Chapter 11

The Fundamental Group

We introduce the fundamental group as the first algebraic invariant associated to a topological space. The fundamental group, $\pi_1(X, x_0)$, is the primary tool in algebraic topology for distinguishing spaces based on the "holes" they contain. We begin by developing the notion of homotopy, which formalizes the notion of continuously deforming one map into another. We give particular emphasis on homotopies of paths and loops, and explain how path homotopy leads naturally to an equivalence relation on loops based at a point. Using concatenation of paths, we show how these equivalence classes acquire a group structure. The central example of the chapter is the circle S^1: Using the lifting properties of the exponential map (the covering map $p : \mathbb{R} \to S^1$) and analyzing how loops wind around the circle, we compute its fundamental group explicitly and show that

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

This computation provides an essential bridge between the abstract machinery of homotopy and concrete algebraic structures. Along the way, we emphasize the geometric meaning of homotopy and winding number, setting the stage for further applications of covering spaces and higher homotopy invariants.

We assume familiarity with basic point-set topology, e.g. Ch. 1–7 of [**Mor24**].

Acknowledgements. The exposition in Section 1 is modeled after [**Mun00**, Sec. 51–52]. Our proof of $\pi_1(S^1) \cong \mathbb{Z}$ follows that of [**Hat02**, Theorem 1.7].

1. Defining the Fundamental Group

Let $I = [0, 1]$ denote the unit interval.

Definition 1.1. Suppose f and f' are continuous maps $X \to Y$. We say that f is *homotopic* to f' if there exists a continuous map

$$F : X \times I \to Y$$

such that $F(x, 0) = f(x)$ and $F(x, 1) = f'(x)$. We say that F is a *homotopy* between f and f', and we write $f \cong f'$ to say that f and f' are homotopic.

We say that a path is *nullhomtopic* if it is homotopic to a constant map. Intuitively, a homotopy represents a way to continuously transform f into f'. Indeed, by considering $F(x, t)$ and varying t from 0 to 1, we start at f and end up at f'.

We will mostly be interested in the case where f and f' are *paths*, i.e. continuous maps $[0, 1] \to X$. For a path f, we define the *initial point* and *final point* in the obvious way: $f(0) = x_0$ is the initial point, and $f(1) = x_1$ is the final point.

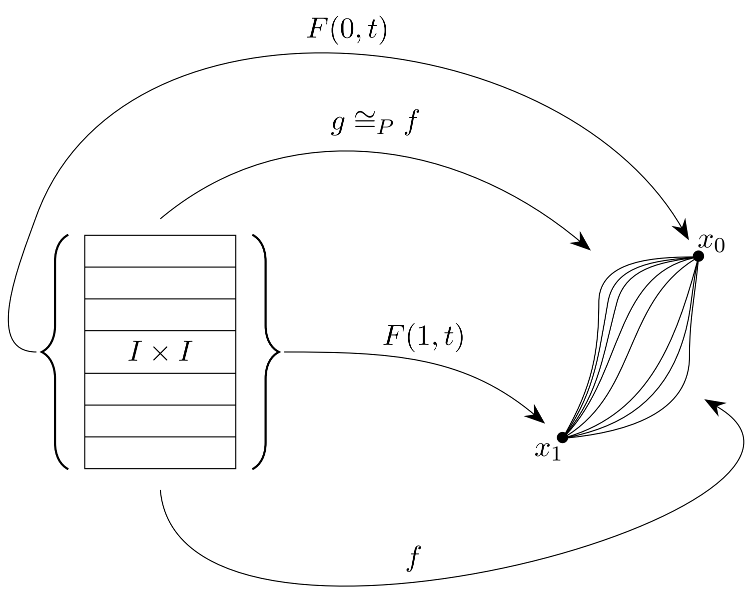

Figure 1. Path homotopy

An interesting question to ask is when two different paths from x_0 to x_1 are "equivalent". This is formalized by the notion of path homotopy, which is stronger than ordinary homotopy.

Definition 1.2. Two paths $f, f' : I \to X$ are said to be *path homotopic* if they have the same initial and final points x_0, y_0, and there exists a homotopy $F : I \times I \to X$ between them such that $F(0,t) = x_0$, $F(1,t) = x_1$ for all $t \in I$. We call such a homotopy a *path homotopy*, and write $f \cong_P f'$ to say that f and f' are homotopic.

So a path homotopy imposes the additional constraint that the endpoints must stay fixed *throughout the continuous transformation.* A visualization of path homotopy is shown in Figure 1.

