Chapter 1

The Fundamental Group

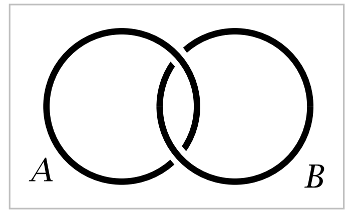
Algebraic topology can be roughly defined as the study of techniques for forming algebraic images of topological spaces. Most often these algebraic images are groups, but more elaborate structures such as rings, modules, and algebras also arise. The mechanisms that create these images — the 'lanterns' of algebraic topology, one might say — are known formally as *functors* and have the characteristic feature that they form images not only of spaces but also of maps. Thus, continuous maps between spaces are projected onto homomorphisms between their algebraic images, so topologically related spaces have algebraically related images.

With suitably constructed lanterns one might hope to be able to form images with enough detail to reconstruct accurately the shapes of all spaces, or at least of large and interesting classes of spaces. This is one of the main goals of algebraic topology, and to a surprising extent this goal is achieved. Of course, the lanterns necessary to do this are somewhat complicated pieces of machinery. But this machinery also has a certain intrinsic beauty.

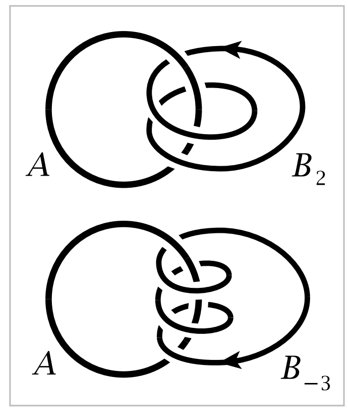
This first chapter introduces one of the simplest and most important functors of algebraic topology, the fundamental group, which creates an algebraic image of a space from the loops in the space, the paths in the space starting and ending at the same point.

The Idea of the Fundamental Group

To get a feeling for what the fundamental group is about, let us look at a few preliminary examples before giving the formal definitions.

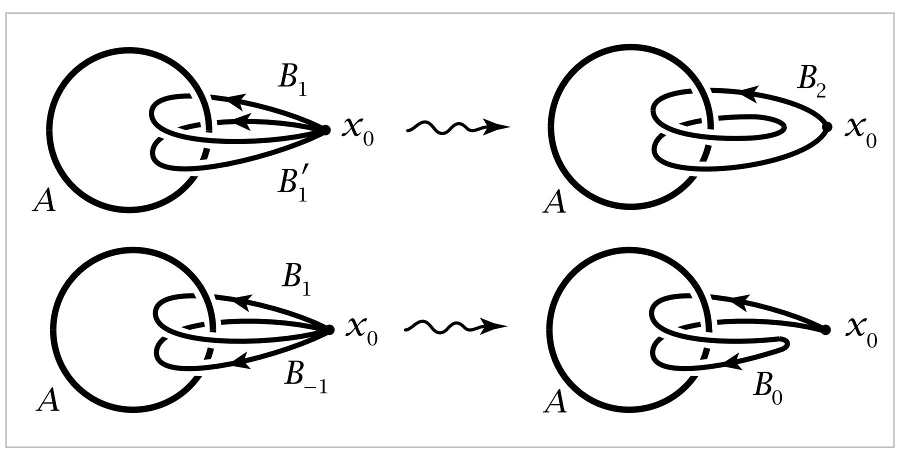


Consider two linked circles A and B in $\mathbb{R}^3$, as shown in the figure. Our experience with actual links and chains tells us that since the two circles are linked, it is impossible to separate B from A by any continuous motion of B, such as pushing, pulling, or twisting. We could even take B to be made of rubber or stretchable string and allow completely general continuous deformations of B, staying in the complement of A at all times, and it would still be impossible to pull B off A. At least that is what intuition suggests, and the fundamental group will give a way of making this intuition mathematically rigorous.



