k-1$ we have either $\gamma(\epsilon_j) \in U \cap V_i$ for some i or $\gamma(\epsilon_j) \in V_i \cap V_{i'}$ for some distinct $i, i' \in I$. This latter condition ensures that $\gamma(\epsilon_j) \notin \{r_i \mid i \in I\}$ for all $0 \leq j \leq k$. Consider some j_0 such that $\gamma([\epsilon_{j_0-1}, \epsilon_{j_0}]) \subset V_{i_0}$ for some i_0:



Since $V_i \setminus r_{i_0}$ is path-connected, there is some path δ in $V_{i_0} \setminus r_{i_i}$ from $\gamma(\epsilon_{j_0-1})$ to $\gamma(\epsilon_{j_0})$. Since V_{i_0} does not contain r_i for any $i \in I$ with $i \neq i_i$, the path δ passes through no point of $\{r_i \mid i \in I\}$. Since V_i is simply-connected, the path obtained by re-parameterizing $\gamma|_{[\epsilon_{j_0-1}, \epsilon_{j_0}]}$ to make its domain $I = [0,1]$ is homotopic to δ. It follows that we can homotope γ to change $\gamma|_{[\epsilon_{j_0-1}, \epsilon_{j_0}]}$ to a suitable re-parametrization of δ. This ensures that the image of $\gamma|_{[\epsilon_{i_0-1}, \epsilon_{i_0}]}$ is disjoint from $\{r_i \mid i \in I\}$. Doing this repeatedly homotopes γ to a path that avoids $\{r_i \mid i \in I\}$, as desired. □

4.1.4. Simple connectivity. We say that a space X is *simply-connected* or 1-*connected* if X is nonempty and path-connected, and for all $p, q \in X$ there is a unique homotopy class of paths from p to q. The following summarizes Example 4.1.3 and Lemma 4.1.4 in this language:

LEMMA 4.1.6. *The sphere $\mathbb{S}^n$ is simply-connected for $n \geq 2$. However, $\mathbb{S}^1$ is not simply-connected.*

4.2. Composing paths

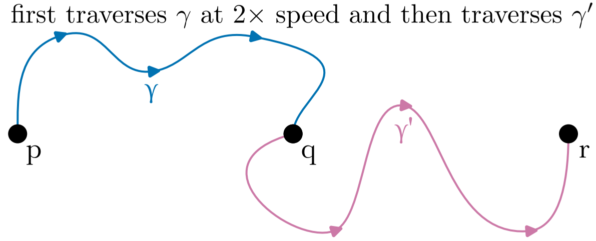
Let X be a space. Our goal is to endow the set of homotopy classes of paths between points of X with an algebraic structure.

4.2.1. Composition. In this structure, only some paths can be "multiplied". The definition is as follows:

DEFINITION 4.2.1. Let $\gamma\colon I \to X$ and $\gamma'\colon I \to X$ be paths between points of X. We say that γ and γ' are *composable* if the terminal point of γ equals the initial point of γ'. If γ and γ' are composable, then $\gamma \cdot \gamma'\colon I \to X$ is the path defined by the formula

$$(\gamma \cdot \gamma')(s) = \begin{cases} \gamma(2s) & \text{if } 0 \leq s \leq 1/2, \\ \gamma'(2s-1) & \text{if } 1/2 \leq s \leq 1. \end{cases} \quad \text{for } s \in I. \qquad \square$$

In other words, $\gamma \cdot \gamma'$ first traverses γ at $2\times$ speed and then traverses γ' at $2\times$ speed:



If γ goes from p to q and γ' goes from q to r, then $\gamma \cdot \gamma'$ goes from p to r. This only makes sense if γ and γ' are composable, and we do not define $\gamma \cdot \gamma'$ if they are not.

REMARK 4.2.2. Being composable is *not* symmetric: if $\gamma \cdot \gamma'$ is defined, then it need not be the case that $\gamma' \cdot \gamma$ is defined. □

4.2.2. Homotopies. The relation of being homotopic is an equivalence relation on the set of paths between points of X. For such a path $\gamma\colon I \to X$, let $[\gamma]$ denote its equivalence class under this relation. The following lemma says that our "multiplication" descends to a multiplication on homotopy classes:

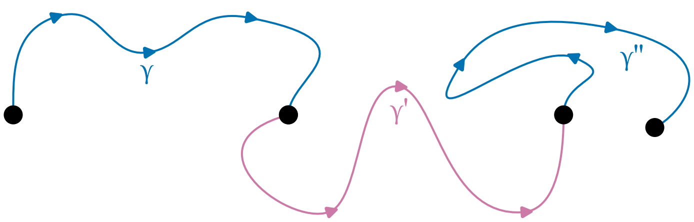
LEMMA 4.2.3. *Let X be a space. Let γ_0 and γ_0' be composable paths in X. Let γ_1 be a path that is homotopic to γ_0 and let γ_1' be a path that is homotopic to γ_0', so $[\gamma_0] = [\gamma_1]$ and $[\gamma_0'] = [\gamma_1']$. Then $[\gamma_0 \cdot \gamma_0'] = [\gamma_0' \cdot \gamma_1']$.*

PROOF. Assume that γ_0 goes from p to q and that γ'_0 goes from q to r. Let γ_t be a homotopy from γ_0 to γ_1 and let γ'_t be a homotopy from γ'_0 to γ'_1. For each $t \in I$, we have

$$\gamma_t(0) = p \quad \text{and} \quad \gamma_t(1) = \gamma'_t(0) = q \quad \text{and} \quad \gamma'_t(1) = r,$$

so γ_t and γ'_t are composable and $\gamma_t \cdot \gamma'_t$ is a well-defined path from p to r. As t varies over I, the paths $\gamma_t \cdot \gamma'_t$ form a homotopy from $\gamma_0 \cdot \gamma'_0$ to $\gamma_1 \cdot \gamma'_1$. □

4.2.3. Associativity. Let γ and γ' and γ'' be paths in X such that γ and γ' are composable and γ' and γ'' are composable. It follows that $\gamma \cdot \gamma'$ and γ'' are composable, and also that γ and $\gamma' \cdot \gamma''$ are composable:



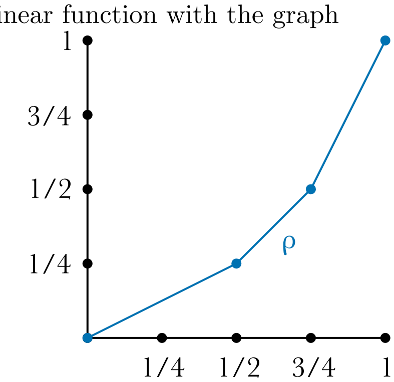
Both $(\gamma \cdot \gamma') \cdot \gamma''$ and $\gamma \cdot (\gamma' \cdot \gamma'')$ thus make sense; however except in degenerate cases we have $(\gamma \cdot \gamma') \cdot \gamma'' \neq \gamma \cdot (\gamma' \cdot \gamma'')$. The following lemma shows that passing to homotopy fixes this:

LEMMA 4.2.4. *Let X be a space. Let γ and γ' and γ'' be paths in X such that γ and γ' are composable and γ' and γ'' are composable. Then $[(\gamma \cdot \gamma') \cdot \gamma''] = [\gamma \cdot (\gamma' \cdot \gamma'')]$.*

PROOF. The paths $f_1 = (\gamma \cdot \gamma') \cdot \gamma''$ and $f_2 = \gamma \cdot (\gamma' \cdot \gamma'')$ are almost the same. They both traverse γ and then γ' and then γ''; however, they do this at different speeds. As functions on $I = [0, 1]$, we have the following:

- The path f_1 traverses γ at $4\times$ speed on the interval $[0, 1/4]$, then γ' at $4\times$ speed on the interval $[1/4, 1/2]$, and then γ'' at $2\times$ speed on the interval $[1/2, 1]$.
- The path f_2 traverses γ at $2\times$ speed on the interval $[0, 1/2]$, then γ' at $4\times$ speed on the interval $[1/2, 3/4]$, and then γ'' at $4\times$ speed on the interval $[3/4, 1]$.

Let $\rho \colon I \to I$ be the piecewise linear function with the graph



We then have $f_2 = f_1 \circ \rho$. The lemma now follows from Lemma 4.2.5 below. □

LEMMA 4.2.5 (Reparameterization lemma). *Let X be a space and $\gamma \colon I \to X$ be a path. Let $\rho \colon I \to I$ be a function such that $f(0) = 0$ and $f(1) = 1$. Then $[\gamma \circ \rho] = [\gamma]$.*

PROOF. The desired homotopy from $\gamma \circ \rho$ to γ is given by

$$\gamma_t(s) = \gamma((1-t)\rho(s) + ts) \quad \text{for } t, s \in I.$$

Here we use the fact that $f(0) = 0$ and $f(1) = 1$ to ensure that the endpoints of γ_t do not move:

$$\gamma_t(0) = \gamma((1-t)\rho(0) + 0) = \gamma(0) \quad \text{and} \quad \gamma_t(1) = \gamma((1-t)\rho(1) + t) = \gamma(1 - t + t) = \gamma(1). \qquad \square$$

4.2.4. Identity. For a point $p \in X$, let $\mathfrak{c}_p \colon I \to X$ be the constant path

$$\mathfrak{c}_p(s) = p \quad \text{for } s \in I.$$

This serves as an identity for our multiplication. However, since we can only multiply composable paths an appropriate $\mathfrak{c}_p$ must be chosen for the left- and right-identities of any given path:

LEMMA 4.2.6. *Let X be a space and let γ be a path in X from p to q. Then $[\gamma \cdot \mathfrak{c}_p] = [\gamma]$ and $[\mathfrak{c}_q \cdot \gamma] = [\gamma]$.*

PROOF. The path $\gamma \cdot \mathfrak{c}_p$ stays at p on the interval $[0, 1/2]$ and then traverses γ at $2\times$ speed:

$$(\gamma \cdot \mathfrak{c}_p)(s) = \begin{cases} p & \text{if } s \in [0,1/2], \\ \gamma(2s-1) & \text{if } s \in [1/2, 1]. \end{cases} \quad \text{for } s \in I.$$

Letting $\rho \colon I \to I$ be the map

$$\rho(s) = \begin{cases} 0 & \text{if } s \in [0,1/2], \\ 2s-1 & \text{if } s \in [1/2, 1] \end{cases} \quad \text{for } s \in I,$$

we thus have $\gamma \cdot \mathfrak{c}_p = \gamma \circ \rho$. Applying Lemma 4.2.5, we see that $[\gamma \cdot \mathfrak{c}_p] = [\gamma \circ \rho] = [\gamma]$, as desired. The proof that $[\mathfrak{c}_q \cdot \gamma] = [\gamma]$ is similar. □

4.2.5. Inverses. Let γ be a path in X from p to q. Define $\overline{\gamma} \colon I \to X$ to be the path that traverses γ in the reverse order:

$$\overline{\gamma}(s) = \gamma(1-s) \quad \text{for } s \in I.$$

The path $\overline{\gamma}$ goes from q to p, and serves as a sort of "inverse" to our multiplication:

LEMMA 4.2.7. *Let X be a space and let γ be a path in X from p to q. Then $[\gamma \cdot \overline{\gamma}] = [\mathfrak{c}_p]$ and $[\overline{\gamma} \cdot \gamma] = [\mathfrak{c}_q]$.*

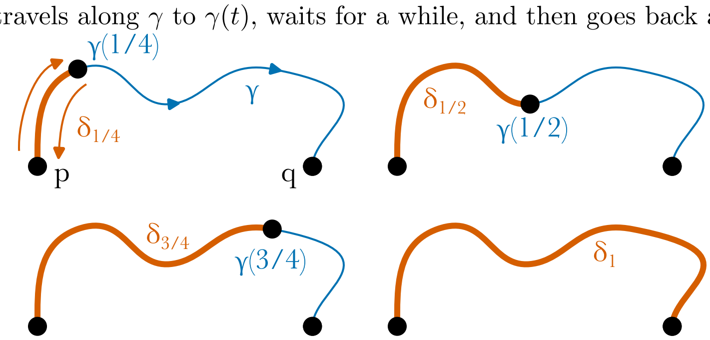
PROOF. The path $\gamma \cdot \overline{\gamma}$ goes from p to p. For $t \in I$, define $\delta_t \colon I \to X$ to be the path

$$\delta_t(s) = \begin{cases} \gamma(2s) & \text{if } s \in [0, t/2], \\ \gamma(t) & \text{if } s \in [t/2, 1-t/2], \\ \gamma(2(1-s)) & \text{if } s \in [1-t/2, 1]. \end{cases} \quad \text{for } s \in I.$$

This makes sense since

$$\gamma(2(t/2)) = \gamma(t) = \gamma(2(1-(1-t/2))).$$

Geometrically, δ_t travels along γ to $\gamma(t)$, waits for a while, and then goes back along $\overline{\gamma}$:



Since δ_t is a homotopy from $\mathfrak{c}_p$ to $\gamma \cdot \overline{\gamma}$, we deduce that $[\mathfrak{c}_p] = [\gamma \cdot \overline{\gamma}]$, as desired. The proof that $[\overline{\gamma} \cdot \gamma] = [\mathfrak{c}_q]$ is similar. □

4.3. Fundamental group and groupoid

Let X be a space. In the previous section, we showed that the set of homotopy classes of paths between points of X has a partially-defined "multiplication" that is associative, has units, and has inverses. What kind of algebraic structure could this be?

4.3.1. Categories. To answer this question, we need the language of category theory. Recall that a category $\mathbf{C}$ consists of the following data:

- A collection of objects. We will write $A \in \mathbf{C}$ to indicate that A is an object of $\mathbf{C}$.
- For all objects $A, B \in \mathbf{C}$, a set $\mathbf{C}(A, B)$ of morphisms. We will often write $f\colon A \to B$ to indicate that f is a morphism from A to B.
- For all objects $A \in \mathbf{C}$, an identity morphism $\mathbb{1}_A\colon A \to A$.

These morphisms can be composed: if $f\colon A \to B$ and $g\colon B \to C$ are morphisms, then we have a morphism $g \circ f\colon A \to C$. This composition should be associative in the sense that if $f\colon A \to B$ and $g\colon B \to C$ and $h\colon C \to D$ are morphisms, then

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

Because of this, there is no need to insert parentheses when composing morphisms. Under this composition, the identity morphisms should be units: if $f\colon A \to B$ is a morphism, then $f \circ \mathbb{1}_A = f$ and $\mathbb{1}_B \circ f = f$.

Example 4.3.1. The collection of all topological spaces and continuous maps forms a category Top. □

Example 4.3.2. The collection of all groups and homomorphisms forms a category Group. □

Example 4.3.3. For a group G, there is a category (also written G) with one object $*$ and with $G(*, *) = G$. □

Remark 4.3.4. The language of category theorem might seem overly abstract, but it turns out to be very useful and clarifying. Fundamentally, it is just a way of organizing information. Typically you cannot prove interesting new theorems by just defining a category, but the language of category theory often suggests useful constructions. □

4.3.2. Fundamental groupoid. We now define the following:

Definition 4.3.5. Let X be a space. The *fundamental groupoid* of X, denoted $\Pi(X)$, is the following category:

- The objects of $\Pi(X)$ are the points of X.
- For points p and q, the $\Pi(X)$-morphisms from p to q are the set of all homotopy classes of paths from p to q. For a path γ from p to q, we will write $[\gamma]\colon p \to q$ for the corresponding morphism from p to q.
- If γ is a path from p to q and γ' is a path from q to r, then the composition of the morphisms $[\gamma]\colon p \to q$ and $[\gamma']\colon q \to r$ is the morphism $[\gamma \cdot \gamma']\colon p \to r$.
- For a point $p \in X$, the identity morphism of p is the constant path $[\mathfrak{c}_p]\colon p \to p$. □

Remark 4.3.6. We now come to an annoying technical point: in a category $\mathbf{C}$, we said that the composition of morphisms $f\colon A \to B$ and $g\colon B \to C$ is the morphism $g \circ f\colon A \to C$. Writing it like this makes sense since in many natural categories the objects are sets with some kind of structure and the morphisms are functions that preserve that structure. However, the objects of the fundamental groupoid are *not* sets with some kind of structure. As is traditional, we compose the morphisms in the fundamental groupoid from left to right (rather than right to left, like functions). □

4.3.3. Fundamental group. Lemma 4.2.7 says that all the morphisms in the fundamental groupoid $\Pi(X)$ are invertible. This is the defining property of a groupoid:

Definition 4.3.7. A *groupoid* is a category $\mathbf{G}$ in which all morphisms are invertible, i.e., such that for all morphisms $\phi\colon A \to B$, there is a morphism $\overline{\phi}\colon B \to A$ with $\overline{\phi} \circ \phi = \mathbb{1}_A$ and $\phi \circ \overline{\phi} = \mathbb{1}_B$. □

Remark 4.3.8. Let $\mathbf{G}$ be a groupoid and $\phi\colon A \to B$ be a morphism in $\mathbf{G}$. In Exercise 4.2, you will prove that the inverse to ϕ is unique in the following sense. Consider $\overline{\phi}, \overline{\phi}'\colon B \to A$. Then $\phi = \overline{\phi}'$ if any of the following conditions are satisfied:

- $\overline{\phi} \circ \phi = \overline{\phi}' \circ \phi = \mathbb{1}_A$; or
- $\phi \circ \overline{\phi} = \phi \circ \overline{\phi}' = \mathbb{1}_B$; or
- $\overline{\phi} \circ \phi = \mathbb{1}_A$ and $\phi \circ \overline{\phi}' = \mathbb{1}_B$.