Example 1.1. To see why this notion is stronger than ordinary homotopy, consider the paths f, f' from $(1,0)$ to $(0,-1)$ in $X = \mathbb{R}^2/\{(0,0)\}$ given by

$$f(x) = (\cos(\pi x), \sin(\pi x)) \quad \text{and} \quad f'(x) = (\cos(\pi x), -\sin(\pi x)).$$

These two paths are homotopic because we can continuously transform one into the other, as shown in Figure 2.

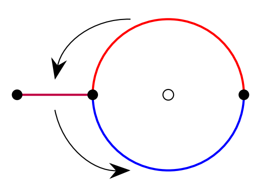

Figure 2. Homotopy vs. path homotopy

However, if the endpoints are fixed, it is impossible to continuously transform f into f' without hitting $(0,0)$, so these two paths are not path homotopic. However, rigorously proving this requires some work. It follows from our computation of the fundamental group of the circle below, along with something called *deformation retracts.*

Next, we have a simple lemma that tells us that homotopy behaves as we would expect it to.

Lemma 1.1. *The homotopy and path homotopy relations $\cong$ and $\cong_P$ are equivalence relations.*

Proof. That $f \cong f$ and $f \cong_P f$ are obvious; the map $F(s,t) = f(s)$ gives the desired (path) homotopy.

If $f \cong g$, corresponding to the homotopy $F : I \times I \to X$, then $F(s, 1-t)$ gives a homotopy $g \cong f$, and similarly for the case of path-homotopy.

Finally, if $f \cong g$ and $g \cong h$, with homotopies F and G, then $H : I \times I \to X$ defined by

$$H(s,t) = \begin{cases} F(s,2t) & t \in [0, \frac{1}{2}] \\ G(s, 2t-1) & t \in [\frac{1}{2}, 1] \end{cases}$$

is a homotopy $f \cong h$. Again, the path-homotopy case is similar. $\square$

Next, we define a way to "glue paths together".

Definition 1.3. Suppose f is a path in X from x_0 to x_1 and g is a path from x_1 to x_2. We define the product $f * g$ by

$$(f * g)(s) = \begin{cases} f(2s) & s \in [0, \frac{1}{2}], \\ g(2s-1) & s \in [\frac{1}{2}, 1]. \end{cases}$$

It is easily checked that the product $f * g$ is a path from x_0 to x_2, and furthermore, this operation is well-defined up to path-homotopy, i.e. if $f \cong_P f'$ and $g \cong_P g'$, then $f * g \cong_P f' * g'$. Hence, using $[f]$ to denote the equivalence class of f under path-homotopy, we can write $[f] * [g] = [f * g]$. Next, we show that the product of (equivalence classes of) paths satisfies the axioms that we would expect a product to have.

Let for $x \in X$, e_x denote the constant path sending all of I to the point x.

Theorem 1.2. *The product operation satisfies the following:*

- *Associativity:*
$$[f] * ([g] * [h]) = ([f] * [g]) * [h],$$
when $f(1) = g(0)$ *and* $g(1) = h(0)$.
- *Identities:*
$$[f] * [e_{x_1}] = [f] \quad \textit{and} \quad [e_{x_0}] * [f] = [f],$$
where f *is a path from* x_0 *to* x_1.
- *Inverses:*
$$[f] * [\overline{f}] = [e_{x_0}] \quad \textit{and} \quad [\overline{f}] * [f] = [e_{x_1}],$$
where f *is a path from* x_0 *to* x_1 *and* $\overline{f}(s) = f(1-s)$ *is the* reverse *of* f.

Proof. These properties are all quite "obvious", but require a bit of care to prove. We freely use the following two easy observations.

First, if $k : X \to Y$ is continuous, and F is a path homotopy in X between two paths f and f', then $k \circ F$ is a path homotopy in Y between the paths $k \circ f$ and $k \circ f'$.

Second, if $k : X \to Y$ is continuous, and f and g are paths in X with $f(1) = g(0)$, then $k \circ (f * g) = (k \circ f) * (k \circ g)$.

Now we prove the three properties.