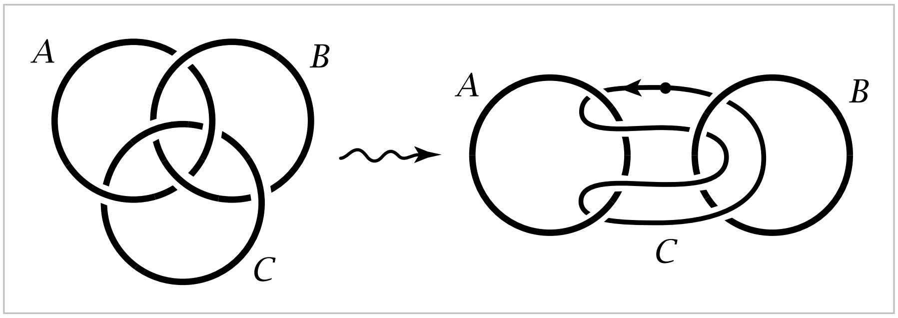
Instead of having B link with A just once, we could make it link with A two or more times, as in the figures to the right. As a further variation, by assigning an orientation to B we can speak of B linking A a positive or a negative number of times, say positive when B comes forward through A and negative for the reverse direction. Thus for each nonzero integer n we have an oriented circle B_n linking A n times, where by 'circle' we mean a curve homeomorphic to a circle. To complete the scheme, we could let B_0 be a circle not linked to A at all.

Now, integers not only measure quantity, but they form a group under addition. Can the group operation be mimicked geometrically with some sort of addition operation on the oriented circles B linking A? An oriented circle B can be thought of as a path traversed in time, starting and ending at the same point x_0, which we can choose to be any point on the circle. Such a path starting and ending at the same point is called a *loop*. Two different loops B and B' both starting and ending at the same point x_0 can be 'added' to form a new loop $B + B'$ that travels first around B, then around B'. For example, if B_1 and B_1' are loops each linking A once in the positive direction, then their sum $B_1 + B_1'$ is deformable to B_2, linking A twice. Similarly, $B_1 + B_{-1}$ can be deformed to the loop B_0, unlinked from A. More generally, we see that $B_m + B_n$ can be deformed to B_{m+n} for arbitrary integers m and n.

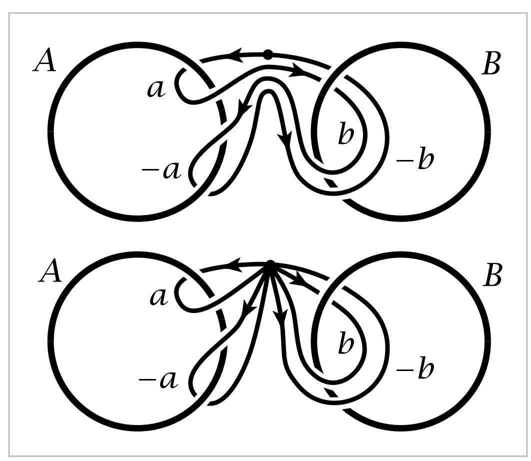


Note that in forming sums of loops we produce loops that pass through the basepoint more than once. This is one reason why loops are defined merely as continuous paths, which are allowed to pass through the same point many times. So if one is thinking of a loop as something made of stretchable string, one has to give the string the magical power of being able to pass through itself unharmed. However, we must be sure not to allow our loops to intersect the fixed circle A at any time, otherwise we could always unlink them from A.

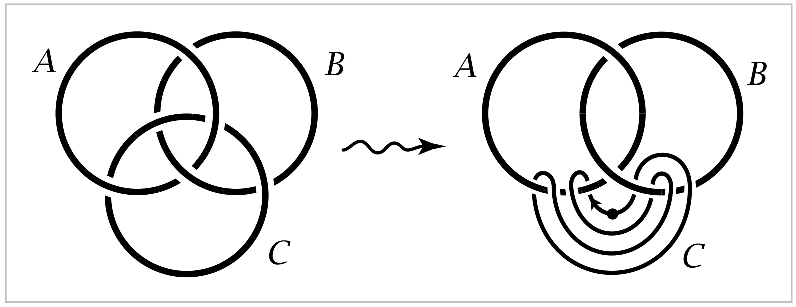
Next we consider a slightly more complicated sort of linking, involving three circles forming a configuration known as the Borromean rings, shown at the left in the figure below. The interesting feature here is that if any one of the three circles is removed, the other two are not linked. In the same spirit as before, let us regard one of the circles, say C, as a loop in the complement of the other two, A and B, and we ask whether C can be continuously deformed to unlink it completely from A and B, always staying in the complement of A and B during the deformation. We can redraw the picture by pulling A and B apart, dragging C along, and then we see C winding back and forth between A and B as shown in the second figure above. In this new position, if we start at the point of C indicated by the dot and proceed in the direction given by the arrow, then we pass in sequence: (1) forward through A, (2) forward through B, (3) backward through A, and (4) backward through B. If we measure the linking of C with A and B by two integers, then the 'forwards' and 'backwards' cancel and both integers are zero. This reflects the fact that C is not linked with A or B individually.