Because of this, we can safely talk about *the* inverse to ϕ. □

As we discussed in Example 4.3.3, a group can be viewed as a category with one object. Under this identification, a group is a groupoid. Conversely, consider a groupoid $\mathbf{G}$. For $A \in \mathbf{G}$, write

$$\mathrm{Aut}_{\mathbf{G}}(A) = \{f \mid \phi\colon A \to A \text{ is a morphism in } \mathbf{G}\}.$$

Since all morphisms in $\mathbf{G}$ are invertible, this is a group. What is more, for a morphism $\psi\colon A \to B$ in $\mathbf{G}$ there is an isomorphism $\psi_*\colon \mathrm{Aut}_{\mathbf{G}}(A) \to \mathrm{Aut}_{\mathbf{G}}(B)$ defined by

$$\psi_*(\phi) = \psi \circ \phi \circ \overline{\psi} \quad \text{for all } \phi\colon A \to A \text{ in } \mathrm{Aut}_{\mathbf{G}}(A).$$

In this way, a groupoid packages together a collection of groups along with certain isomorphisms between them.

We now return to the fundamental groupoid. For $x_0 \in X$ the *fundamental group* of X with basepoint x_0, denoted $\pi_1(X, x_0)$, is

$$\pi_1(X, x_0) = \mathrm{Aut}_{\Pi(X)}(x_0).$$

In other words, $\pi_1(X, x_0)$ is the group whose objects are homotopy classes of *loops based at* x_0, i.e., paths γ from x_0 to itself. If α is a path from x_0 to x_0', then we get an isomorphism

$$\alpha_*\colon \pi_1(X, x_0') \to \pi_1(X, x_0)$$

defined by

$$\alpha_*([\gamma]) = [\alpha{\cdot}\gamma{\cdot}\overline{\alpha}] \quad \text{for all } [\gamma] \in \pi_1(X, x_0').$$

From these isomorphisms, we see the following:

Lemma 4.3.9. *Let X be a path-connected space. Then for all $x_0, x_0' \in X$ we have $\pi_1(X, x_0) \cong \pi_1(X, x_0')$.*

Proof. Just use the above isomorphism associated to a path from x_0 to x_0'. □

4.3.4. Commentary. We will give many computations of $\pi_1(X, x_0)$ over the next few sections. For X path-connected, Lemma 4.3.9 says that the isomorphism type of $\pi_1(X, x_0)$ is independent of the basepoint x_0. The isomorphism type of $\pi_1(X, x_0)$ is thus a useful invariant of path-connected spaces, i.e., if two path-connected spaces have different fundamental groups, then they are not homeomorphic. The fundamental groupoid is not so useful as an invariant since it knows far too much about the space; for instance, its objects are literally the points of the space.

You might wonder why we bothered to introduce the fundamental groupoid at all. There are two reasons:

- While for a path-connected space the isomorphism type of the fundamental group does not depend on the basepoint, the isomorphisms between the fundamental groups at different basepoints are not canonical. The fundamental groupoid packages them all together, and is present at least implictly in all serious treatements of the fundamental group. It seems perverse to refuse to give a name to a structure you use.
- There are many constructions in topology that are most naturally phrased in terms of the fundamental groupoid. For instance, the most general form of the classification of covering spaces uses the fundamental groupoid (see §YYY). Later volumes of this book will contain other examples.

We remark that serious applications of $\pi_1(X, x_0)$ often require a careful treatment of the basepoint x_0. Simply identifying the fundamental group at different basepoints will quickly lead you astray. This is analogous to the fact that while all finite-dimensional vector spaces over a field $\mathbf{k}$ are isomorphic to $\mathbf{k}^n$ for some $n \geq 0$, one cannot simply identify vector spaces with $\mathbf{k}^n$. Such an identification requires a choice of basis, and often there is no natural choice. Much of linear algebra focuses on carefully choosing bases adapted to different situations and studying how all these different bases are related.

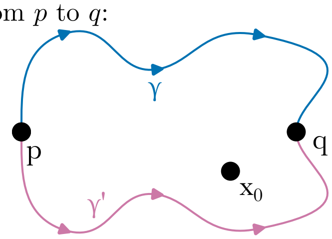
4.4. Simple connectivity, deformation retracts, and contractibility

Let X be a space and $x_0 \in X$. We start our discussion of ways to calculate $\pi_1(X, x_0)$ by discussing cases where it is trivial.

4.4.1. Simple connectivity. Recall that X is said to be simply-connected or 1-connected if X is nonempty and path-connected, and there is a unique homotopy class of paths between any two points of X. This has a nice interpretation in terms of the fundamental group:[2]

LEMMA 4.4.1. *Let X be a path-connected space and let $x_0 \in X$. Then X is simply-connected if and only if $\pi_1(X, x_0) = 1$.*

PROOF. If X is simply-connected, then in particular there is only one homotopy class of paths from x_0 to itself, so $\pi_1(X, x_0) = 1$. Conversely, assume that $\pi_1(X, x_0) = 1$. Let p and q be two points of X, and let γ and γ' be paths from p to q:



Since X is path-connected, Lemma 4.3.9 implies that $\pi_1(X, p) = 0$. The path $\gamma \cdot \overline{\gamma}'$ is a path from p to p, so $[\gamma \cdot \overline{\gamma}'] \in \pi_1(X, p)$ must be trivial. We therefore have

$$[\gamma'] = 1[\gamma'] = [\gamma \cdot \overline{\gamma}'][\gamma'] = [\gamma][\overline{\gamma}'][\gamma'] = [\gamma],$$

as desired. □

This implies:

LEMMA 4.4.2. *For $n \geq 2$ and $x_0 \in \mathbb{S}^n$ arbitrary, we have $\pi_1(\mathbb{S}^n, x_0) = 1$.*

PROOF. We proved in Lemma 4.1.6 that $\mathbb{S}^n$ is simply-connected for $n \geq 2$, so by Lemma 4.4.1 we deduce that $\pi_1(\mathbb{S}^n, x_0) = 1$ for all $x_0 \in \mathbb{S}^n$. □

REMARK 4.4.3. We proved in Lemma 4.1.6 that $\mathbb{S}^1$ is *not* simply-connected, so $\pi_1(\mathbb{S}^1, x_0) \neq 1$. In fact, we will later show that $\pi_1(\mathbb{S}^1, x_0) \cong \mathbb{Z}$ via an analysis similar to the one we used to prove that $\mathbb{S}^1$ is not simply-connected. □

REMARK 4.4.4. Let X be a path-connected space and let $p, q \in X$ be points. Let $\mathcal{P}(p, q)$ be the set of morphisms in $\Pi(X)$ from p to q, i.e., the set of homotopy classes of paths in X from p to q. The proof of Lemma 4.4.1 shows how to relate $\mathcal{P}(p, q)$ to $\pi_1(X, p)$. Indeed, $\pi_1(X, p)$ acts on the set $\mathcal{P}(p, q)$ via

$$[\gamma][\delta] = [\gamma \cdot \delta] \quad \text{for } [\gamma] \in \pi_1(X, p) \text{ and } [\delta] \in \mathcal{P}(p, q).$$

This action has two key properties, both of which follow from easy calculations in the fundamental groupoid $\Pi(X)$:

- It is *faithful*, i.e., for $[\gamma] \in \pi_1(X, p)$ and $[\delta] \in \mathcal{P}(p, q)$ if $[\gamma][\delta] = [\delta]$ then $[\gamma] = 1$. Indeed:
$$[\gamma] = [\gamma \cdot \delta \cdot \overline{\delta}] = [\gamma \cdot \delta][\overline{\delta}] = [\delta][\overline{\delta}] = 1.$$
- It is *transitive*, i.e., for all $[\delta], [\delta'] \in \mathcal{P}(p, q)$ there exists some $[\gamma] \in \pi_1(X, p)$ with $[\gamma][\delta] = [\delta']$. Indeed, just take $\gamma = \delta' \cdot \overline{\delta}$:
$$[\delta' \cdot \overline{\delta}][\delta] = [\delta'][\overline{\delta} \cdot \delta] = [\delta'].$$

Fixing some $[\delta_0] \in \mathcal{P}(p, q)$, there is thus a bijection between $\pi_1(X, p)$ and $\mathcal{P}(p, q)$ taking $[\gamma] \in \pi_1(X, p)$ to $[\gamma \cdot \delta_0] \in \mathcal{P}(p, q)$. One might think that this bijection allows us to turn $\mathcal{P}(p, q)$ into a group; however, this is not a good idea since the bijection depends on the choice of $[\delta_0]$, and is thus not canonical. Instead, this makes $\mathcal{P}(p, q)$ into what is called a *torsor* for the group $\pi_1(X, p)$. □

[2] This lemma explains why this is sometimes called being 1-connected. In Chapter YYY we will define groups $\pi_n(X, x_0)$ for all $n \geq 1$, and a nonempty path-connected space is said to be n-connected if $\pi_d(X, x_0) = 0$ for $d \leq n$.

4.4.2. Deformation retracts. Let $A \subset X$ be a subspace. A *retract* of X to A is a map $r\colon X \to A$ such that $r(a) = a$ for all $a \in A$, i.e., such that $r|_A = \mathbb{1}$. A *deformation retraction* of X to A is a homotopy $r_t\colon X \to X$ from the identity $\mathbb{1}\colon X \to X$ to a map $r_1\colon X \to X$ such that:

- the map r_1 is a retraction of X to A; and
- for all $t \in I$ and $a \in A$, we have $r_t(a) = a$.

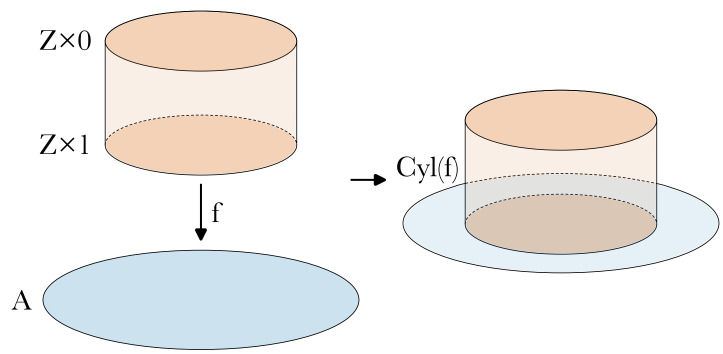
If there exists a deformation retraction of X to A, then we say that A is a *deformation retract* of X and that X *deformation retracts* to A. Here are several examples:

EXAMPLE 4.4.5. Let $U \subset \mathbb{R}^n$ be a set that is *star-shaped*, i.e., such that there exists a point $p_0 \in U$ such that for all $x \in U$ the line segment from p_0 to x is contained in U. For instance, U might be convex. We claim that U deformation retracts to p_0. Indeed, the maps $r_t\colon U \to U$ defined by

$$r_t(x) = (1-t)x + tp_0 \quad \text{for } x \in U \text{ and } t \in I$$

form a deformation retraction. □

EXAMPLE 4.4.6. Let $f\colon Z \to A$ be a map between spaces. The *mapping cylinder* of f, denoted $\mathrm{Cyl}(f)$, is the quotient of the disjont union $(Z \times I) \sqcup A$ that identifies $(z,1) \in Z \times I$ with $f(z) \in A$ for all $z \in Z$:

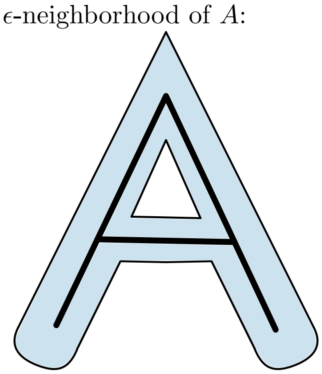


For $z \in Z$ and $s \in I$, let $\overline{(z,s)}$ be the image of $(z,s) \in Z \times I$ in $\mathrm{Cyl}(f)$. The space $\mathrm{Cyl}(f)$ deformation retracts to A via the deformation retract $r_t\colon\ \mathrm{Cyl}(f) \to A$ defined by

$$\begin{cases} r_t\overline{(z,s)} = \overline{(z,(1-t)s)} & \text{for } (z,s) \in Z \times [0,1], \\ r_t(a) = a & \text{for } a \in A. \end{cases}$$

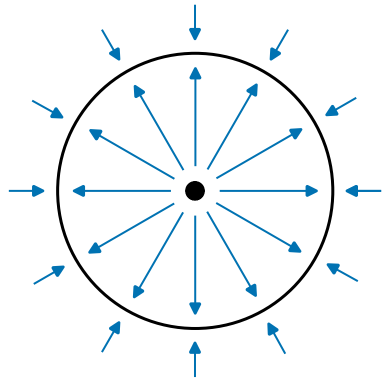
The reader can easily check that this makes sense and is continuous. □

EXAMPLE 4.4.7. As in the following figure, let A be the letter A embedded in the plane and for some small $\epsilon > 0$ let X be a closed ϵ-neighborhood of A:



Then X deformation retracts to A via a deformation retraction during which points travel along straight line segments to A. In fact, this is a special case of the previous example: the boundary of X consists of two circles $\mathbb{S}^1 \sqcup \mathbb{S}^1$, and X is homeomorphic to the mapping cylinder of a map $f\colon \mathbb{S}^1 \sqcup \mathbb{S}^1 \to A$. □

EXAMPLE 4.4.8. We claim that $\mathbb{S}^{n-1}$ is a deformation retract of $\mathbb{R}^n \setminus 0$. Geometrically, the picture is as follows, where the blue arrows show the paths points of $\mathbb{R}^n \setminus 0$ travel during the deformation retraction:



In formulas, this deformation retraction is given by the maps $r_t \colon \mathbb{R}^n \setminus 0 \to \mathbb{R}^n \setminus 0$ defined by

$$r_t(x) = \left((1-t) + \frac{t}{\|x\|}\right) x \quad \text{for } x \in \mathbb{R}^n \setminus 0 \text{ and } t \in I. \qquad \square$$

Assume now that $A \subset X$ is a subspace and $a_0 \in A$. Loops in A based at a_0 can also be regarded as loops in X based at a_0, and if two loops in A based at p are homotopic in A then are also homotopic in X. This gives a natural homomorphism[3] $\pi_1(A, a_0) \to \pi_1(X, a_0)$. This homomorphism need not be injective since it is possible for two loops in A based at a_0 to be homotopic in X but not in A. However:

LEMMA 4.4.9. *Let X be a space, let $A \subset X$ be a subspace, and let $a_0 \in A$. Then:*

(i) If A is a retract of X, then the map $\pi_1(A, a_0) \to \pi_1(X, a_0)$ is injective.

(ii) If A is a deformation retraction of X, then the map $\pi_1(A, a_0) \to \pi_1(X, a_0)$ is an isomorphism.

PROOF. We start with (i). Let $r \colon X \to A$ be a retraction. Let $[\gamma] \in \pi_1(A, a_0)$ lie in the kernel of the map $\pi_1(A, a_0) \to \pi_1(X, a_0)$. There is thus a homotopy $\gamma_t \colon I \to X$ with $\gamma_0 = \gamma$ and $\gamma_1 = \mathfrak{c}_{a_0}$. Since $r(a_0) = a_0$, the homotopy $r \circ \gamma_t \colon I \to A$ is also a homotopy of paths from γ to $\mathfrak{c}_{a_0}$, so $[\gamma]$ is trivial in $\pi_1(A, a_0)$, as desired.