- Associativity: Consider the function p whose graph is shown in blue in Figure 3. This function can be thought of as a path in I from 0 to 1. It is clearly homeomorphic to the identity map $i : I \to I$ (which is the red graph). A homeomorphism can be written down explicitly, but it is much easier to see this visually. Then, applying $(f * g) * h$ to these two

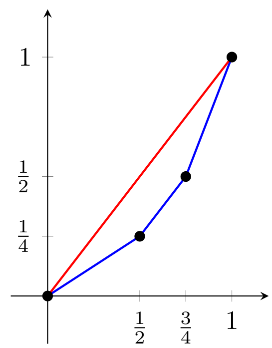

Figure 3. Associativity

  paths, we get $(f*g)*h = ((f*g)*h) \circ i$ and $f*(g*h) = ((f*g)*h) \circ p$. Hence, the paths $(f * g) * h$ and $f * (g * h)$ are homotopic.

- Identity: Consider the two paths in I from 0 to 1 [Figure 4]. It is clear that these are homotopic. Applying the continuous map $f : I \to X$, it follows that f and $e_{x_0} * f$ are homotopic. We have a similar proof for the second statement.

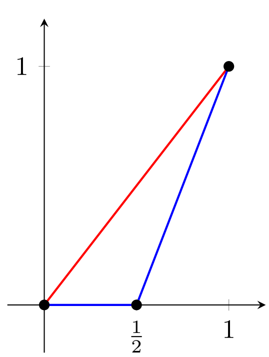

Figure 4. Identities

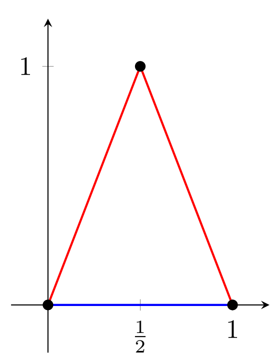

Figure 5. Inverses

- Inverse: Consider the two paths in I from 0 to 0 [Figure 5]. As in the previous case, they are evidently homotopic, and applying $f : I \to X$, the red path maps to $f * \overline{f}$, and the blue path maps to e_{x_0}. Hence $f * \overline{f}$ and e_{x_0} are homotopic.

This completes the proof. $\square$

We can finally define the fundamental group.

Let X be a space, and suppose x_0 is a point of X. We call a path on X that starts and ends at x_0 a *loop* based at x_0. Note that $*$ defines a binary operation on the set of loops. By Theorem 1.2, it in fact defines a group structure, with e_{x_0} as the identity.

Definition 1.4. The *fundamental group* of X relative to the *base point* x_0 is the set of loops based at x_0, with the operation $*$. We use $\pi_1(X, x_0)$ to denote this group.

This is also called the *first homotopy groups*. There exist higher-order homotopy groups; in fact there is a $\pi_n(X, x_0)$ for every positive integer n. We will not study these here.

As one can probably guess from considering various examples of X, such as the unit circle, the figure-eight, or a torus, the fundamental groups should not really depend on the base point.

This turns out to be the case. Suppose x_0, x_1 are two distinct points in X, and we have a path α from x_0 to x_1. Then, the map

$$\widehat{\alpha} : \pi_1(X, x_0) \to \pi_1(X, x_1)$$

given by

$$\widehat{\alpha}([f]) = [\overline{\alpha}] * [f] * [\alpha]$$

defines an isomorphism between the two groups. Before we prove this, we remark that the definition of this homomorphism resembles the conjugation action in a group; and indeed the proof that this is an isomorphism mirrors the proof that conjugation is an automorphism.

To check that this is a homomorphism, note that

$$\begin{aligned}\widehat{\alpha}([f] * [g]) &= [\overline{\alpha}] * [f] * [g] * [\alpha] \\ &= [\overline{\alpha}] * [f] * [\alpha] * [\overline{\alpha}] * [g] * [\alpha] \\ &= \widehat{\alpha}([f]) * \widehat{\alpha}([g])/\end{aligned}$$

Finally, note that $\widehat{\overline{\alpha}}$ is an inverse to $\widehat{\alpha}$, so α is an isomorphism.

Consequently, we have

Proposition 1.3. *If X is path-connected, then all the fundamental groups $\pi_1(X, x_0)$ are isomorphic.*

Since we can split any topological space into path-connected components, it suffices to consider path-connected spaces when dealing with the fundamental group.

For spaces like $\mathbb{R}^n$, or a n-dimensional ball, (or in fact any convex subset of $\mathbb{R}^n$) it is not too hard to see that any loop is nullhomotopic, which implies that the fundamental group is trivial. Such spaces have a name.

Definition 1.5. A space X is called *simply connected* if it is path-connected and its fundamental group is trivial.

2. The Fundamental Group of the circle

Let S^1 denote the circle (the subspace of $\mathbb{R}^2$ of points that are a distance 1 from the origin). In this section, we prove that the fundamental group of the circle is isomorphic to $\mathbb{Z}$.