To get a more accurate measure of how C links with A and B together, we regard the four parts (1)–(4) of C as an ordered sequence. Taking into account the directions in which these segments of C pass through A and B, we may deform C to the sum $a + b - a - b$ of four loops as in the figure. We write the third and fourth loops as the negatives of the first two since they can be deformed to the first two, but with the opposite orientations, and as we saw in the preceding example, the sum of two oppositely oriented loops is deformable to a trivial loop, not linked with anything. We would like to view the expression $a + b - a - b$ as lying in a nonabelian group, so that it is not automatically zero. Changing to the more usual multiplicative notation for nonabelian groups, it would be written $aba^{-1}b^{-1}$, the commutator of a and b.



To shed further light on this example, suppose we modify it slightly so that the circles A and B are now linked, as in the next figure. The circle C can then be deformed into the position shown at the right, where it again represents the composite loop $aba^{-1}b^{-1}$, where a and b are loops linking A and B. But from the picture on the left it is apparent that C can actually be unlinked completely from A and B. So in this case the product $aba^{-1}b^{-1}$ should be trivial.



The fundamental group of a space X will be defined so that its elements are loops in X starting and ending at a fixed basepoint $x_0 \in X$, but two such loops are regarded as determining the same element of the fundamental group if one loop can be continuously deformed to the other within the space X. (All loops that occur during deformations must also start and end at x_0.) In the first example above, X is the complement of the circle A, while in the other two examples X is the complement of the two circles A and B. In the second section in this chapter we will show:

- The fundamental group of the complement of the circle A in the first example is infinite cyclic with the loop B as a generator. This amounts to saying that every loop in the complement of A can be deformed to one of the loops B_n, and that B_n cannot be deformed to B_m if $n \neq m$.
- The fundamental group of the complement of the two unlinked circles A and B in the second example is the nonabelian free group on two generators, represented by the loops a and b linking A and B. In particular, the commutator $aba^{-1}b^{-1}$ is a nontrivial element of this group.
- The fundamental group of the complement of the two linked circles A and B in the third example is the free abelian group on two generators, represented by the loops a and b linking A and B.

As a result of these calculations, we have two ways to tell when a pair of circles A and B is linked. The direct approach is given by the first example, where one circle is regarded as an element of the fundamental group of the complement of the other circle. An alternative and somewhat more subtle method is given by the second and third examples, where one distinguishes a pair of linked circles from a pair of unlinked circles by the fundamental group of their complement, which is abelian in one case and nonabelian in the other. This method is much more general: One can often show that two spaces are not homeomorphic by showing that their fundamental groups are not isomorphic, since it will be an easy consequence of the definition of the fundamental group that homeomorphic spaces have isomorphic fundamental groups.

1.1 Basic Constructions

This first section begins with the basic definitions and constructions, and then proceeds quickly to an important calculation, the fundamental group of the circle, using notions developed more fully in §1.3. More systematic methods of calculation are given in §1.2. These are sufficient to show for example that every group is realized as the fundamental group of some space. This idea is exploited in the Additional Topics at the end of the chapter, which give some illustrations of how algebraic facts about groups can be derived topologically, such as the fact that every subgroup of a free group is free.