We now prove (ii). Let $r_t \colon X \to X$ be a deformation retract. In light of (i), it is enough to prove that the map $\pi_1(A, a_0) \to \pi_1(X, a_0)$ is surjective. Consider a loop $\delta \colon I \to X$ based at p. We must prove that δ can be homotoped to a loop lying in A. Since $r_t(a_0) = a_0$ for all $t \in I$, we have a homotopy of paths $r_t \circ \delta$. Since $r_1(X) \subset A$, the image of the endpoint $r_1 \circ \delta$ of this homotopy lies in A, as desired. $\square$

Here is one consequence:

LEMMA 4.4.10. *Let $n \geq 3$. For $x_0 \in \mathbb{R}^n \setminus 0$, we have $\pi_1(\mathbb{R}^n \setminus 0, x_0) = 0$.*

PROOF. Since $\mathbb{R}^n \setminus 0$ is path-connected, we can change x_0 without changing the fundamental group. Choose x_0 such that $x_0 \in \mathbb{S}^{n-1} \subset \mathbb{R}^n \setminus 0$. Since $\mathbb{R}^n \setminus 0$ deformation retracts to $\mathbb{S}^{n-1}$, Lemma 4.4.9 implies that

$$\pi_1(\mathbb{R}^n \setminus 0, x_0) \cong \pi_1(\mathbb{S}^{n-1}, x_0).$$

Since $n \geq 3$, this vanishes by Lemma 4.4.2. $\square$

4.4.3. Contractibility. A nonempty space X is said to be *contractible* if the identity map $\mathbb{1} \colon X \to X$ is homotopic to a constant map. This holds, for instance, if X deformation retracts to any one-point subspace x_0. Star-shaped or convex subspaces of $\mathbb{R}^n$ are therefore contractible. However, being contractible is more general than this since none of the points of X need to be fixed during the contraction. See Exercise 4.4 for an example where these are genuinely different notions.

If a space X deformation retracts to a point $x_0 \in X$, then it follows from Lemma 4.4.9 that

$$\pi_1(X, x_0) \cong \pi_1(x_0, x_0) = 1.$$

[3] This is an instance of the functorality of the fundamental group, which we will discuss more carefully in §4.6.

The following shows that this vanishing holds more generally if X is merely contractible. Contrasting the proof of this with that of Lemma 4.4.9 illustrates the care that must sometimes be taken with the basepoint:

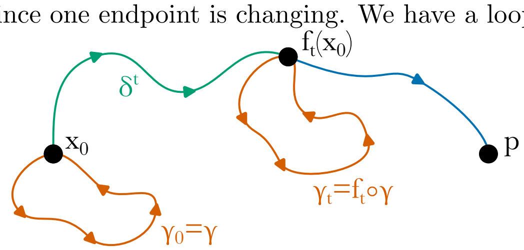
Lemma 4.4.11. *Let X be a contractible space and let $x_0 \in X$. Then $\pi_1(X, x_0) = 1$.*

Proof. Let $f_t \colon X \to X$ be a homotopy from the identity $\mathbb{1} \colon X \to X$ to a constant map. Let $p \in X$ be the constant value of f_1. Consider $[\gamma] \in \pi_1(X, x_0)$ We must prove that γ is homotopic to a constant path. A first idea is to consider the loop $\gamma_t = f_t \circ \gamma$. However, this does not work since $\gamma_t(0) = \gamma_t(1)$ need not equal x_0, and in fact as t varies in I the point $\gamma_t(0) = \gamma_t(1)$ traverses a path from x_0 to $f_1(x_0) = p$.

To fix this, for $t \in I$ let $\delta^t \colon I \to X$ be the path defined by

$$\delta^t(s) = f_{ts}(x_0) \quad \text{for all } s \in I.$$

This path goes from x to $f_t(x_0)$. We put the t in the superscript to remind the reader that this is not a homotopy of paths since one endpoint is changing. We have a loop $\delta^t \cdot \gamma_t \cdot \overline{\delta}^t$ based at x_0:



Recall that for $x \in X$ we denote by $\mathfrak{c}_x$ the constant path at x. For $t = 0$, we have

$$[\delta^0 \cdot \gamma_0 \cdot \delta^0] = [\delta^0][\gamma_0][\overline{\delta}^0] = [\mathfrak{c}_{x_0}][\gamma][\mathfrak{c}_{x_0}] = [\gamma],$$

and for $t = 1$ we have

$$[\delta^1 \cdot \gamma_1 \cdot \delta^1] = [\delta^1][\gamma_1][\overline{\delta}^1] = [\delta^1][\mathfrak{c}_p][\overline{\delta}^1] = [\delta^1 \cdot \overline{\delta}^1] = 1.$$

Using this homotopy, we thus see that $[\gamma] = 1$, as desired. □

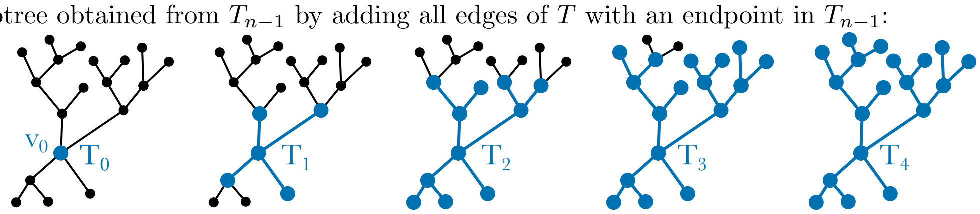
Here is another important example. Recall that we discussed graphs in §1.6. A tree is a nonempty connected graph with no cycles. We will prove:

Lemma 4.4.12. *Let T be a tree and let v_0 be a vertex of T. Then T deformation retracts to v_0, and in particular T is contractible.*

Proof. We will omit some of the point-set details, and invite the reader in Exercise 4.5 to verify that all the maps we construct are continuous. Inductively define subtrees

$$T_0 \subset T_1 \subset T_2 \subset \cdots$$

of T in the following way. Start by letting $T_0 = v_0$. Next, if T_{n-1} has been constructed, let T_n be the subtree obtained from T_{n-1} by adding all edges of T with an endpoint in T_{n-1}:



Since T is a tree, each new edge e added to T_{n-1} to form T_n has the property that exactly one endpoint of e lies in T_{n-1}; otherwise, e would form part of a cycle in T_n. This implies that T_n deformation retracts to T_{n-1} via a deformation retract where the points of these new edges e move along e to the vertex lying in T_{n-1}. Let $r_t^n \colon T_n \to T_n$ be this deformation retract. Since T is connected, we have

$$T = \bigcup_{n=0}^{\infty} T_n.$$

For each $n \geq 1$ and $m \geq 0$, consider the retractions

$$R_m^n = r_1^n \circ \cdots \circ r_1^{n+m} \colon T_{n+m} \to T_{n-1}.$$

For $m_1 \geq m_2 \geq 0$, the retractions $R_{m_1}^n$ and $R_{m_2}^n$ agree where they both are defined, namely on T_{n+m_2}. It follows that for a fixed $n \geq 1$ the different R_m^n glue together to give a retraction $R^n \colon T \to T_{n-1}$.

Assume first that $T = T_{n_1}$ for some $n_1 \gg 0$ (which holds, for instance, if T is a finite tree). In this case, we can deformation retract $T = T_{n_1}$ to $T_0 = v_0$ by first using $r_t^{n_1}$ to deformation retract T_{n_1} to T_{n_1-1}, then using $r_t^{n_1-1}$ to deformation retract T_{n_1-1} to T_{n_1-2}, etc. For the general case, we have to be a bit more careful. Write

$$I = \{0\} \cup \bigcup_{n=1}^{\infty} I_n \quad \text{with } I_n = [1/2^n, 1/2^{n-1}],$$

so I_n has length $1/2^n$. Define $r_t \colon T \to T$ in the following way:

- For $t \in I_n$ and $x \in T$, let $r_t(x) = r^n_{2^n(t-1/2^n)}(R^{n+1}(x))$.
- For $t = 0$ and $x \in T$, define $r_0(x) = x$.

The reader will check in Exercise 4.5 that this definition makes sense and is continuous. By definition we have $r_0 = \mathbb{1}$, and since $1 \in I_1$ we have

$$r_1(x) = r_1^1(R^2(x)) = v_0 \quad \text{for } x \in T,$$

where we recall that T_0 is the vertex v_0. It follows that r_t is a deformation retraction of T to v_0, as desired. □

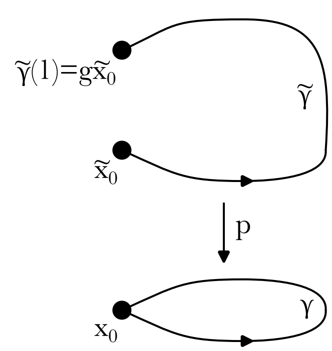
4.5. Calculating the fundamental group using covering spaces

Thus far we have not calculated any nontrivial fundamental groups. In this section, we describe the most important tool for making these calculations.

4.5.1. Regular covers and the fundamental group. Recall that we discussed regular covers in §1.5. Our main result is:

THEOREM 4.5.1. *Let $p \colon \widetilde{X} \to X$ be a regular cover with $\widetilde{X}$ simply-connected. Set $G = \mathrm{Deck}(\widetilde{X})$. Then for $x_0 \in X$ we have $\pi_1(X, x_0) \cong G$.*

PROOF. Pick $\widetilde{x}_0 \in \widetilde{X}$ with $p(\widetilde{x}_0) = x_0$. We define a set map $f \colon \pi_1(X, x_0) \to G$ as follows. Consider $[\gamma] \in \pi_1(X, x_0)$. By path lifting (Lemma 2.3.1), we can lift γ to a path $\widetilde{\gamma}$ in $\widetilde{X}$ starting at $\widetilde{x}_0$. By homotopy lifting (Lemma 2.5.1), the homotopy class of $\widetilde{\gamma}$ only depends on the homotopy class of γ. In particular, $\widetilde{\gamma}(1) \in \widetilde{X}$ only depends on $[\gamma] \in \pi_1(X, x_0)$. The point $\widetilde{\gamma}(1)$ projects to $\gamma(1) = x_0$, so $\widetilde{\gamma}(1)$ lies in the fiber $p^{-1}(x_0)$:

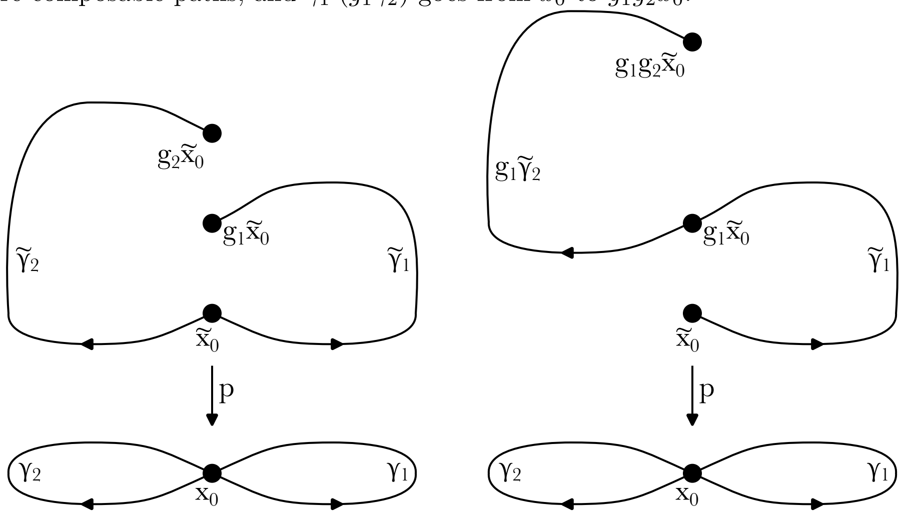


Since $\widetilde{X}$ is a path-connected regular cover of X, there exists a unique $g \in G$ with $g\widetilde{x}_0 = \widetilde{\gamma}(1)$; see Lemma 1.4.3. Define $f([\gamma]) = g$.

To prove the theorem, it is enough to prove that f is a group homomorphism and also that f is both injective and surjective. We do this in the following three claims:

CLAIM 1. *The set map f is a group homomorphism.*

Consider $[\gamma_1], [\gamma_2] \in \pi_1(X, x_0)$. For $i = 1, 2$, let $\widetilde{\gamma}_i$ be the lift of γ_i to $\widetilde{X}$ with $\widetilde{\gamma}_i(0) = \widetilde{x}_0$. Letting $g_i = f([\gamma_i])$, the path $\widetilde{\gamma}_i$ thus goes from $\widetilde{x}_0$ to $g_i\widetilde{x}_0$. The deck group G acts not only on $\widetilde{X}$, but also on paths in $\widetilde{X}$. Under this group action, the path $g_1\widetilde{\gamma}_2$ goes from $g_1\widetilde{x}_0$ to $g_1g_2\widetilde{x}_0$. It follows that $\widetilde{\gamma}_1$ and $g_1\widetilde{\gamma}_2$ are composable paths, and $\widetilde{\gamma}_1 \cdot (g_1\widetilde{\gamma}_2)$ goes from $\widetilde{x}_0$ to $g_1g_2\widetilde{x}_0$:



The path $\widetilde{\gamma}_1 \cdot (g_1\widetilde{\gamma}_2)$ is the lift of $\gamma_1 \cdot \gamma_2$, so by definition this implies that $f([\gamma_1 \cdot \gamma_2]) = g_1g_2$, as desired.

Claim 2. *The homomorphism f is surjective.*

Consider $g \in G$. Since $\widetilde{X}$ is path-connected, we can find a path $\widetilde{\gamma}$ in $\widetilde{X}$ from $\widetilde{x}_0$ to $g\widetilde{x}_0$. The path $\widetilde{\gamma}$ projects to a path γ in X from x_0 to x_0, so we have an element $[\gamma] \in \pi_1(X, x_0)$. By definition, $f([\gamma]) = g$.

Claim 3. *The homomorphism f is injective.*

Consider $[\gamma] \in \pi_1(X, x_0)$ such that $f([\gamma]) = 1$. Let $\widetilde{\gamma}$ be the lift of γ to $\widetilde{X}$ with $\widetilde{\gamma}(0) = \widetilde{x}_0$. Since $f([\gamma]) = 1$, we must have $\widetilde{\gamma}(1) = \widetilde{x}_0$, so $\widetilde{\gamma}$ is a loop based at $\widetilde{x}_0$. Since $\widetilde{X}$ is simply-connected, the loop $\widetilde{\gamma}$ is homotopic to a constant loop. Composing this homotopy with the map $p\colon \widetilde{X} \to X$, we obtain a homotopy from γ to a constant loop, so $[\gamma] = 1$, as desired. □

4.5.2. Understanding isomorphism. As in Theorem 4.5.1, let $p\colon \widetilde{X} \to X$ be a regular cover with $\widetilde{X}$ simply-connected. Let $\widetilde{x}_0 \in \widetilde{X}$, and set $G = \text{Deck}(\widetilde{X})$ and $x_0 = p(\widetilde{x}_0)$. Theorem 4.5.1 says that $\pi_1(X, x_0) \cong G$. Examining its proof, this isomorphism is as follows:

- Consider $g \in G$. Let $\widetilde{\gamma}$ be a path in $\widetilde{X}$ from $\widetilde{x}_0$ to $g\widetilde{x}_0$, and let γ be the projection of $\widetilde{\gamma}$ to X. Then the element of $\pi_1(X, x_0)$ corresponding to g is $[\gamma] \in \pi_1(X, x_0)$.

4.5.3. Examples. We now give five calculations of $\pi_1(X, x_0)$ using Theorem 4.5.1. The first is important enough to separate it out as a lemma. Recall that we identify $\mathbb{S}^1$ with a subset of $\mathbb{C}$, so $1 \in \mathbb{S}^1$.