It is clear that for each positive integer n, looping counterclockwise around the unit circle n times gives a loop $[f_n] \in \pi_n$. Similarly, we can define a loop $[f_{-n}]$ which loops n times in the clockwise direction. Adding the identity loop f_0, we get a loop f_n for all $n \in \mathbb{Z}$, and furthermore it is clear that

$$[f_{m+n}] = [f_m] * [f_n],$$

so the group operation matches that of $\mathbb{Z}$. However, we are not even close to done! We need to show that (a) these loops are non-homotopic and (b) there are no other homotopy classes of loops. To do this, we use a very clever idea called *covering maps*.

The essence of this idea is that we use the homotopy structure of a space we are familiar with (like $\mathbb{R}$) to understand the homotopy structure of a more complicated group.

Consider the map $p : \mathbb{R} \to S^1$ given by

$$p(x) = (\cos(2\pi x), \sin(2\pi x)).$$

This is evidently a continuous map. Intuitively, we can think of p as "wrapping" $\mathbb{R}$ around the unit circle. In another way, by plotting the points $(x, p(x))$ in $\mathbb{R}^3$, we get a helix, and we are "collapsing" this helix onto the unit circle, as shown in Figure 6.

Given a path $f : I \to S^1$, we define a *lifting* of f to $\mathbb{R}$ to be a path $\widetilde{f}$ such that $p \circ \widetilde{f} = f$.

Proposition 2.1. *We have the following.*

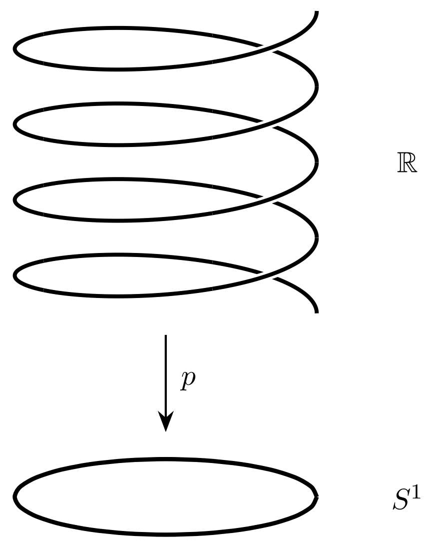

Figure 6. Collapsing the real line ($\mathbb{R}$) onto a circle (S^1)

(1) Any path $f : I \to S^1$ starting at $(1, 0)$ has a unique lifting to a path $\widetilde{f} : I \to \mathbb{R}$ that starts at 0.

(2) For any path homotopy $F : I \times I \to S^1$ between paths starting at $(1, 0)$, there is a unique lifting to a path homotopy $\widetilde{F} : I \times I \to \mathbb{R}$.

These are both direct consequences of the following, more general result.

Theorem 2.2. *Suppose we have maps $F\colon Y \times I \to S^1$ and $\widetilde{F}\colon Y \times \{0\} \to \mathbb{R}$ so that $\widetilde{F}$ lifts $F|_{Y\times\{0\}}$. Then there is a unique map $\widetilde{F}\colon Y \times I \to \mathbb{R}$ that lifts F and restricts to the given $\widetilde{F}$ on $Y \times \{0\}$.*

Proof. This only relies on one important property of the projection $p : \mathbb{R} \to S^1$, which is the following.

($*$) *There is an open cover $\{U_\alpha\}_{\alpha\in A}$ of S^1 such that for each α, $p^{-1}(U_\alpha)$ can partitioned into (disjoint) open sets such that each of them is mapped homeomorphically onto U_α by p.*

In our specific case, we could take U_1, U_2 to be two open arcs that cover S^1.

The idea behind the proof is as follows: due to **($*$)**, if we "zoom in" on a sufficiently small neighborhood, p acts like a homeomorphism, so on this neighborhood, defining $\widetilde{F}$ is easy. Then, we can "patch together" all of the neighborhoods to get the function $\widetilde{F}$.

We start with some point $y_0 \in Y$, and we will define $\widetilde{F}$ on $N \times I$, for some neighborhood N of y_0 in Y. Since F is continuous, every point (y_0, t) has a neighborhood of the form $N_t \times (a_t, b_t)$ for some neighborhood N_t of y_0. Since $\{y_0\} \times I$ is compact, we can pick finitely many of these that cover it. Intersecting all the N_t corresponding to these neighborhoods, we get a neighborhood N of y_0 and some partition $0 = t_0 < t_1 < \cdots < t_m = 1$ such that $F(N \times [t_k, t_{k+1}])$ is contained in some U_α for all k. The neighborhood N will not be the final neighborhood we end up with, because we will be forced to make it smaller as we apply the "patching-together argument", starting from $\widetilde{F}(N \times \{0\})$ (which is given to us since $\widetilde{F}$ is already defined on $Y \times \{0\}$).