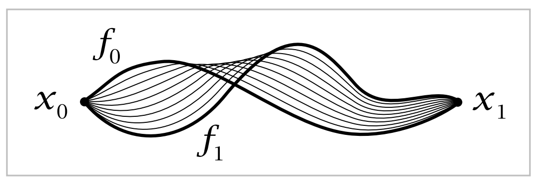
Paths and Homotopy

The fundamental group will be defined in terms of loops and deformations of loops. Sometimes it will be useful to consider more generally paths and their deformations, so we begin with this slight extra generality.

By a **path** in a space X we mean a continuous map $f:I\to X$ where I is the unit interval $[0,1]$. The idea of continuously deforming a path, keeping its endpoints fixed, is made precise by the following definition. A **homotopy** of paths in X is a family $f_t:I\to X$, $0\le t\le 1$, such that

(1) The endpoints $f_t(0)=x_0$ and $f_t(1)=x_1$ are independent of t.

(2) The associated map $F:I\times I\to X$ defined by $F(s,t)=f_t(s)$ is continuous.



When two paths f_0 and f_1 are connected in this way by a homotopy f_t, they are said to be **homotopic**. The notation for this is $f_0\simeq f_1$.

Example 1.1: Linear Homotopies. Any two paths f_0 and f_1 in $\mathbb{R}^n$ having the same endpoints x_0 and x_1 are homotopic via the homotopy $f_t(s)=(1-t)f_0(s)+tf_1(s)$. During this homotopy each point $f_0(s)$ travels along the line segment to $f_1(s)$ at constant speed. This is because the line through $f_0(s)$ and $f_1(s)$ is linearly parametrized as $f_0(s)+t[f_1(s)-f_0(s)]=(1-t)f_0(s)+tf_1(s)$, with the segment from $f_0(s)$ to $f_1(s)$ covered by t values in the interval from 0 to 1. If $f_1(s)$ happens to equal $f_0(s)$ then this segment degenerates to a point and $f_t(s)=f_0(s)$ for all t. This occurs in particular for $s=0$ and $s=1$, so each f_t is a path from x_0 to x_1. Continuity of the homotopy f_t as a map $I\times I\to\mathbb{R}^n$ follows from continuity of f_0 and f_1 since the algebraic operations of vector addition and scalar multiplication in the formula for f_t are continuous.

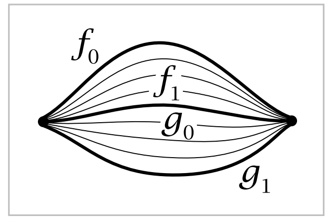
This construction shows more generally that for a convex subspace $X\subset\mathbb{R}^n$, all paths in X with given endpoints x_0 and x_1 are homotopic, since if f_0 and f_1 lie in X then so does the homotopy f_t.

Before proceeding further we need to verify a technical property:

Proposition 1.2. *The relation of homotopy on paths with fixed endpoints in any space is an equivalence relation.*

The equivalence class of a path f under the equivalence relation of homotopy will be denoted $[f]$ and called the **homotopy class** of f.

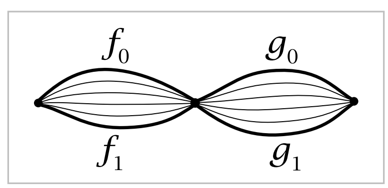
Proof: Reflexivity is evident since $f \simeq f$ by the constant homotopy $f_t = f$. Symmetry is also easy since if $f_0 \simeq f_1$ via f_t, then $f_1 \simeq f_0$ via the inverse homotopy f_{1-t}. For transitivity, if $f_0 \simeq f_1$ via f_t and if $f_1 = g_0$ with $g_0 \simeq g_1$ via g_t, then $f_0 \simeq g_1$ via the homotopy h_t that equals f_{2t} for $0 \le t \le {}^1\!/_2$ and g_{2t-1} for ${}^1\!/_2 \le t \le 1$. These two definitions agree for $t = {}^1\!/_2$ since we assume $f_1 = g_0$. Continuity of the associated map $H(s,t) = h_t(s)$ comes from the elementary fact, which will be used frequently without explicit mention, that a function defined on the union of two closed sets is continuous if it is continuous when restricted to each of the closed sets separately. In the case at hand we have $H(s,t) = F(s,2t)$ for $0 \le t \le {}^1\!/_2$ and $H(s,t) = G(s,2t-1)$ for ${}^1\!/_2 \le t \le 1$ where F and G are the maps $I \times I \to X$ associated to the homotopies f_t and g_t. Since H is continuous on $I \times [0, {}^1\!/_2]$ and on $I \times [{}^1\!/_2, 1]$, it is continuous on $I \times I$. $\square$