Lemma 4.5.2 (Circle). *We have $\pi_1(\mathbb{S}^1, 1) \cong \mathbb{Z}$, where $n \in \mathbb{Z}$ corresponds to the loop $\gamma_n\colon I \to \mathbb{S}^1$ defined by*

$$\gamma_n(s) = e^{2\pi i n s} \quad \text{for } s \in I.$$

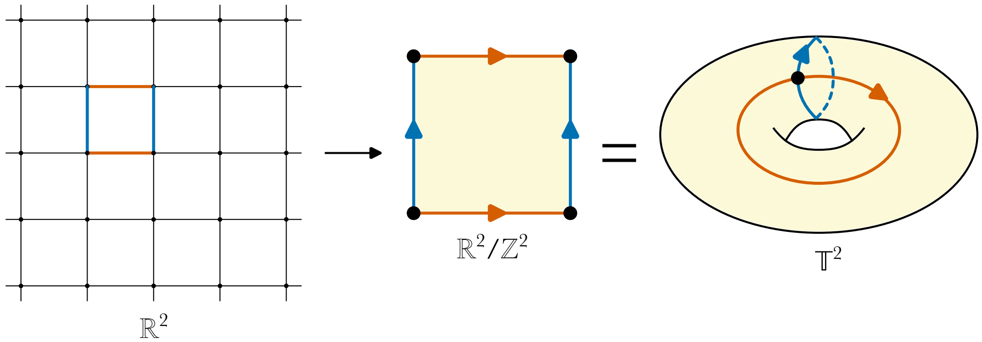
Proof. Consider the universal cover $p\colon \mathbb{R} \to \mathbb{S}^1$ of $\mathbb{S}^1$, so $p(\theta) = e^{2\pi i\theta}$. As we observed in Example 1.4.4, this is a regular cover with deck group $\mathbb{Z}$, which acts on $\mathbb{R}$ by integer translations. Since $\mathbb{R}$ is contractible, it is simply-connected. We can therefore apply Theorem 4.5.1 to see that

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

That $n \in \mathbb{Z}$ corresponds to the loop γ_n is immediate from the description of this isomorphism in §4.5.2. □

Our next example generalizes this:

Example 4.5.3 (Torus). As in Example 1.1.10 let $\mathbb{Z}^n$ act on $\mathbb{R}^n$ by integer translations and identify the quotient $\mathbb{R}^n/\mathbb{Z}^n$ with the n-dimensional torus $\mathbb{T}^n = (\mathbb{S}^1)^{\times n}$:



This figure shows the case $n = 2$. The projection $p\colon \mathbb{R}^n \to \mathbb{Z}^n$ is a regular cover with deck group $\mathbb{Z}^n$. Set $x_0 = p(0)$. Since $\mathbb{R}^n$ is contractible, it is simply-connected. We can thus apply Theorem 4.5.1 to see that

$$\pi_1(\mathbb{T}^n, x_0) = \pi_1((\mathbb{S}^1)^{\times n}, x_0) \cong \mathbb{Z}^n.$$

More generally, you will show in Exercise 4.3 that if X and Y are spaces with basepoints $x_0 \in X$ and $y_0 \in Y$, then

$$\pi_1(X \times Y, (x_0, y_0)) \cong \pi_1(X, x_0) \times \pi_1(Y, y_0). \qquad \square$$

We now turn to real projective space:

Example 4.5.4 (Real projective space). Let $n \geq 2$. Recall that $\mathbb{RP}^n$ is the space of lines through the origin in $\mathbb{R}^{n+1}$. As we described in Example 1.1.11, there is a 2-fold cover $p\colon \mathbb{S}^n \to \mathbb{RP}^n$ taking $x \in \mathbb{S}^n$ to the line through x. This is a regular cover with deck group the cyclic group C_2 of order 2, which acts on $\mathbb{S}^n$ by the antipodal map $x \mapsto -x$. Fix some $x_0 \in \mathbb{S}^n$, and let $\ell_0 = p(x_0) \in \mathbb{RP}^n$. Since $n \geq 2$, we have $\pi_1(\mathbb{S}^n, x_0) = 1$ (Lemma 4.4.2). We can therefore apply Theorem 4.5.1 and see that

$$\pi_1(\mathbb{RP}^n, \ell_0) \cong C_2.$$

From the description of this isomorphism in §4.5.2, we see that the generator of C_2 corresponds to the loop in $\pi_1(\mathbb{RP}^n, \ell_0)$ that rotates the line ℓ_0 around an axis by an angle of π, coming back to itself but with the reversed orientation. $\square$

Remark 4.5.5. We have $\mathbb{RP}^1 \cong \mathbb{S}^1$, so $\pi_1(\mathbb{RP}^1, \ell_0) \cong \mathbb{Z}$. $\square$

Here is a variant on this:

Example 4.5.6 (Finite cyclic groups). Let $d \geq 2$ and let $\zeta \in \mathbb{C}$ be a primitive d^{th} root of unity. For $m \in C_d$, there is a well-defined complex number ζ^m. Regarding $\mathbb{S}^3$ as a subspace of $\mathbb{C}^2$, the group C_d acts on $\mathbb{S}^3$ via

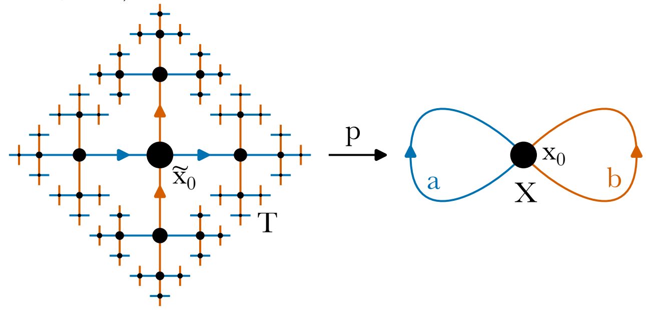
$$m(z_1, z_2) = (\zeta^m z_1, \zeta^m z_2) \quad \text{for } m \in C_d \text{ and } (z_1, z_2) \in \mathbb{S}^3 \subset \mathbb{C}^2.$$

The action is free, and since C_d is finite this is a covering space action (Exercise 1.8). Fixing a basepoint $x_0 \in \mathbb{S}^3/C_d$, we can therefore apply Theorem 4.5.1 and see that

$$\pi_1(\mathbb{S}^3/C_d, x_0) \cong C_d. \qquad \square$$

Remark 4.5.7. Let A be a finitely generated abelian group. We can find a space X and $x_0 \in X$ with $\pi_1(X, x_0) \cong A$ as follows. Write $A \cong \mathbb{Z}^n \times F$ with F a product of finite cyclic groups. For X, we can then take the product of a torus $\mathbb{T}^r$ with a product of quotients of $\mathbb{S}^3$ by these finite cyclic groups as in Example 4.5.6. $\square$

Example 4.5.8 (Free group). As in Example 1.6.9, consider the cover $p\colon T \to X$ shown here:



Each of the blue horizontal edges of T maps to the blue loop labeled a on the left side of X, and each of the orange vertical edges of T maps to the orange loop labeled b on the right side of X. Let $x_0 \in X$ be the vertex, and let $\widetilde{x}_0 \in T$ be the indicated lift of x_0. This is a regular cover (Exercise 1.11), and since T is a tree it is contractible (Lemma 4.4.12). Letting G be the deck group of $p\colon T \to X$, we can therefore apply Theorem 4.5.1 to see that

$$\pi_1(X, x_0) \cong G.$$

What kind of group is G? It acts simply transitively on the vertices of T, so each vertex is of the form $g\widetilde{x}_0$ for some unique $g \in G$. From the description of the isomorphism in §4.5.2, the element of $\pi_1(X, x_0)$ corresponding to $g \in G$ is the homotopy class of the loop in X based at x_0 obtained by taking a path in T from $\widetilde{x}_0$ to $g\widetilde{x}_0$ and projecting it to X.

Consider the elements $[a], [b] \in \pi_1(X, x_0)$. Let $\mathcal{W}$ be the set of *reduced* words in $[a]$ and $[b]$, that is, products of $[a]$ and $[a]^{-1} = [\overline{a}]$ and $[b]$ and $[b]^{-1} = [\overline{b}]$ such that:

- neither $[a][a]^{-1}$ nor $[a]^{-1}[a]$ appears as a subword; and
- neither $[b][b]^{-1}$ nor $[b]^{-1}[b]$ appears as a subword.

For instance, $\mathcal{W}$ contains $[b][a][a][b]^{-1}[a]^{-1}$. The empty word is allowed, so $1 \in \mathcal{W}$. Since each element of $\mathcal{W}$ is a product of elements of $\pi_1(X, x_0)$, there is a set map $\mathcal{W} \to \pi_1(X, x_0)$.

For each vertex $\widetilde{x}_1$ of T, there is a unique sequence of edges connecting $\widetilde{x}_0$ to $\widetilde{x}_1$ that does not backtrack, that is, traverse an edge in one direction and then go backwards along the same edge. This non-backtracking condition is exactly what is needed to ensure that this edge-path corresponds to an element of $\pi_1(X, x_0)$ represented by a reduced word. In this way, we see that the set map $\mathcal{W} \to \pi_1(X, x_0)$ is a bijection.

To sum up, $\pi_1(X, x_0)$ is generated by $[a]$ and $[b]$, and each element in it can be uniquely written as reduced word in $[a]$ and $[b]$. This implies that $\pi_1(X, x_0)$ is the free group on $[a]$ and $[b]$. In fact, for a set S some authors *define* the free group $F(S)$ on S to be the set of reduced words on elements of S. They construct a product on $F(S)$ by concatanating reduced words and then cancelling terms of the form ss^{-1} and $s^{-1}s$ to obtain a reduced word.

This is not a particularly good construction of a free group; for instance, it takes work to show that this product is associative since there are choices to be made as to what order to cancel terms to reduce a non-reduced word. More conceptual constructions first define free groups in a way that is manifestly a group and then prove that each element can be uniquely expressed as a reduced word. There are algebraic approaches to this (see Chapter YYY). However, in §5.4 we will show how to do this using the geometry of trees. □

4.6. Functoriality

We now study the ways in which the fundamental group and groupoid of X depends on X.

4.6.1. Initial thoughts. We start with the fundamental group. For a path-connected space X and $x_0 \in X$, Lemma 4.3.9 implies that up to isomorphism $\pi_1(X, x_0)$ does not depend on the choice of x_0. The isomorphism class of $\pi_1(X, x_0)$ therefore is an invariant of the X in the sense that if Y is another path-connected space with a basepoint y_0 and $\pi_1(X, x_0) \ncong \pi_1(Y, y_0)$, then $X \ncong Y$. For instance, since $\pi_1(\mathbb{S}^1, 1) \cong \mathbb{Z}$ but $\pi_1(\mathbb{S}^2, x_0) = 1$, it follows that $\mathbb{S}^1$ is not homeomorphic to $\mathbb{S}^2$.

REMARK 4.6.1. Of course, this can be proved in other ways as well; for instance, removing any two points from $\mathbb{S}^1$ gives a disconnected space, but removing a finite collection of points from $\mathbb{S}^2$ does not disconnect it. We will soon see results that can be proved with π_1 but seem resistant to more simple-minded ideas. □

4.6.2. Induced maps. Now consider a map $f\colon X \to Y$ and $x_0 \in X$. If γ is a loop in X based at x_0, then $f \circ \gamma$ is a loop in Y based at $f(x_0)$. The loop $[f \circ \gamma]$ only depends on the homotopy class of γ, and the map $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ defined by

$$f_*([\gamma]) = [f \circ \gamma] \quad \text{for all } [\gamma] \in \pi_1(X, x_0)$$

is a homomorphism called the homomorphism *induced* by f. We have already seen an example of this when we discussed retractions in §4.4.2: if A is a subspace of X and and $a_0 \in A$, then the inclusion map $\iota\colon A \hookrightarrow X$ induces a homomorphism $\iota_*\colon \pi_1(A, a_0) \to \pi_1(X, a_0)$.

4.6.3. Pointed spaces and homotopies. Since the fundamental group is the group of homotopy classes of loops, one expects the homomorphism induced by a map to only depend on the homotopy class of the map. However, this is not quite right since we have to be careful about the basepoint. To state things properly, we introduce the following terminology:

DEFINITION 4.6.2. A *pointed space* is a pair (X, x_0) with X a space and $x_0 \in X$. A map between pointed space (X, x_0) and (Y, y_0) is a map $f\colon X \to Y$ such that $f(x_0) = y_0$. We will denote such a map by $f\colon (X, x_0) \to (Y, y_0)$. A homotopy of maps from (X, x_0) to (Y, y_0) is a homotopy $f_t\colon X \to Y$ such that $f_t(x_0) = y_0$ for all $t \in I$. Just like for maps, we will denote this by $f_t\colon (X, x_0) \to (Y, y_0)$, and if such an f_t exists we will say that $f_0\colon (X, x_0) \to (Y, y_0)$ and $f_1\colon (X, x_0) \to (Y, y_0)$ are homotopic. □

With this setup, a map $f\colon (X, x_0) \to (Y, y_0)$ between pointed spaces induces a homomorphism $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, y_0)$, and if $f\colon (X, x_0) \to (Y, y_0)$ and $g\colon (X, x_0) \to (Y, Y_0)$ are homotopic maps between pointed spaces then $f_* = g_*$.

4.6.4. Functors. The homomorphisms induced by maps of pointed spaces have the following two simple properties:

- for maps of pointed space $f\colon (X, x_0) \to (Y, y_0)$ and $g\colon (Y, y_0) \to (Z, z_0)$, we have $(g \circ f)_* = g_* \circ f_*$; and
- the identity map $\mathbb{1}\colon (X, x_0) \to (X, x_0)$ induces the identity homomorphism, i.e., $\mathbb{1}_* = \mathbb{1}$.

All of this can be summarized in categorical language as follows. Recall that if $\mathbf{C}$ and $\mathbf{D}$ are categories, then a *functor* $F\colon \mathbf{C} \to \mathbf{D}$ consists of the following data:

- For all objects $C \in \mathbf{C}$, an object $F(D) \in \mathbf{D}$.
- For all morphisms $f\colon C_1 \to C_2$ between objects of $\mathbf{C}$, a morphism $F(f)\colon F(C_1) \to F(C_2)$.

These are required to satisfy:

- for all morphisms $f\colon C_1 \to C_2$ and $g\colon C_2 \to C_3$ between objects of $\mathbf{C}$, we have $F(g \circ f) = F(g) \circ F(f)$; and
- for all identity morphisms $\mathbb{1}_C\colon C \to C$ in $\mathbf{C}$, we have $F(\mathbb{1}_C) = \mathbb{1}_{F(C)}$.

To fit the fundamental group into this, let Top_* be the category of pointed spaces, so the objects of Top_* are pointed spaces (X, x_0) and the morphisms in Top_* are the map $f\colon (X, x_0) \to (Y, y_0)$ between pointed spaces. We can then summarize our discussion by:

LEMMA 4.6.3. *The fundamental group is a functor* $\pi_1\colon \text{Top}_* \to \text{Group}$.

REMARK 4.6.4. It will play less of a role in this book, but see §4.6.6 below for a discussion of the functoriality of the fundamental groupoid. □

4.6.5. Illustration of functoriality. Using functoriality, we can use the fundamental group to obstruct the existence of maps between spaces. As a classic example of this, we give another proof of the following, which we originally proved using winding numbers in §3.2:

PROPOSITION 4.6.5. *There does not exist a retraction* $r\colon \mathbb{D}^2 \to \mathbb{S}^1$.

Proof. Assume that a retraction $r\colon \mathbb{D}^2 \to \mathbb{S}^1$ exists. Let $\iota\colon \mathbb{S}^1 \to \mathbb{D}^2$ be the inclusion, so $r \circ \iota = \mathbb{1}_{\mathbb{S}^1}$. Letting $x_0 \in \mathbb{S}^1$ be a basepoint, we have $\iota(x_0) = r(x_0) = x_0$. The maps on fundamental groups induced by ι and r are thus

$$\begin{array}{ccccc} \pi_1(\mathbb{S}^1, x_0) & \xrightarrow{\iota_*} & \pi_1(\mathbb{D}^1, x_0) & \xrightarrow{r_*} & \pi_1(\mathbb{S}^1, x_0). \\ \| & & \| & & \| \\ \mathbb{Z} & & 0 & & \mathbb{Z} \end{array}$$

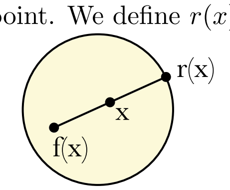
Since the central group is 0, this composition is the 0 map. However, since $r \circ \iota = \mathbb{1}_{\mathbb{S}^1}$ this composition is also the identity map $\mathbb{1}\colon \mathbb{Z} \to \mathbb{Z}$, giving a contradiction. □

Remark 4.6.6. Geometrically, this proof relies on the same key facts as our proof in §3.2; however, they are packaged in different ways. □

For the reader's convenience, we explain again why this implies the two-dimensional Brouwer fixed point theorem:

Theorem 4.6.7 (Two-dimensional Brouwer fixed point theorem). *Let $f\colon \mathbb{D}^2 \to \mathbb{D}^2$ be a continuous map. Then f has a fixed point, i.e., there exists some $x \in \mathbb{D}^2$ with $f(x) = x$.*

Proof. Assume that f has no fixed points. Define a function $r\colon \mathbb{D}^2 \to \mathbb{S}^1$ as follows. For $x \in \mathbb{D}^2$, consider the ray starting at $f(x)$ and passing through x. This is well-defined since $f(x) \neq x$, and it intersects the boundary $\mathbb{S}^1$ in a single point. We define $r(x)$ to be that intersection point:



For $x \in \mathbb{S}^1$, we have $r(x) = x$. In other words, r is a retraction from $\mathbb{D}^2$ to its boundary $\mathbb{S}^1$, contradicting Proposition 4.6.5. □

4.6.6. Fundamental groupoid and functoriality. For completeness, we now explain how to think about the fundamental groupoid as a functor. Recall that a groupoid is a category in which all morphisms are invertible. For groupoids $\mathbf{G}_1$ and $\mathbf{G}_2$, a groupoid homomorphism from $\mathbf{G}_1$ to $\mathbf{G}_2$ is a functor $F\colon \mathbf{G}_1 \to \mathbf{G}_2$. Unpacking this, F consists of the following data:

- For each object $p \in \mathbf{G}_1$, an object $F(p) \in \mathbf{G}_2$.
- For each morphism $\phi\colon p \to q$ in $\mathbf{G}_1$, a morphism $F(\phi)\colon F(p) \to F(q)$ in $\mathbf{G}_2$.