Assume inductively that we have defined $\widetilde{F}$ on $N \times [0, t_k]$,, for $0 \leq k \leq m-1$. Then $F(N \times [t_k, t_{k+1}]) \subset U_\alpha$ for some α, and $\widetilde{F}(y_0, t_k) \in V$, where V is one of the homeomorphic images of U_α in $p^{-1}(U_\alpha)$ coming from the property **(∗)**. By replacing N with $N \cap \widetilde{F}[N \times \{t_k\}]^{-1}(V))$ (i.e. the set of $n \in N$ so that $\widetilde{F}(n, t_k) \in V$), we may assume that $\widetilde{F}(N \times \{t_k\}) \subset V$. Now, defining $\widetilde{F}$ on $N \times [t_k, t_{k+1}]$ is easy: we can simply set it to be the composition of the maps F and $p^{-1} : U \to V$ (which exists since $p|_V : V \to U$ is a homeomorphism). Repeating this process finitely many times, we get the desired lift $\widetilde{F} : N \times I \to \mathbb{R}$.

Now, we would like to be able to path all of the $\widetilde{F}$ constructed above together to get the final function $\widetilde{F}$. However, to do this, we need to know that if two such neighborhoods intersect, then the corresponding $\widetilde{F}$ must agree. To do this, we first prove that the map $\widetilde{F}$ is unique when Y is a point.

In this case, we can just consider $\widetilde{F}$ and $\widetilde{F}'$ as maps $I \to S^1$ and $I \to \mathbb{R}$, respectively, such that $\tilde{F}(0)$ is given. Suppose we have two lifts $\widetilde{F}$ and $\widetilde{F}'$; then $\widetilde{F}(0) = \widetilde{F}(0)'$. Then, running a very similar version of above argument, we can get a partition

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

such that for each k, $F([t_k, t_{k+1}])$ is contained in some U_α. Then, assume inductively that $F = \widetilde{F}$ on $[0, t_k]$. Since $\widetilde{F}([t_k, t_{k+1})]$ is connected, it must be contained in some V appearing in the partition of $p^{-1}(U_\alpha)$ so that $p_V : V \to U$ is a homemomorphism. Furthermore, this V is forced by the value of $\widetilde{F}(t_k) = \widetilde{F}'(t_k)$. It similarly follows that $\widetilde{F}'([t_k, t_{k+1}])$ is also contained in V. Since p is a bijection $V \to U$ and both $\widetilde{F}$ and $\widetilde{F}'$ are lifts of F, it follows that $\widetilde{F} = \widetilde{F}'$ on $[t_k, t_{k+1}]$. By induction, it follows that $\widetilde{F} = \widetilde{F}'$.

Now, getting back to the general case, note that $\widetilde{F}(y_0 \times I)$ is a valid lift of $F(y_0 \times I)$, and corresponds to the case when y is a point. Hence $\widetilde{F}(y_0 \times I)$ is unique. Thus whenever two neighborhoods of the form $N \times I$ intersect, their corresponding $\widetilde{F}$ defined above must agree on the intersection. This allows us to "patch together" these $\widetilde{F}$ to get a valid lifting of F. Finally, this function is unique since each $\widetilde{F}(y_0 \times I)$ is. □

Theorem 2.3. *The fundamental group of S^1 is isomorphic to $\mathbb{Z}$.*

Proof. For a path $f : I \to \mathbb{R}$ from 0 to some $n \in \mathbb{Z}$, the path $p \circ f$ is a loop based at $(1, 0)$ in S^1. Since p is continuous, homotopic paths f give homotopic loops, so we can define

$$\Phi(n) = [p \circ f].$$

We claim that this is an isomorphism $\mathbb{Z} \to \pi_1(S^1, (1, 0))$. First, we check that this is a homomorphism. If f is a path in $\mathbb{R}$ from 0 to n and g is a path from 0 to m, we prove that

$$\Phi(m + n) = \Phi(m) * \Phi(n) = (p \circ f) * (p \circ g)$$

as follows. Let $h = f * (g + n)$, where $g + n$ is the path given by

$$(g + n)(s) = g(s) + n.$$

Note that $p(h) = (p \circ f) * (p \circ (g + n))$. However, $p \circ (g + n) = p \circ g$ since p is 1-periodic. Hence, $\Phi(m + n) = [p(h)] = (p \circ f) * (p \circ g)$.