Given two paths $f,g : I \to X$ such that $f(1) = g(0)$, there is a **composition** or **product path** $f \cdot g$ that traverses first f and then g, defined by the formula

$$f \cdot g(s) = \begin{cases} f(2s), & 0 \le s \le {}^1\!/_2 \\ g(2s-1), & {}^1\!/_2 \le s \le 1 \end{cases}$$

Thus f and g are traversed twice as fast in order for $f \cdot g$ to be traversed in unit time. This product operation respects homotopy classes since if $f_0 \simeq f_1$ and $g_0 \simeq g_1$ via homotopies f_t and g_t, and if $f_0(1) = g_0(0)$ so that $f_0 \cdot g_0$ is defined, then $f_t \cdot g_t$ is defined and provides a homotopy $f_0 \cdot g_0 \simeq f_1 \cdot g_1$.



In particular, suppose we restrict attention to paths $f : I \to X$ with the same starting and ending point $f(0) = f(1) = x_0 \in X$. Such paths are called **loops**, and the common starting and ending point x_0 is referred to as the **basepoint**. The set of all homotopy classes $[f]$ of loops $f : I \to X$ at the basepoint x_0 is denoted $\pi_1(X, x_0)$.

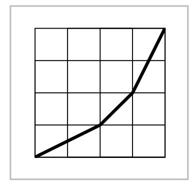
Proposition 1.3. *$\pi_1(X,x_0)$ is a group with respect to the product $[f][g] = [f \cdot g]$.*

This group is called the **fundamental group** of X at the basepoint x_0. We will see in Chapter 4 that $\pi_1(X,x_0)$ is the first in a sequence of groups $\pi_n(X,x_0)$, called homotopy groups, which are defined in an entirely analogous fashion using the n-dimensional cube I^n in place of I.

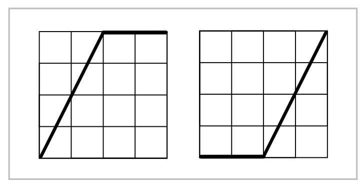
Proof: By restricting attention to loops with a fixed basepoint $x_0 \in X$ we guarantee that the product $f \cdot g$ of any two such loops is defined. We have already observed that the homotopy class of $f \cdot g$ depends only on the homotopy classes of f and g, so the product $[f][g] = [f \cdot g]$ is well-defined. It remains to verify the three axioms for a group.

As a preliminary step, define a **reparametrization** of a path f to be a composition $f\varphi$ where $\varphi : I \to I$ is any continuous map such that $\varphi(0) = 0$ and $\varphi(1) = 1$. Reparametrizing a path preserves its homotopy class since $f\varphi \simeq f$ via the homotopy $f\varphi_t$ where $\varphi_t(s) = (1-t)\varphi(s) + ts$ so that $\varphi_0 = \varphi$ and $\varphi_1(s) = s$. Note that $(1-t)\varphi(s) + ts$ lies between $\varphi(s)$ and s, hence is in I, so the composition $f\varphi_t$ is defined.

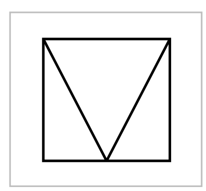
If we are given paths f, g, h with $f(1) = g(0)$ and $g(1) = h(0)$, then both products $(f \cdot g) \cdot h$ and $f \cdot (g \cdot h)$ are defined, and $f \cdot (g \cdot h)$ is a reparametrization of $(f \cdot g) \cdot h$ by the piecewise linear function φ whose graph is shown in the figure at the right. Hence $(f \cdot g) \cdot h \simeq f \cdot (g \cdot h)$. Restricting attention to loops at the basepoint x_0, this says the product in $\pi_1(X, x_0)$ is associative.