The morphisms $F(\phi)$ must respect composition in the obvious sense. For each $p \in \mathbf{G}_1$, we have the group $\mathrm{Aut}_{\mathbf{G}_1}(p)$, and $F\colon \mathbf{G}_1 \to \mathbf{G}_2$ induces a group homomorphism $F_*\colon \mathrm{Aut}_{\mathbf{G}_1}(p) \to \mathrm{Aut}_{\mathbf{G}_2}(f(p))$. If we think of a groupoid as a collection of groups connected by isomorphisms, the homomorphism $F\colon \mathbf{G}_1 \to \mathbf{G}_2$ can be regarded as a collection of group homomorphisms that respect the given isomorphisms.

Let Groupoid be the category whose objects are groupoids and whose objects are groupoid homomorphisms. The fundamental groupoid can then be regarded as a functor $\Pi\colon \mathrm{Top} \to \mathrm{Groupoid}$:

- For a space X, we have the groupoid $\Pi(X)$.
- For a map of space $f\colon X \to Y$, we have the groupoid homomorphism $f_*\colon \Pi(X) \to \Pi(Y)$ defined as follows:
  - An object of $\Pi(X)$ is a point $p \in X$, and $f_*(p) = f(p) \in Y$.
  - A morphism in $\Pi(X)$ from $p \in X$ to $q \in X$ is the homotopy class of a path γ from p to q, and $f_*([\gamma]) = [f \circ \gamma]$.

We remark that unlike for the fundamental group, the groupoid homomorphisms $f_*\colon \Pi(X) \to \Pi(Y)$ are not homotopy invariant, at least not in a naive sense. See Exercise 4.6 for one way to think about this.

4.7. Exercises

EXERCISE 4.1. Let X be a path-connected space and let $\gamma_0, \gamma_1 : I \to X$ be two paths in X. Prove that γ_0 is homotopic to γ_1 if we do not require the homotopy to fix the endpoints of the path. □

EXERCISE 4.2. Let $\mathbf{G}$ be a groupoid and $\phi\colon A \to B$ be a morphism in $\mathbf{G}$. Consider $\overline{\phi}, \overline{\phi}' : B \to A$. Then $\phi = \overline{\phi}'$ if any of the following conditions are satisfied:

- $\overline{\phi} \circ \phi = \overline{\phi}' \circ \phi = \mathbb{1}_A$; or
- $\phi \circ \overline{\phi} = \phi \circ \overline{\phi}' = \mathbb{1}_B$; or
- $\overline{\phi} \circ \phi = \mathbb{1}_A$ and $\phi \circ \overline{\phi}' = \mathbb{1}_B$. □

EXERCISE 4.3. Let X and Y be spaces with basepoints $x_0 \in X$ and $y_0 \in Y$. Prove that

$$\pi_1(X \times Y, (x_0, y_0)) \cong \pi_1(X, x_0) \times \pi_1(Y, y_0).$$ □

EXERCISE 4.4. TOWRITE: an example of a space that is contractible but does not deformation retract to any point. □

EXERCISE 4.5. Verify that the maps constructed in the proof of Lemma 4.4.12 are well-defined and continuous. □

EXERCISE 4.6. Let $f_t\colon X \to Y$ be a homotopy of maps between spaces. Prove that f_t induces a natural isomorphism between the functors $f_0\colon \Pi_1(X) \to \Pi_1(Y)$ and $f_1\colon \Pi_1(X) \to \Pi_1(Y)$ giving the induced maps between fundamental groupoids. Here recall that if $F, G\colon \mathbf{C} \to \mathbf{D}$ are functors between categories $\mathbf{C}$ and $\mathbf{D}$, then a *natural isomorphism* $\Psi\colon F \to G$ consists of the following data:

- For all objects A of $\mathbf{C}$, a $\mathbf{D}$-isomorphism $\Psi(A)\colon F(A) \to G(A)$.

These must satisfy the following:

- For all morphisms $\lambda\colon A \to B$ between objects of $\mathbf{C}$, the diagram

$$\begin{array}{ccc} F(A) & \xrightarrow{F(\lambda)} & F(B) \\ \downarrow{\scriptstyle \Psi(A)} & & \downarrow{\scriptstyle \Psi B} \\ G(A) & \xrightarrow{G(\lambda)} & G(B) \end{array}$$

  must commute. □

EXERCISE 4.7. Fix a basepoint $x_0 \in \mathbb{S}^1$ and let $f\colon (\mathbb{S}^1, x_0) \to (\mathbb{S}^1, x_0)$ be a pointed map. Recall that we defined $\deg(f) \in \mathbb{Z}$ in §2.6.4. Let $d = \deg(f)$. Prove that the induced map $f_*\colon \pi_1(\mathbb{S}^1, x_0) \to \pi_1(\mathbb{S}^1, x_0)$ from $\mathbb{Z}$ to $\mathbb{Z}$ is[4] multiplication by d. □

[4]Here we are choosing the same isomorphism $\pi_1(\mathbb{S}^1, x_0) \cong \mathbb{Z}$ for the domain and codomain. If we chose different isomorphisms with $\mathbb{Z}$ for the domain and codomain, it might be multiplication by $-d$.

CHAPTER 5

Fundamental groups: homotopy equivalences and more examples

This chapter continues our development of the fundamental group. The topics include homotopy equivalences and a variety of examples, including a proof that all groups appear as fundamental groups of spaces.

5.1. Homotopy equivalences

It is often useful to regard homeomorphic spaces as being the same. In this section, we discuss a weakening of this that plays an important role in algebraic topology.

5.1.1. Pointed homotopy equivalences. A map $f\colon (X, x_0) \to (Y, y_0)$ between pointed spaces is a *homotopy equivalence* if there exists a map $g\colon (Y, y_0) \to (X, x_0)$ such that $g \circ f\colon (X, x_0) \to (X, x_0)$ and $f \circ g\colon (Y, y_0) \to (Y, y_0)$ are both homotopic to the identity. We call g a *homotopy inverse* to f, and if there is a homotopy equivalence between (X, x_0) and (Y, y_0) then we will say that (X, x_0) is *homotopy equivalent* to (Y, y_0). Here is an example:

Example 5.1.1. Let X be a space, let $A \subset X$ be a subspace, and let $r_t\colon X \to X$ be a deformation retract to A. We can therefore regard r_1 as a retraction $r_1\colon X \to A$. Pick a basepoint $a_0 \in A$, and let $\iota\colon (A, a_0) \to (X, a_0)$ be the inclusion. Then ι is a homotopy equivalence with homotopy inverse $r_1\colon (X, a_0) \to (A, a_0)$. Indeed, $r_1 \circ \iota\colon (A, a_0) \to (A, a_0)$ is literally the identity, and r_t is a homotopy from the identity $r_0 = \mathbb{1}_X\colon (X, x_0) \to (X, x_0)$ to $r_1 = \iota \circ r_1$. □

It will become more and more clear as we delve deeper into algebraic topology that homotopy equivalent pointed spaces are in many ways the "same" from the perspective of the tools of the subject. Here is one easy way in which this is true, which generalizes the corresponding fact for deformation retracts (Lemma 4.4.9):

Lemma 5.1.2. *Let $f\colon (X, x_0) \to (Y, y_0)$ be a homotopy equivalence between pointed spaces. Then $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, y_0)$ is an isomorphism.*

Proof. Let $g\colon (Y, y_0) \to (X, x_0)$ be a homotopy inverse to f. Since $g \circ f\colon (X, x_0) \to (X, x_0)$ and $f \circ g\colon (Y, y_0) \to (Y, y_0)$ are homotopic to the identity, the induced maps $(g \circ f)_*\colon \pi_1(X, x_0) \to \pi_1(X, x_0)$ and $(f \circ g)_*\colon \pi_1(Y, y_0) \to \pi_1(Y, y_0)$ are the identity. Functorality implies that

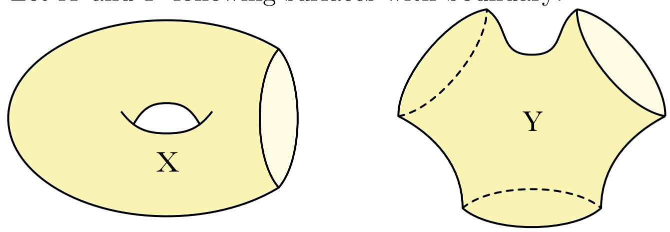
$$\mathbb{1} = (g \circ f)_* = g_* \circ f_* \quad \text{and}\, \mathbb{1} = (f \circ g)_* = f_* \circ g_*,$$

so f_* and g_* are inverses to each other. This implies that that f_* and g_* are isomorphisms. □

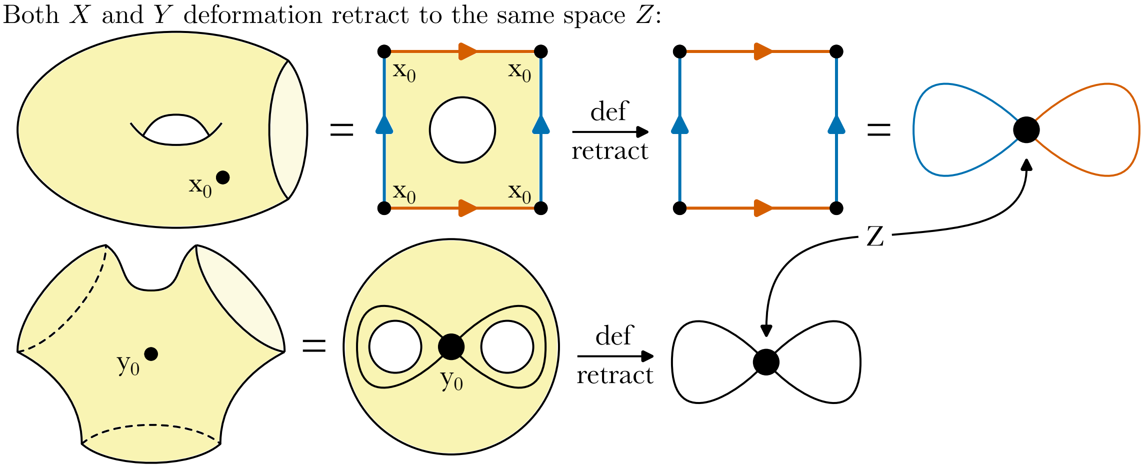
5.1.2. Composing deformation retractions. As we already noted, if X deformation retracts to a subspace Y and $y_0 \in Y$, then the inclusion $\iota\colon (Y, y_0) \to (X, y_0)$ is a homotopy equivalence. Being homotopy equivalent is an equivalence relation (see Exercise 5.1), so by applying this multiple times we can get interesting homotopy equivalences. For instance, if Y is a subspace of both X and Z and both X and Z deformation retract to Y, then (X, y_0) is homotopy equivalent to (Z, y_0) even though neither X or Z is contained in the other.

Here is an example of a non-obvious homotopy equivalence proved using this approach:

EXAMPLE 5.1.3. Let X and Y following surfaces with boundary:



Both X and Y deformation retract to the same space Z:



Letting $x_0 \in X$ and $y_0 \in Y$ be as indicated, it follows that (X, x_0) is homotopy equivalent to (Y, y_0). Moreover, since the fundamental group of Z is a free group on two generators (Example 4.5.8), it follows that $\pi_1(X, x_0)$ and $\pi_1(Y, y_0)$ are also free groups on two generators. □

5.1.3. Unpointed homotopy equivalences. We now explain how this works without basepoints. A map $f\colon X \to Y$ between spaces is a *homotopy equivalence* if there exists a map $g\colon Y \to X$ such that $g \circ f\colon X \to X$ and $f \circ g\colon Y \to Y$ are homotopic to the identity. The difference between this and the pointed case is that these homotopies need not fix a basepoint. We call g a *homotopy inverse* to f, and if a homotopy equivalence from X to Y exists we say that X and Y are *homotopy equivalent*.

In Lemma 4.4.11, we proved that even though contractions need not fix a basepoint, it is still true that contractible spaces have trivial fundamental groups. By being similarly careful with the basepoint, we prove the following:

LEMMA 5.1.4. *Let $f\colon X \to Y$ be a homotopy equivalence and let $x_0 \in X$. Set $y_0 = f(x_0)$. Then $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, y_0)$ is an isomorphism.*

PROOF. Let $g\colon Y \to X$ be a homotopy inverse to f. Set $x_1 = g(y_0)$ and $y_1 = f(x_1)$. The naive thing to do would be to prove that the maps $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, y_0)$ and $g_*\colon \pi_1(Y, y_0) \to \pi_1(X, x_1)$ were inverses to each other. However, this does not make sense since the domain $\pi_1(X, x_0)$ of f_* is not the same as the codomain $\pi_1(X, x_1)$ of g_*.

Instead, what we will prove is that

$$(g \circ f)_*\colon \pi_1(X, x_0) \to \pi_1(X, x_1) \quad \text{and} \quad (f \circ g)_*\colon \pi_1(Y, y_0) \to \pi_1(Y, y_1) \tag{5.1.1}$$

are both isomorphisms. This will imply that $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, y_0)$ and $g_*\colon \pi_1(Y, y_0) \to \pi_1(X, x_1)$ are both injections. Since f_* and g_* are injections and $g_* \circ f_* = (g \circ f)_*$ is an isomorphism, it is immediate that f_* is also a surjection,[1] so f_* is an isomorphism.

It remains to prove that the two maps in (5.1.1) are isomorphisms. The proofs are the same, so we will give the details for $(g \circ f)_*\colon \pi_1(X, x_0) \to \pi_1(X, x_1)$. Since g is a homotopy inverse to f, the

[1]Here are more details. Consider some $\zeta \in \pi_1(Y, y_0)$. We want to find some $\eta \in \pi_1(X, x_0)$ with $f_*(\eta) = \zeta$. Since $g_* \circ f_*$ is an isomorphism, there exists some $\eta \in \pi_1(X, x_0)$ such that $g_*(f_*(\eta)) = g_*(\zeta)$. Since g_* is an injection, we must have $f_*(\eta) = \zeta$.

map $g \circ f\colon X \to X$ is homotopic to the identity. Let $h_t\colon X \to X$ be a homotopy from $g \circ f$ to $\mathbb{1}_X$. Let $\delta\colon I \to X$ be the path

$$\delta(s) = h_s(x_0) \quad \text{for } s \in I.$$

We have

$$\delta(0) = h_0(x_0) = g(f(x_0)) = x_1 \quad \text{and} \quad \delta(1) = h_1(x_0) = \mathbb{1}_X(x_0) = x_0,$$

so δ is a path from x_1 to x_0. As in the proof of Lemma 4.3.9, the path δ induces an isomorphism $\delta_*\colon \pi_1(X, x_0) \to \pi_1(X, x_1)$ defined by

$$\delta_*([\gamma]) = [\delta \cdot \gamma \cdot \overline{\delta}] \quad \text{for all } [\gamma] \in \pi_1(X, x_0).$$

We will prove that $(g \circ f)_*$ equals δ_*.