Next, we prove bijectivity. By Theorem 2.1, every loop in $\pi_1(S^1, (1, 0))$ has a unique lifting to a path $\widetilde{f} : I \to \mathbb{R}$ that starts at 0. The ending point of this path must be some integer n, and it follows that $p \circ \widetilde{f} = f$, so $f \in \Phi(n)$. Hence, Φ is surjective. Finally, suppose that $\Phi(n) = \Phi(m)$. Then, $f_0 \in \Phi(n)$ and $f_1 = \Phi(m)$ are homotopic. By part (2) of Theorem 2.1, this homotopy lifts to a unique path homotopy $\widetilde{F} : I \times I \to \mathbb{R}$. By the definition of a path homotopy, it follows that $\widetilde{F}(s, 0)$ and $\widetilde{F}(s, 1)$ have the same ending points. These are lifts of f_0 and f_1, respectively. By uniqueness of these lifts, the ending points must be m and n, so $m = n$. Hence, Φ is injective, and this completes the proof. □

Remark 2.1. Theorem 2.3 has many nice consequences. For example, it can be used to give a very clean proof of the Fundamental Theorem of Algebra, which states that every nonconstant polynomial $p(x)$ over $\mathbb{C}$ has a root in $\mathbb{C}$. Given a proposed counterexample to this statement, the idea is to consider the function

$$f_r(s) = \frac{p(re^{2\pi i s})/p(r)}{|p(re^{2\pi i s})/p(r)|}$$

which defines a loop in S^1 for any $r \geq 0$. If p has no roots, then $f_r(s)$ continuously varies as r does, and it follows that all the f_r are homotopic. However, as $r \to \infty$, the lower-order terms become insignificant and $f_r(s)$ approaches the loop $\Phi(n)$, where n is the degree of f. However, as $r \to 0$, $f_r(s)$ approaches the trivial loop. But by Theorem 2.3, these two loops are not homeomorphic in S^1, which yields a contradiction.

For another application, we can use Theorem 2.3 to prove that any continuous mapping of the closed unit disk (in $\mathbb{R}^2$) to itself has a fixed point. This is called the Brouwer fixed point theorem (in dimension 2).

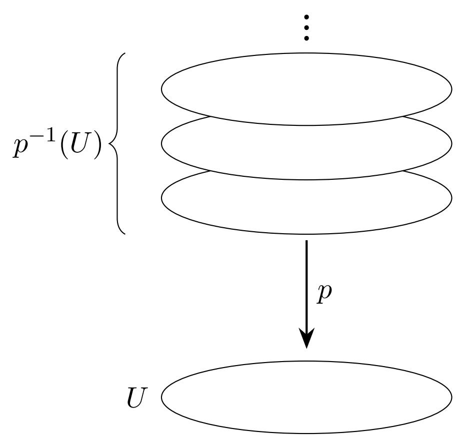

Figure 7. Covering maps

The map p in the above argument is an example of a *covering map*. In general, a covering map is a continuous map $p : E \to B$ satisfying the property $(*)$, with $\mathbb{R}$ replaced by E and S^1 replaced by B. We say that E is a *covering space* of B. See Figure 7 for a visualization of a covering map. The proof of Theorem 2.2 works just as well for arbitrary covering maps, and indeed, this is the fundamental property of covering maps that makes them so powerful. In general, we get a correspondence $\pi_1(B, b_0) \to p^{-1}(b_0)$ that sends a loop to the ending point of its image from the generalized version of Theorem 2.2. If the space E is simply connected, then the ending point uniquely specifies the homotopy class of this path, and hence it uniquely specifies the loop. Therefore, this map is in fact a bijection. It follows that if we are able to find a simply connected covering space, we have essentially understood the fundamental group.

Indeed, this technique can be used to deduce a lot of useful information about fundamental groups. For another example, a well-chosen covering map tells us that the fundamental group of the figure-eight is *not* abelian (and hence not isomorphic to the fundamental group of the circle).

Bibliography

[Hat02] A. Hatcher, *Algebraic topology*, Cambridge University Press, 2002.

[Mor24] Sidney A. Morris, *Topology without tears*, 2024. Free online textbook, [Link].

[Mun00] James R. Munkres, *Topology*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 2000.