Given a path $f : I \to X$, let c be the constant path at $f(1)$, defined by $c(s) = f(1)$ for all $s \in I$. Then $f \cdot c$ is a reparametrization of f via the function φ whose graph is shown in the first figure at the right, so $f \cdot c \simeq f$. Similarly, $c \cdot f \simeq f$ where c is now the constant path at $f(0)$, using the reparametrization function in the second figure. Taking f to be a loop, we deduce that the homotopy class of the constant path at x_0 is a two-sided identity in $\pi_1(X, x_0)$.



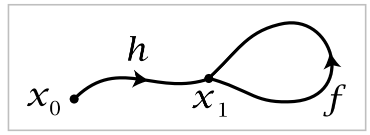
For a path f from x_0 to x_1, the **inverse path** $\overline{f}$ from x_1 back to x_0 is defined by $\overline{f}(s) = f(1-s)$. To see that $f \cdot \overline{f}$ is homotopic to a constant path we use the homotopy $h_t = f_t \cdot g_t$ where f_t is the path that equals f on the interval $[0, 1-t]$ and that is stationary at $f(1-t)$ on the interval $[1-t, 1]$, and g_t is the inverse path of f_t. We could also describe h_t in terms of the associated function $H : I \times I \to X$ using the decomposition of $I \times I$ shown in the figure. On the bottom edge of the square H is given by $f \cdot \overline{f}$ and below the 'V' we let $H(s,t)$ be independent of t, while above the 'V' we let $H(s,t)$ be independent of s. Going back to the first description of h_t, we see that since $f_0 = f$ and f_1 is the constant path c at x_0, h_t is a homotopy from $f \cdot \overline{f}$ to $c \cdot \overline{c} = c$. Replacing f by $\overline{f}$ gives $\overline{f} \cdot f \simeq c$ for c the constant path at x_1. Taking f to be a loop at the basepoint x_0, we deduce that $[\overline{f}\,]$ is a two-sided inverse for $[f]$ in $\pi_1(X, x_0)$. $\square$



Example 1.4. For a convex set X in $\mathbb{R}^n$ with basepoint $x_0 \in X$ we have $\pi_1(X, x_0) = 0$, the trivial group, since any two loops f_0 and f_1 based at x_0 are homotopic via the linear homotopy $f_t(s) = (1-t)f_0(s) + tf_1(s)$, as described in Example 1.1.

It is not so easy to show that a space has a nontrivial fundamental group since one must somehow demonstrate the nonexistence of homotopies between certain loops. We will tackle the simplest example shortly, computing the fundamental group of the circle.

It is natural to ask about the dependence of $\pi_1(X,x_0)$ on the choice of the basepoint x_0. Since $\pi_1(X,x_0)$ involves only the path-component of X containing x_0, it is clear that we can hope to find a relation between $\pi_1(X,x_0)$ and $\pi_1(X,x_1)$ for two basepoints x_0 and x_1 only if x_0 and x_1 lie in the same path-component of X. So let $h:I\to X$ be a path from x_0 to x_1, with the inverse path $\overline{h}(s)=h(1-s)$ from x_1 back to x_0. We can then associate to each loop f based at x_1 the loop $h\cdot f\cdot\overline{h}$ based at x_0.



Strictly speaking, we should choose an order of forming the product $h\cdot f\cdot\overline{h}$, either $(h\cdot f)\cdot\overline{h}$ or $h\cdot(f\cdot\overline{h})$, but the two choices are homotopic and we are only interested in homotopy classes here. Alternatively, to avoid any ambiguity we could define a general n-fold product $f_1\cdot\,\cdots\,\cdot f_n$ in which the path f_i is traversed in the time interval $[\frac{i-1}{n},\frac{i}{n}]$. Either way, we define a **change-of-basepoint** map $\beta_h:\pi_1(X,x_1)\to\pi_1(X,x_0)$ by $\beta_h[f]=[h\cdot f\cdot\overline{h}]$. This is well-defined since if f_t is a homotopy of loops based at x_1 then $h\cdot f_t\cdot\overline{h}$ is a homotopy of loops based at x_0.