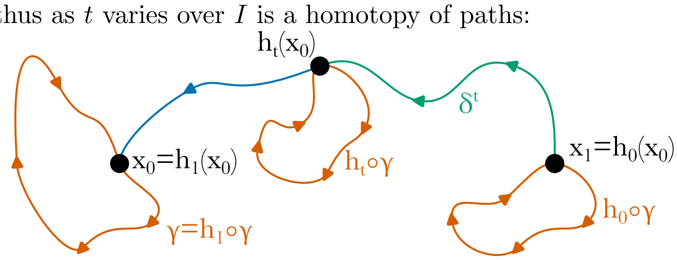
To see this, consider $[\gamma] \in \pi_1(X, x_0)$. Since $h_0 = g \circ f$, we must prove that

$$[h_0 \circ \gamma] = [\delta \cdot \gamma \cdot \overline{\delta}] \quad \text{in } \pi_1(X, x_1).$$

The maps $h_t \circ \gamma\colon I \to X$ do not form a homotopy of paths since the basepoint x_0 moves. To fix this, define $\delta^t\colon I \to X$ via the formula

$$\delta^t(s) = h_{ts}(x_0) \quad \text{for all } s, t \in I.$$

The path δ^t thus goes from $h_0(x_0) = g \circ f(x_0) = x_1$ to $h_t(x_0)$. It follows that $\delta^t \cdot (h_t \circ \gamma) \cdot \overline{\delta}^t$ is a path from x_1 to x_1, and thus as t varies over I is a homotopy of paths:



Recalling that $\mathfrak{c}_{x_1}$ is the constant path at x_1, we deduce that

$$[\delta^0 \cdot (h_0 \circ \gamma) \cdot \overline{\delta}^0] = [\mathfrak{c}_{x_1} \cdot (h_0 \circ \gamma) \cdot \mathfrak{c}_{x_1}] = [h_0 \circ \gamma]$$

equals

$$[\delta^1 \cdot (h_1 \circ \gamma) \cdot \overline{\delta}^1] = [\delta \cdot \gamma \cdot \overline{\delta}],$$

as desired. □

5.2. Collapsing contractible subspaces

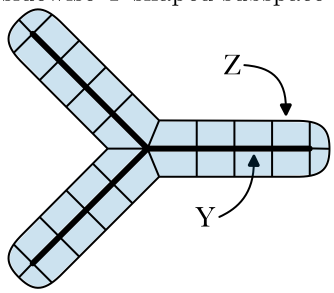
Let X be a space and Y be a contractible subspace of X. Consider the quotient map $q\colon X \to X/Y$. It turns out that in many cases q is a homotopy equivalence. Roughly speaking, this holds as long as Y is embedded into X with reasonable local properties. There are a number of conditions that ensure this. In this section, we give one that is fairly easy to state and prove.

5.2.1. Mapping cylinder neighborhoods. Recall from Example 4.4.6 that for a map $f\colon Z \to Y$ between spaces, the mapping cylinder of f is the space $\text{Cyl}(f)$ obtained by quotienting the disjoint union $(Z \times I) \sqcup Y$ to identify $(z, 1) \in Z \times I$ with $f(z) \in Y$ for all $z \in Z$. For $z \in Z$ and $s \in I$, let $\overline{(z, s)}$ be the image of $(z, s) \in Z \times I$ in $\text{Cyl}(f)$. We now define:

Definition 5.2.1. Let X be a space and let $Y \subset X$ be a subspace. A *mapping cylinder neighborhood* of Y is a closed subset N of X containing Y along with a closed subset $Z \subset N$ such that:

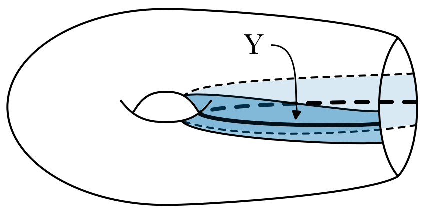
- $N \setminus Z$ is an open neighborhood of Y in X; and
- there exists a map $f\colon Z \to Y$ and a homeomorphism $\phi\colon \text{Cyl}(f) \to N$ such that $f(\overline{(z, 0)}) = z$ and $f(y) = y$ for all $z \in Z$ and $y \in Y$. □

Example 5.2.2. Let Y be the sidewise-Y shaped subspace of $\mathbb{R}^2$ shown here:

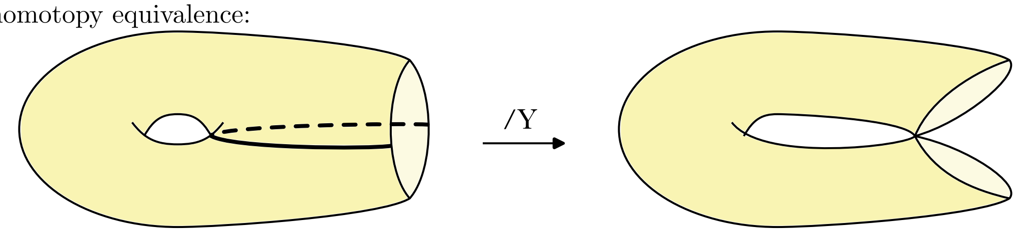


The subspace Y of $\mathbb{R}^2$ has a mapping cylinder neighborhood N indicated in blue. This blue subspace is the mapping cylinder of a map $f\colon Z \to Y$ with $Z \cong \mathbb{S}^1$ the indicated subspace. The lines connect points $z \in Z$ with their images $f(z) \in Y$. Since Y is contractible and has a mapping cylinder neighborhood, it will follow from Theorem 5.2.5 below that the quotient map $\mathbb{R}^2 \to \mathbb{R}^2/Y$ is a homotopy equivalence. In fact, $\mathbb{R}^2/Y \cong \mathbb{R}^2$ (see Exercise 5.4). □

Example 5.2.3. Let X be the following surface with boundary and let $Y \cong I$ be the indicated arc in X:



A mapping cylinder neighborhood N of Y is drawn in blue. Here $N \cong Y \times I$, and N is homeomorphic to the mapping cylinder of the projection $f\colon Y \sqcup Y \to Y$. Since Y is contractible and has a mapping cylinder neighborhood, it will follow from Theorem 5.2.5 below that the quotient map $q\colon X \to X/Y$ is a homotopy equivalence:



In this case, X and X/Y are not homeomorphic; indeed, X/Y is not a manifold around the point that is the image of Y. We showed in Example 5.1.3 that the fundamental group of X is the free group on two generators, so the same is true for X/Y. □

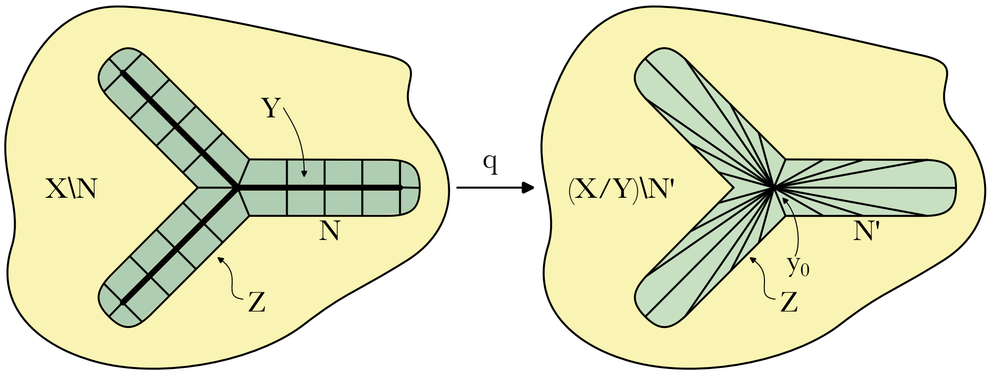
Example 5.2.4. For readers who are familiar with smooth manifolds, here is an important example. Let M^n be a smooth manifold with boundary and let N^d be a properly embedded submanifold of M^n. A closed tubular neighborhood of N^d is then a mapping cylinder neighborhood of N^d. Example 5.2.3 is a special case of this. □

5.2.2. Collapsing subspaces with mapping cylinder neighborhoods.

We now prove:

Theorem 5.2.5. *Let X be a space and let $Y \subset X$ be a contractible subspace with a mapping cylinder neighborhood. Then the quotient map $q\colon X \to X/Y$ is a homotopy equivalence.*

Proof. We must use the hypotheses to construct a homotopy inverse $g\colon X/Y \to X$ to q. Let N be a mapping cylinder neighborhood of Y. Identify N with $\mathrm{Cyl}(f)$ for some $Z \subset N$ and some map $f\colon Z \to Y$. Let y_0 be the point of X/Y corresponding to Y, let $f'\colon Z \to y_0$ be the projection, and let $N' = \mathrm{Cyl}(f')$. We can identify N' with N/Y, and after making this identification N' is a mapping cylinder neighborhood of y_0 in X/Y with $X \setminus N = (X/Y) \setminus N'$. See here:



To construct a continuous map $X/Y \to X$, it is enough to construct a continuous map $N' \to N$ that is the identity on Z and then extend $N' \to N$ to X/Y by the identity.[2] In a similar way, we can construct continuous maps $X \to X$ (resp. $X/Y \to X/Y$) by constructing continuous maps $N \to N$ (resp. $N' \to N$) and extending by the identity.

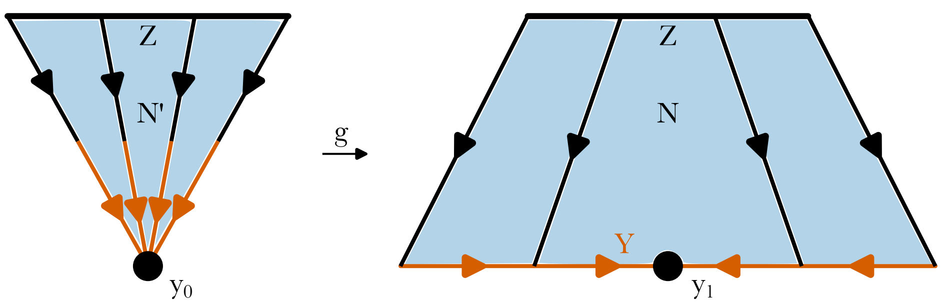
For $z \in Z$ and $s \in I$, let $\overline{(z,s)} \in N$ and $\overline{(z,s)}' \in N'$ be the corresponding points. We now divide the proof into three steps:

Step 1. *We construct the purported homotopy inverse $g\colon X/Y \to X$.*

Since Y is contractible, there is a homotopy $h_t\colon Y \to Y$ from $\mathbb{1}_Y$ to a constant map. Let $y_1 \in Y$ be the constant value of h_1. Define $g\colon X/Y \to X$ via the formulas

$$\begin{cases} g(\overline{(z,s)}') = \overline{(z,2s)} & \text{for } z \in Z \text{ and } s \in [0,1/2], \\ g(\overline{(z,s)}') = h_{2s-1}(f(z)) & \text{for } z \in Z \text{ and } s \in [1/2,1], \\ g(y_0) = y_1 & \\ g(x) = x & \text{for } x \in (X/Y) \setminus N' = X \setminus N. \end{cases}$$

See here, where the indicated map takes each line segment from $z \in Z$ to y_0 to the path that first goes to $f(z)$ (in black) and then in Y to y_1 (in orange):



By what we said above, this map $g\colon X/Y \to X$ is continuous.

Step 2. *We prove that the composition $g \circ q\colon X \to X$ is homotopic to the identity.*

The map $g \circ q\colon X \to X$ is given by the formulas

$$\begin{cases} g \circ q(\overline{(z,s)}) = \overline{(z,2s)} & \text{for } z \in Z \text{ and } s \in [0,1/2], \\ g \circ q(\overline{(z,s)}) = h_{2s-1}(f(z)) & \text{for } z \in Z \text{ and } s \in [1/2,1], \\ g \circ q(y) = y_1 & \text{for } y \in Y, \\ g \circ q(x) = x & \text{for } x \in X \setminus N. \end{cases}$$

[2] To see that this is continuous, note that the map $X/Y \to X$ is continuous on the closed sets N' and $(X/Y)\setminus(N'\setminus Z)$; indeed, on the latter set it is the identity. These cover X/Y. Now apply the fact that if $\psi\colon A \to B$ is a map of sets between spaces and $\{C_1, \ldots, C_n\}$ is a cover of A by closed sets such that each $\psi|_{C_i}$ is continuous, then ψ is continuous. Note that this would be false if our cover had infinitely many closed sets in it.

This is homotopic to the identity via the homotopy $\phi_t \colon X \to X$ given by the formulas

$$\begin{cases} \phi_t(\overline{(z,s)}) = \overline{(z,(2-t)s)} & \text{for } z \in Z \text{ and } s \in [0, 1/(2-t)], \\ \phi_t(\overline{(z,s)}) = h_{(2-t)s-1}(f(z)) & \text{for } z \in Z \text{ and } s \in [1/(2-t), 1], \\ \phi_t(y) = h_{1-t}(y) & \text{for } y \in Y, \\ \phi_t(x) = x & \text{for } x \in X \setminus N. \end{cases}$$

By the discussion at the beginning of the proof this is continuous.

Step 3. *We prove that the composition $q \circ g \colon X/Y \to X/Y$ is homotopic to the identity.*

The map $q \circ g \colon X/Y \to X/Y$ is given by the formulas

$$\begin{cases} q \circ g(\overline{(z,s)}') = \overline{(z,2s)}' & \text{for } z \in Z \text{ and } s \in [0,1/2], \\ q \circ g(\overline{(z,s)}') = y_0 & \text{for } z \in Z \text{ and } s \in [1/2, 1], \\ q \circ g(y_0) = y_0 & \\ g \circ q(x) = x & \text{for } x \in (X/Y) \setminus N'. \end{cases}$$

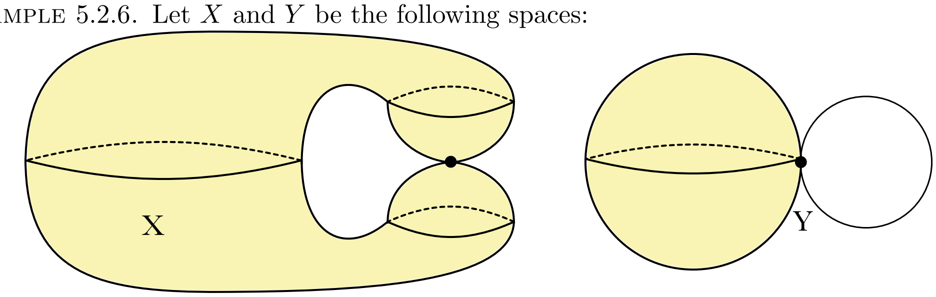
This is homotopic to the identity via the homotopy $\psi_t \colon X/Y \to X/Y$ given by the formulas

$$\begin{cases} \psi_t(\overline{(z,s)}') = \overline{(z,(2-t)s)}' & \text{for } z \in Z \text{ and } s \in [0, 1/(2-t)], \\ \psi_t(\overline{(z,s)}') = y_0 & \text{for } z \in Z \text{ and } s \in [1/(2-t), 1], \\ \psi_t(y_0) = y_0 & \\ \psi_t(x) = x & \text{for } x \in X \setminus N. \end{cases}$$

By the discussion at the beginning of the proof this is continuous. □

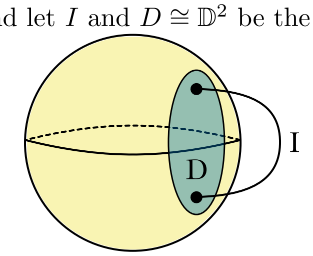
5.2.3. An example. We now give an example of how to this to analyze an interesting example.

Example 5.2.6. Let X and Y be the following spaces:



The space X is obtained by quotienting $\mathbb{S}^2$ to identify two points together,[3] and the space Y is obtained by gluing $\mathbb{S}^2$ and $\mathbb{S}^1$ together at a single point. We will prove that X and Y are homotopy equivalent, and then we will prove that their fundamental groups are isomorphic to $\mathbb{Z}$.

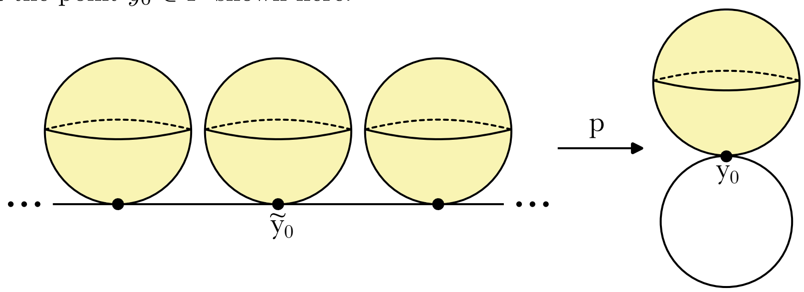
Let Z be the following space and let I and $D \cong \mathbb{D}^2$ be the indicated subspaces of Z:



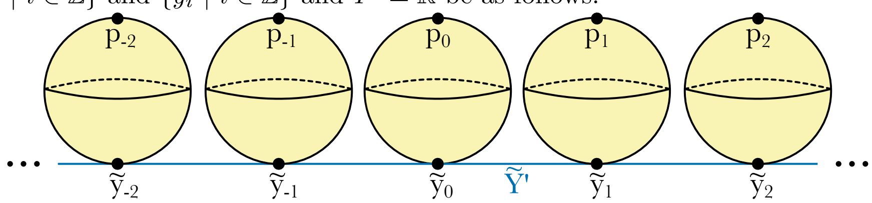
We have $Z/I \cong X$ and $Z/D \cong Y$. Since I is contractible and has a mapping cylinder neighborhood in Z, it follows that Z is homotopy equivalent to X. Similarly, since D is contractible and has a mapping cylinder neighborhood in Z, it follows that Z is homotopy equivalent to Y. We conclude that X and Y are homotopy equivalent, as claimed.

[3]It does not matter which two points are identified. Indeed, any two points "look the same" in the sense that they differ by a homeomorphism of $\mathbb{S}^2$. More generally, if M^n is a connected n-manifold with $n \geq 2$ and $\{p_1, \ldots, p_k\}$ and $\{q_1, \ldots, q_k\}$ are two sets of k distinct points on M^n, then there exists a homeomorphism $f \colon M^n \to M^n$ with $f(p_i) = q_i$ for all $1 \leq i \leq k$. See Exercise 5.5.

In particular, the fundamental groups of X and Y are the same. It remains to prove that the fundamental group of Y is isomorphic to $\mathbb{Z}$. Consider the basepoint $y_0 \in Y$, the covering space $p\colon \widetilde{Y} \to Y$, and the point $\widetilde{y}_0 \in \widetilde{Y}$ shown here:



The map p wraps each segment of the horizontal line connecting two spheres around the copy of $\mathbb{S}^1$ in Y. This is a regular cover with deck group $\mathbb{Z}$, which acts on $\widetilde{Y}$ by horizontal translations. By Theorem 4.5.1, to prove that $\pi_1(Y, y_0) \cong \mathbb{Z}$ it is enough to prove that $\widetilde{Y}$ is simply-connected. Since $\widetilde{Y}$ is path-connected, this is equivalent to showing that $\pi_1(\widetilde{Y}, \widetilde{y}_0) = 1$. Consider some $[\gamma] \in \pi_1(\widetilde{Y}, \widetilde{y}_0)$. Let $P = \{p_i \mid i \in \mathbb{Z}\}$ and $\{\widetilde{y}_i \mid i \in \mathbb{Z}\}$ and $\widetilde{Y}' \cong \mathbb{R}$ be as follows:



Just like when we proved that $\mathbb{S}^n$ is simply-connected for $n \geq 2$, we can apply Lemma 4.1.5 (general position) to homotope γ such that it lies in $\widetilde{Y} \setminus P$. To prove that $[\gamma] = 1$, it is therefore enough to prove that $\pi_1(\widetilde{Y} \setminus P, \widetilde{y}_0) = 1$. If $\widetilde{Y}_i$ is the sphere in $\widetilde{Y}$ containing $\{p_i, \widetilde{y}_i\}$, then $\widetilde{Y}_i \setminus p_i$ deformation retracts to $\widetilde{y}_i$. Using these deformation retractions, $\widetilde{Y} \setminus P$ deformation retracts to $\widetilde{Y}'$, which deformation retracts to $\widetilde{y}_0$. This implies that $\pi_1(\widetilde{Y} \setminus P, \widetilde{y}_0) = 1$, as desired. □

5.3. Every group is the fundamental group of a space

In this section, we prove that every group G is the fundamental group of some space.

5.3.1. Joins. This requires some topological preliminaries. Let X and Y be spaces. Recall that $X \sqcup Y$ is the disjoint union of X and Y. The *join* of X and Y, denoted $X * Y$, is the quotient space

$$X * Y = X \sqcup Y \sqcup (X \times Y \times I) \,/ \sim,$$

where $\sim$ makes the following identifications:

- For $x \in X$ and $y \in Y$, we identify $(x, y, 0) \in X \times Y \times I$ with $x \in X$.
- For $x \in X$ and $y \in Y$, we identify $(x, y, 1) \in X \times Y \times I$ with $y \in Y$.

The space $X * Y$ contains subspaces X and Y, and the other points of $X * Y$ are all of the form (x, y, s) with $x \in X$ and $y \in Y$ and $s \in (0, 1)$. One should view $X * Y$ as consisting of X and Y as well as line segments connecting points of X to points of Y. For $x \in X$ and $y \in Y$ and $s \in (0, 1)$, the point (x, y, s) lies on the line segment connecting x to y. We introduce the following notation:

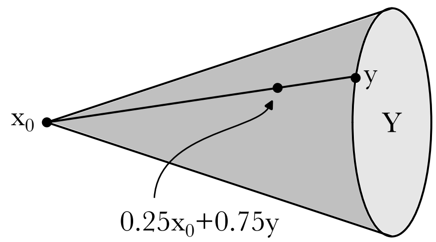
NOTATION 5.3.1. For $x \in X$ and $y \in Y$ and $s_1, s_2 \in I$ with $s_1 + s_2 = 1$, define $s_1 x + s_2 y$ to be the image of $(x, y, s_2) \in X \times Y \times I$ in $X * Y$. □

In this notation, we can move the point $s_1 x + s_2 y$ around by either:

- moving x around in X and y around in Y; or
- varying the coefficients s_1 and s_2. If s_1 goes to 0, then necessarily s_2 goes to 1. In this case, $s_1 x$ disappears and we are left with $y \in Y$. Similarly, if s_2 goes to 0 then necessarily s_1 goes to 1. In this case, $s_2 y$ disappears and we are left with $x \in X$.

Here are some examples:

Example 5.3.2. Let $X = \{x_0\}$ be a one-point space. Then $X * Y$ is the cone on Y with cone point x_0:



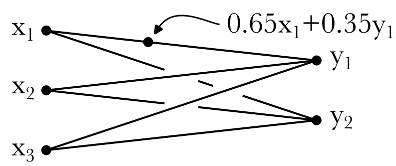
We denote this by $\text{Cone}_{x_0}(Y)$. This space deformation retracts to the cone point x_0 via the deformation retract $f_t\colon \text{Cone}_{x_0}(Y) \to \text{Cone}_{x_0}(Y)$ defined by

$$f_t(s_1 x_0 + s_2 y) = (s_1 + t s_2)x_0 + (s_2 - t s_2)y \quad \text{for } t \in I \text{ and } y \in Y \text{ and } s_1, s_2 \in I \text{ with } s_1 + s_2 = 1.$$

In particular, for $y \in Y$ and $s_1, s_2 \in I$ with $s_1 + s_2 = 1$ we have

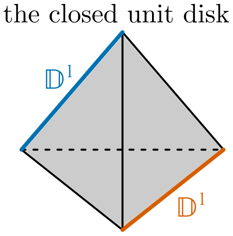
$$f_1(s_1 x_0 + s_2 y) = (s_1 + s_2)x_0 + 0y = x_0.$$ □

Example 5.3.3. Let X and Y be discrete sets. Then $X * Y$ is the complete bipartite graph with vertices X and Y:



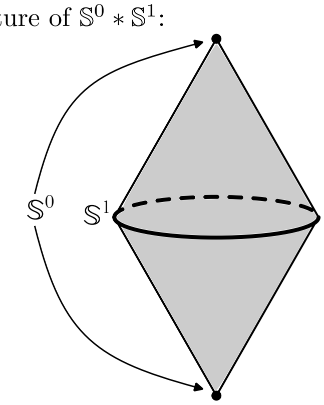
For later use, note that this is connected as long as X and Y are both nonempty. □

Example 5.3.4. Recall that $\mathbb{D}^n$ is the closed unit disk in $\mathbb{R}^n$. Here is a picture of $\mathbb{D}^1 * \mathbb{D}^1$:



As is clear from this figure, we have $\mathbb{D}^1 * \mathbb{D}^1 \cong \mathbb{D}^3$. More generally, we have $\mathbb{D}^n * \mathbb{D}^m \cong \mathbb{D}^{n+m+1}$; see Exercise 5.2. □

Example 5.3.5. Here is a picture of $\mathbb{S}^0 * \mathbb{S}^1$:

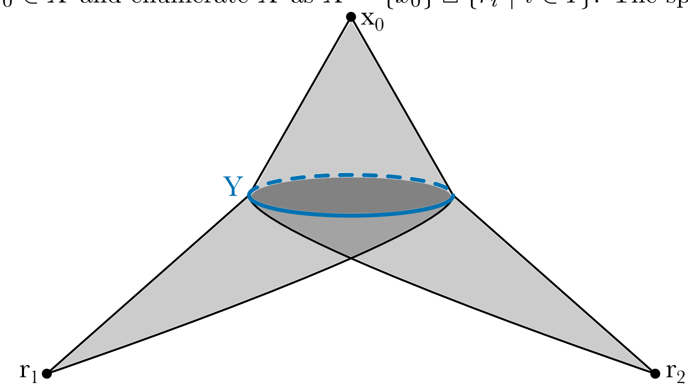


As is clear from this figure, we have $\mathbb{S}^0 * \mathbb{S}^1 \cong \mathbb{S}^2$. More generally, we have $\mathbb{S}^n * \mathbb{S}^m \cong \mathbb{S}^{n+m+1}$; see Exercise 5.3. □

5.3.2. Simple connectivity of join. We now prove:

Lemma 5.3.6. *Let X be a nonempty discrete space and let Y be a path-connected space. Then $X * Y$ is simply-connected.*

Proof. Pick $x_0 \in X$ and enumerate X as $X = \{x_0\} \sqcup \{r_i \mid i \in I\}$. The space $X * Y$ looks like:



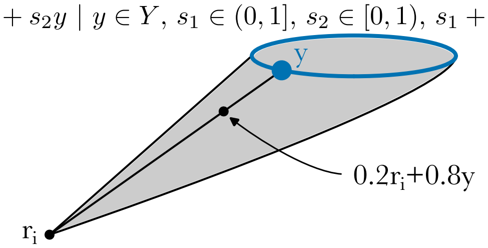
It is clear that $X * Y$ is path-connected, so it is enough to prove that $\pi_1(X * Y, x_0) = 1$. Consider some $[\gamma] \in \pi_1(X * Y, x_0)$. We will prove that γ can be homotoped to a constant map by a general position argument similar to the one we used in Lemma 4.1.4 to prove that $\mathbb{S}^n$ is simply-connected for $n \geq 2$. Our argument will have two steps:

(i) The loop $\gamma\colon I \to X * Y$ can be homotoped to lie in $U = X * Y \setminus \{r_i \mid i \in I\}$.
(ii) The subspace U of $X * Y$ deformation retracts to x_0.

These two facts will imply the lemma; indeed, (i) implies that $[\gamma]$ is in the image of the map $\pi_1(U, x_0) \to \pi_1(X * Y, x_0)$, and (ii) implies that $\pi_1(U, x_0) = 1$ and hence that $[\gamma] = 1$. Here are the proofs of (i) and (ii):

Step 1. *For $[\gamma] \in \pi_1(X * Y, x_0)$, the loop $\gamma\colon I \to X * Y$ can be homotoped to lie in $U = X * Y \setminus \{r_i \mid i \in I\}$.*

For $i \in I$, let $V_i = \{s_1 r_i + s_2 y \mid y \in Y,\ s_1 \in (0,1],\ s_2 \in [0,1),\ s_1 + s_2 = 1\}$:



Note that we do not allow $s_2 = 1$, so V_i does not contain Y. The set V_i is an open neighborhood of r_i that deformation retracts to r_i via the deformation retraction $r_t\colon V_i \to V_i$ defined by

$$r_t(s_1 r_i + s_2 y) = (s_1 + t s_2) r_i + (s_2 - t s_2) y \quad \text{for } t \in I \text{ and } s_1 r_i + s_2 y \in V_i.$$

It follows that V_i is contractible and hence simply-connected. Also, since Y is path-connected the space $V_i \setminus \{r_i\} \cong Y \times (0,1)$ is path-connected. These are exactly the conditions we need to apply Lemma 4.1.5 (general position), which shows that we can homotope γ so that its image does not contain any of the r_i and thus lies in U.

Step 2. *The subspace U of $X * Y$ deformation retracts to x_0.*

Define

$$W = \{s_1 x_0 + s_2 y \mid y \in Y,\ s_1, s_2 \in I,\ s_1 + s_2 = 1\},$$
$$\overline{V}_i = \{s_1 r_i + s_2 y \mid y \in Y,\ s_1, s_2 \in I,\ s_1 + s_2 = 1\}.$$

Both W and each $\overline{V}_i$ contain Y, and $X * Y$ is the union of W and the $\overline{V}_i$. Additionally, the subspace U is the union of W and the $\overline{V}_i \setminus \{r_i\}$. We have $\overline{V}_i \setminus \{r_i\} \cong Y \times (0,1]$, so $\overline{V}_i$ deformation retracts to Y. Combining all these deformation retractions gives a deformation retraction of U to W. Since $W \cong \mathrm{Cone}_{x_0}(Y)$, the space W deformation retracts to x_0 (see Example 5.3.2). The step follows. ☐

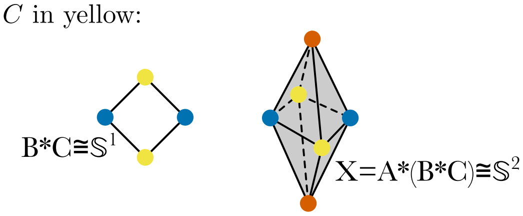
5.3.3. Realizing fundamental groups. All the tools needed to prove that every group is the fundamental group of some space are in place:

THEOREM 5.3.7. *Let G be a group. Then there exists a pointed space (X, x_0) with $\pi_1(X, x_0) \cong G$.*

PROOF. By Theorem 4.5.1, it is enough to construct a simply-connected space Z equipped with a covering space action (see §1.7.1) by G. The desired space X will then be Z/G with an arbitrary basepoint x_0. Since G has no structure, the only easily-identified set on which it acts is itself. We therefore must somehow build Z out of G.

Let A and B and C be three copies of G, regarded as sets with a left G-action. Endow A and B and C with the discrete topology, so each is a discrete space on which G acts freely. The join $B * C$ is the bipartite graph with vertices B and C (see Example 5.3.3), so $B * C$ is path-connected. Lemma 5.3.6 therefore implies that $X = A * (B * C)$ is simply-connected. The free actions of G on A and B and C induce a free action of G on X. You will verify in Exercise 5.6 that this action is a covering space action. □

REMARK 5.3.8. The above proof might seem a little abstract, so we will spell it out for G the cyclic group C_2 of order 2. Here are pictures in this case of $B * C$ and $X = A * (B * C)$, with A in orange and B in blue and C in yellow:



The action of C_2 on X interchanges the vertices of the same color. In fact, identifying X with $\mathbb{S}^2$, this is exactly the antipodal action of C_2 on $\mathbb{S}^2$ with quotient $\mathbb{RP}^2$, which we recall has fundamental group C_2. □

5.4. Free groups

We now explain how to construct free groups as fundamental groups of graphs. Our results give just a taste of what is possible. See Chapter YYY for far more, including algebraic proofs of the results we prove geometrically.

5.4.1. Definition and uniqueness of free groups. Let S be a set. Roughly speaking, a free group on S is a group that is easy to map out of. One need only say where the elements of S must go. Here is the formal definition:[4]

DEFINITION 5.4.1. Let S be a set. A *free group* on S is a group $F(S)$ containing S such that for all groups G, all set maps $h \colon S \to G$ extend uniquely to homomorphisms $H \colon F(S) \to G$. The set S is called a *free basis* for $F(S)$. □

It is not obvious that a free group on a set S exists. We will soon construct one, but first we prove that they are unique in the sense that any two free groups on S are isomorphic:[5]

LEMMA 5.4.2. *Let S be a set and let $F(S)$ and $F'(S)$ be free groups on S. Then there is an isomorphism $\phi \colon F(S) \to F'(S)$ with $\phi(s) = s$ for all $s \in S$.*

PROOF. By the definition of a free group, the inclusion $S \hookrightarrow F'(S)$ extends to a homomorphism $H \colon F(S) \to F'(S)$. Similarly, the inclusion $S \hookrightarrow F(S)$ extends to a homomorphism $H' \colon F'(S) \to F(S)$. Both $H' \circ H \colon F(S) \to F(S)$ and the identity $\mathbb{1}_{F(S)} \colon F(S) \to F(S)$ extend the inclusion $S \hookrightarrow F(S)$, so by the uniqueness in the definition of a free group we must have $H' \circ H = \mathbb{1}_{F(S)}$. Similarly, $H \circ H' = \mathbb{1}_{F'(S)}$, so H and H' are isomorphisms. □

[4]It would be better to not require that S is a subset of $F(S)$, but merely that there is a set map $\iota \colon S \to F(S)$. The definition then is that for all set maps $h \colon S \to G$ there is a homomorphism $H \colon F(S) \to G$ with $h = H \circ \iota$. One then proves that this implies that ι is injective, so S can be regarded as a subset of $F(S)$. We chose to make $S \subset F(S)$ part of the definition to make it easier to grasp for beginners.