Proposition 1.5. *The map $\beta_h:\pi_1(X,x_1)\to\pi_1(X,x_0)$ is an isomorphism.*

Proof: We see first that β_h is a homomorphism since $\beta_h[f\cdot g]=[h\cdot f\cdot g\cdot\overline{h}]=[h\cdot f\cdot\overline{h}\cdot h\cdot g\cdot\overline{h}]=\beta_h[f]\beta_h[g]$. Further, β_h is an isomorphism with inverse $\beta_{\overline{h}}$ since $\beta_h\beta_{\overline{h}}[f]=\beta_h[\overline{h}\cdot f\cdot h]=[h\cdot\overline{h}\cdot f\cdot h\cdot\overline{h}]=[f]$, and similarly $\beta_{\overline{h}}\beta_h[f]=[f]$. ☐

Thus if X is path-connected, the group $\pi_1(X,x_0)$ is, up to isomorphism, independent of the choice of basepoint x_0. In this case the notation $\pi_1(X,x_0)$ is often abbreviated to $\pi_1(X)$, or one could go further and write just $\pi_1 X$.

In general, a space is called **simply-connected** if it is path-connected and has trivial fundamental group. The following result explains the name.

Proposition 1.6. *A space X is simply-connected iff there is a unique homotopy class of paths connecting any two points in X.*

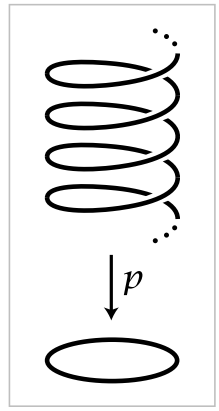
Proof: Path-connectedness is the existence of paths connecting every pair of points, so we need be concerned only with the uniqueness of connecting paths. Suppose $\pi_1(X)=0$. If f and g are two paths from x_0 to x_1, then $f\simeq f\cdot\overline{g}\cdot g\simeq g$ since the loops $\overline{g}\cdot g$ and $f\cdot\overline{g}$ are each homotopic to constant loops, using the assumption $\pi_1(X,x_0)=0$ in the latter case. Conversely, if there is only one homotopy class of paths connecting a basepoint x_0 to itself, then all loops at x_0 are homotopic to the constant loop and $\pi_1(X,x_0)=0$. ☐

The Fundamental Group of the Circle

Our first real theorem will be the calculation $\pi_1(S^1) \approx \mathbb{Z}$. Besides its intrinsic interest, this basic result will have several immediate applications of some substance, and it will be the starting point for many more calculations in the next section. It should be no surprise then that the proof will involve some genuine work.

Theorem 1.7. *$\pi_1(S^1)$ is an infinite cyclic group generated by the homotopy class of the loop $\omega(s) = (\cos 2\pi s, \sin 2\pi s)$ based at $(1,0)$.*

Note that $[\omega]^n = [\omega_n]$ where $\omega_n(s) = (\cos 2\pi ns, \sin 2\pi ns)$ for $n \in \mathbb{Z}$. The theorem is therefore equivalent to the statement that every loop in S^1 based at $(1,0)$ is homotopic to ω_n for a unique $n \in \mathbb{Z}$. To prove this the idea will be to compare paths in S^1 with paths in $\mathbb{R}$ via the map $p:\mathbb{R}\to S^1$ given by $p(s) = (\cos 2\pi s, \sin 2\pi s)$. This map can be visualized geometrically by embedding $\mathbb{R}$ in $\mathbb{R}^3$ as the helix parametrized by $s \mapsto (\cos 2\pi s, \sin 2\pi s, s)$, and then p is the restriction to the helix of the projection of $\mathbb{R}^3$ onto $\mathbb{R}^2$, $(x,y,z) \mapsto (x,y)$. Observe that the loop ω_n is the composition $p\widetilde{\omega}_n$ where $\widetilde{\omega}_n : I \to \mathbb{R}$ is the path $\widetilde{\omega}_n(s) = ns$, starting at 0 and ending at n, winding around the helix $|n|$ times, upward if $n > 0$ and downward if $n < 0$. The relation $\omega_n = p\widetilde{\omega}_n$ is expressed by saying that $\widetilde{\omega}_n$ is a **lift** of ω_n.