[5]The definition of a free group is what is called a *universal mapping property*, and our proof of uniqueness works for anything defined by a universal mapping property.

5.4.2. Existence and reduced words. Let S be a set. A *word* in S is a formal expression $w = s_1^{\epsilon_1} \cdots s_n^{\epsilon_n}$ with $s_i \in S$ and $\epsilon_i \in \{\pm 1\}$ for all $1 \le i \le n$. This word is *reduced* if for all $1 \le i < n$ we do *not* have

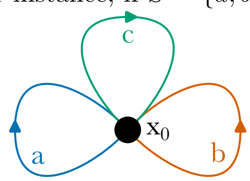
$$s_i^{\epsilon_i} s_{i+1}^{\epsilon_{i+1}} \in \left\{ss^{-1},\, s^{-1}s \mid s \in S\right\}.$$

By cancelling terms of the form ss^{-1} and $s^{-1}s$ with $s \in S$ we can reduce any word to a reduced word. If S is a subset of a group Γ, then we can regard words in S as elements of Γ and cancelling terms of the form ss^{-1} and ss^{-1} does not change the associated element of Γ. We will prove:

THEOREM 5.4.3. *Let S be a set. Then:*

- *there exists a free group $F(S)$ on S; and*
- *every element of $F(S)$ can be represented by a unique reduced word in S.*

PROOF. If $S = \emptyset$ then we can take $F(S) = 1$, so assume that $S \neq \emptyset$. Recall our convention from §1.6 that all graphs are oriented. Let X_S be a graph with one vertex x_0 and with $|S|$ oriented edges, each labeled with an element of S. For instance, if $S = \{a, b, c\}$ then X_S is



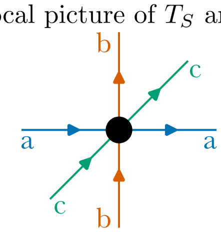
Each element of S corresponds to an oriented loop in X_S based at x_0, so we can regard $s \in S$ as an element of $\pi_1(X_S, x_0)$. This allows us to regard words in S as elements of $\pi_1(X_S, x_0)$. We now divide the proof into two steps:

STEP 1. *Each element of group $\pi_1(X_S, x_0)$ can be represented by a unique reduced word in S. Consequently, S is a subset*[6] *of $\pi_1(X_S, x_0)$ and $\pi_1(X_S, x_0)$ is generated by S.*

Define T_S to be an infinite tree each of whose vertices has valence $2|S|$. The oriented edges of T_S are labeled by elements of S, and for each vertex v of T_S there are:

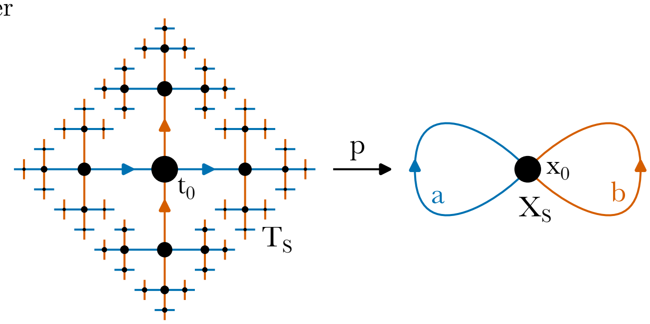
- $|S|$ edges coming out of v labeled by elements of S; and
- $|S|$ edges going into v labeled by elements of S.

For instance, if $S = \{a, b, c\}$ then the local picture of T_S around v looks like



This uniquely specifies T_S.

There is a covering space $p\colon T_S \to X_S$ taking each vertex of T_S to x_0 and each oriented edge of T_S labeled by $s \in S$ to the corresponding loop in X_S labeled by s. For instance, if $S = \{a, b\}$ this is the now-familiar cover



[6]Before this step we explained how to regard elements of S as elements of $\pi_1(X_S, x_0)$, but until this step is complete it is possible that some elements of S might be the trivial element of $\pi_1(X_S, x_0)$ or that two distinct elements of S might represent the same element of $\pi_1(X_S, x_0)$.

Just like in the $S = \{a, b\}$ case, the cover $p\colon T_S \to X_S$ is regular.[7] Since T_S is a tree, it is contractible and hence simply-connected (cf. Lemma 4.4.12).

Fix a vertex t_0 of T_S. Letting G be the deck group of $p\colon T_S \to X$, we can apply Theorem 4.5.1 to see that

$$\pi_1(X_S, x_0) \cong G.$$

The group G acts simply transitively on the vertices of T, so each vertex is of the form gt_0 for some unique $g \in G$. From the description of the isomorphism in §4.5.2, the element of $\pi_1(X, x_0)$ corresponding to $g \in G$ is the homotopy class of the loop in X based at x_0 obtained by taking a path in T from t_0 to gt_0 and projecting it to X.

For each vertex t_1 of T, there is a unique sequence of edges connecting t_0 to t_1 that does not backtrack, that is, traverse an edge in one direction and then go backwards along the same edge. This non-backtracking condition is exactly what is needed to ensure that this edge-path corresponds to an element of $\pi_1(X_S, x_0)$ represented by a reduced word. In this way, we see that every element of $\pi_1(X_S, x_0)$ is represented by a unique reduced word in S, as desired.

STEP 2. *The group $\pi_1(X_S, x_0)$ is a free group on S.*

This could be proved algebraically from Step 1, but we will instead give a geometric proof. All we will use from Step 1 is[8] that $S \subset \pi_1(X_S, x_0)$ and that S generates $\pi_1(X_S, x_0)$. Consider a group G and a set map $h\colon S \to G$. Our goal is to prove that h extends uniquely to a homomorphism $H\colon \pi_1(X_S, x_0) \to G$. Since S generates $\pi_1(X_S, x_0)$, uniqueness is immediate and we only need to prove existence.

By Theorem 5.3.7, we can find a pointed space (Y, y_0) with $\pi_1(Y, y_0) \cong G$. Identify $\pi_1(Y, y_0)$ with G. For each $s \in S$, write $h(s) = [\gamma_s] \in \pi_1(Y, y_0)$. We can then define a map $\phi\colon (X_S, x_0) \to (Y, y_0)$ by requiring $\phi(x_0) = y_0$ and then letting ϕ map the loop labeled by s to γ_s. By construction, $H = \phi_*\colon \pi_1(X_S, x_0) \to \pi_1(Y, y_0) = G$ extends h. □

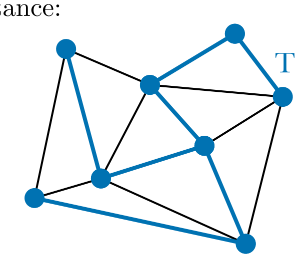
5.4.3. Graphs with one vertex. For later use, we extract the following result from the proof of Theorem 5.4.3:

COROLLARY 5.4.4. *Let S be a set and let X_S be a graph with one vertex x_0 and with $|S|$ oriented edges, each labeled with an element of S. Identify each $s \in S$ with the corresponding loop in X_S based at x_0. Then $\pi_1(X_S, x_0)$ is a free group on $\{[s] \mid s \in S\}$.*

5.5. Fundamental groups of graphs

We close this chapter by showing how to calculate the fundamental group of an arbitrary graph.

5.5.1. Maximal trees. Recall that a tree is a nonempty connected graph with no cycles. Each tree is contractible (Lemma 4.4.12). For a graph X, a *maximal tree* in X is a subtree T of X that contains every vertex of X. For instance:



These always exist:

LEMMA 5.5.1. *Let X be a nonempty connected graph. Then X contains a maximal tree.*

[7]The point here is that T is a tree each of whose vertices has the same local picture, so there are edge-label preserving graph automorphisms of T taking any vertex to any other vertex. These are deck transformations.

[8]If as described in the footnote before the definition of a free group we did not require the inclusion $S \hookrightarrow F(S)$ to be injective, then we could get away here with only knowing that the image of S generates $\pi_1(X_S, x_0)$.

Proof. Inductively define subtrees

$$T_0 \subset T_1 \subset T_2 \subset \cdots$$

of X in the following way. Start by choosing a vertex v_0 of X and letting $T_0 = v_0$. Next, if T_{n-1} has been constructed, let T_n be the subtree obtained from T_{n-1} as follows:

- For each vertex v of X that does not lie in T_{n-1} but is connected by an edge to T_{n-1}, choose an edge connecting T_{n-1} to v and add it to T_n.

Now define

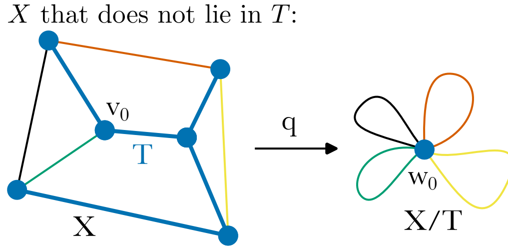
$$T = \bigcup_{n=0}^{\infty} T_n.$$

This is a subgraph of G. Since a cycle in T only involves finitely many edges, a cycle of T must be contained in some T_n. Since each T_n is a tree, it follows that T has no cycles, so T is a tree. Since X is connected, each vertex of X must lie in T, so T is a maximal tree. □

5.5.2. Graphs. Using maximal trees, we will prove:

Theorem 5.5.2. *Let X be a connected graph and let v_0 be a vertex of X. Then $\pi_1(X, v_0)$ is a free group.*

Proof. Let T be a maximal tree in X. The quotient graph X/T contains a single vertex w_0 and a loop for each edge of X that does not lie in T:



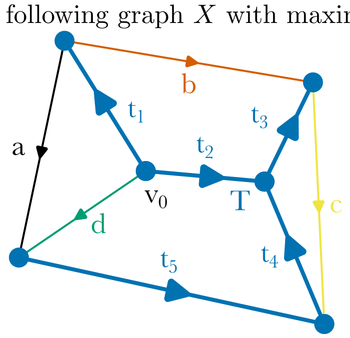
We saw in the proof of Theorem 5.4.3 that $\pi_1(X, w_0)$ is a free group. Since T is a contractible subspace of X with a mapping cylinder neighborhood, the quotient map $q\colon X \to X/T$ is a homotopy equivalence. The induced map $q_*\colon \pi_1(X, v_0) \to \pi_1(W, w_0)$ is therefore an isomorphism, so $\pi_1(X, v_0)$ is also a free group. □

5.5.3. Free bases. Let X be a connected graph and let v_0 be a vertex of X. By analyzing the proof of Theorem 5.5.2, we can construct a free basis for the free group $\pi_1(X, v_0)$. Begin by choosing a maximal tree T of X. Let w_0 be the single vertex of X/T. As in the proof of Theorem 5.5.2, the quotient map $q\colon X \to X/T$ induces an isomorphism $q_*\colon \pi_1(X, v_0) \to \pi_1(X/T, w_0)$.

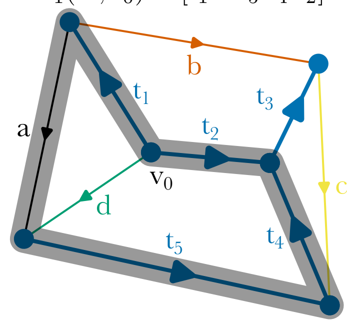
Let $\{e_i \mid i \in I\}$ be the edges of X that do *not* lie in T. The map q maps each e_i to a loop $\overline{e}_i$ in X/T that is based at w_0. Recall our convention that each edge of a graph is oriented (cf. §1.6). Using the orientation on $\overline{e}_i$, we get an element $[\overline{e}_i] \in \pi_1(X/T, w_0)$. By Corollary 5.4.4, the set $\{[\overline{e}_i] \mid i \in I\}$ is a basis for the free group $\pi_1(X/T, w_0)$.

For each $i \in I$, we must lift $\overline{e}_i$ to a loop in X that is based at v_0. To do this, let t_i be a path in T from v_0 to the initial vertex of the edge e_i and let t_i' be a path in T from the terminal vertex of e_i back to v_0. We then have a loop $t_i \cdot e_i \cdot t_i'$ in X based at v_0, and $q_*([t_i \cdot e_i \cdot t_i']) = [\overline{e}_i]$. We deduce that $\{[t_i \cdot e_i \cdot t_i'] \mid i \in I\}$ is a basis for the free group $\pi_1(X, v_0)$.

Example 5.5.3. Consider the following graph X with maximal tree T:



We have labeled and shown the orientation on each edge of X. Following the above algorithm, we obtain a free basis for $\pi_1(X, v_0)$. There is one element of this basis for each edge $\{a, b, c, d\}$. For the edge a, the corresponding element of $\pi_1(X, v_0)$ is $[t_1 \cdot a \cdot t_5 \cdot t_4 \cdot \bar{t}_2]$:



We similarly get basis elements corresponding to b and c and d. In summary, the following is a free basis for $\pi_1(X_S, v_0)$:

$$\{[t_1 \cdot a \cdot t_5 \cdot t_4 \cdot \bar{t}_2], [t_1 \cdot b \cdot \bar{t}_3 \cdot \bar{t}_2], [t_2 \cdot t_3 \cdot c \cdot \bar{t}_4 \cdot \bar{t}_2], [d \cdot t_5 \cdot t_4 \cdot \bar{t}_2]\}. \qquad \square$$

We will give more examples of this in the next chapter when we discuss covers of graphs.

5.6. Exercises

EXERCISE 5.1. Prove that being homotopy equivalent is an equivalence relation on pointed spaces and on spaces. □

EXERCISE 5.2. Define

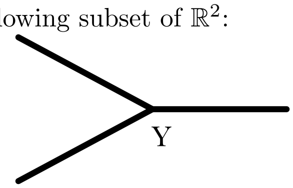
$$\Delta^n = \left\{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1, \ldots, x_{n+1} \geq 0 \text{ and } x_1 + \cdots + x_{n+1} = 1\right\}.$$

The space Δ^n is called an *n-simplex*. Prove the following:

(a) $\Delta^n \cong \mathbb{D}^n$.
(b) $\Delta^n * \Delta^m \cong \Delta^{n+m+1}$. Hint: recall that elements of the join $\Delta^n * \Delta^m$ can be written as formal sums $s_1 x + s_2 y$ with $x \in \Delta^n$ and $y \in \Delta^m$ and $s_1, s_2 \in I$ such that $s_1 + s_2 = 1$, where $s_1 x$ disappears if $s_1 = 0$ and $s_2 y$ disappears if $s_2 = 0$. Use this to write an explicit homeomorphism $\phi\colon \Delta^n * \Delta^m \to \Delta^{n+m+1}$. □

EXERCISE 5.3. Prove that $\mathbb{S}^n * \mathbb{S}^m \cong \mathbb{S}^{n+m+1}$. Hint: this is similar to Exercise 5.2. □

EXERCISE 5.4. Let Y be the following subset of $\mathbb{R}^2$:



Prove that $\mathbb{R}^2/Y \cong \mathbb{R}^2$. □

EXERCISE 5.5. Let M^n be a connected n-manifold with $n \geq 2$. Let $\{p_1, \ldots, p_k\}$ and $\{q_1, \ldots, q_k\}$ be two sets of k distinct points on M^n. Prove that there exists a homeomorphism $f\colon M^n \to M^n$ such that $f(p_i) = q_i$ for all $1 \leq i \leq k$. Hint: define

$$X_k(M^n) = \left\{(x_1, \ldots, x_k) \in (M^n)^{\times k} \mid x_i \neq x_j \text{ for all } 1 \leq i, j \leq k \text{ distinct}\right\}.$$

Prove that $X_k(M^n)$ is connected (this is where you will use $n \geq 2$). Next, prove that the orbit of $(p_1, \ldots, p_k) \in X_k(M^n)$ under the homeomorphism group of M^n is both open and closed (reduce this to a local statement about a single point in $\mathbb{R}^n$), and thus is all of $X_k(M^n)$. □

EXERCISE 5.6. Prove that the free action of G on X we constructed in the proof of Theorem 5.3.7 is a covering space action (see §1.7.1 for the definition). □

Part 2

Essays