We will prove the theorem by studying how paths in S^1 lift to paths in $\mathbb{R}$. Most of the arguments will apply in much greater generality, and it is both more efficient and more enlightening to give them in the general context. The first step will be to define this context.

Given a space X, a **covering space** of X consists of a space $\widetilde{X}$ and a map $p:\widetilde{X}\to X$ satisfying the following condition:

$(*)$ For each point $x \in X$ there is an open neighborhood U of x in X such that $p^{-1}(U)$ is a union of disjoint open sets each of which is mapped homeomorphically onto U by p.

Such a U will be called **evenly covered.** For example, for the previously defined map $p:\mathbb{R}\to S^1$ any open arc in S^1 is evenly covered.

To prove the theorem we will need just the following two facts about covering spaces $p:\widetilde{X}\to X$.

(a) For each path $f:I\to X$ starting at a point $x_0 \in X$ and each $\widetilde{x}_0 \in p^{-1}(x_0)$ there is a unique lift $\widetilde{f}:I\to\widetilde{X}$ starting at $\widetilde{x}_0$.

(b) For each homotopy $f_t:I\to X$ of paths starting at x_0 and each $\widetilde{x}_0 \in p^{-1}(x_0)$ there is a unique lifted homotopy $\widetilde{f}_t:I\to\widetilde{X}$ of paths starting at $\widetilde{x}_0$.

Before proving these facts, let us see how they imply the theorem.

Proof of Theorem 1.7: Let $f:I\to S^1$ be a loop at the basepoint $x_0 = (1,0)$, representing a given element of $\pi_1(S^1,x_0)$. By (a) there is a lift $\widetilde{f}$ starting at 0. This path $\widetilde{f}$ ends at some integer n since $p\widetilde{f}(1) = f(1) = x_0$ and $p^{-1}(x_0) = \mathbb{Z}\subset\mathbb{R}$. Another path in $\mathbb{R}$ from 0 to n is $\widetilde{\omega}_n$, and $\widetilde{f}\simeq\widetilde{\omega}_n$ via the linear homotopy $(1-t)\widetilde{f}+t\widetilde{\omega}_n$. Composing this homotopy with p gives a homotopy $f\simeq\omega_n$ so $[f]=[\omega_n]$.

To show that n is uniquely determined by $[f]$, suppose that $f\simeq\omega_n$ and $f\simeq\omega_m$, so $\omega_m\simeq\omega_n$. Let f_t be a homotopy from $\omega_m = f_0$ to $\omega_n = f_1$. By (b) this homotopy lifts to a homotopy $\widetilde{f}_t$ of paths starting at 0. The uniqueness part of (a) implies that $\widetilde{f}_0 = \widetilde{\omega}_m$ and $\widetilde{f}_1 = \widetilde{\omega}_n$. Since $\widetilde{f}_t$ is a homotopy of paths, the endpoint $\widetilde{f}_t(1)$ is independent of t. For $t=0$ this endpoint is m and for $t=1$ it is n, so $m=n$.

It remains to prove (a) and (b). Both statements can be deduced from a more general assertion about covering spaces $p:\widetilde{X}\to X$:

(c) Given a map $F:Y\times I\to X$ and a map $\widetilde{F}:Y\times\{0\}\to\widetilde{X}$ lifting $F|Y\times\{0\}$, then there is a unique map $\widetilde{F}:Y\times I\to\widetilde{X}$ lifting F and restricting to the given $\widetilde{F}$ on $Y\times\{0\}$.

Statement (a) is the special case that Y is a point, and (b) is obtained by applying (c) with $Y = I$ in the following way. The homotopy f_t in (b) gives a map $F:I\times I\to X$ by setting $F(s,t) = f_t(s)$ as usual. A unique lift $\widetilde{F}:I\times\{0\}\to\widetilde{X}$ is obtained by an application of (a). Then (c) gives a unique lift $\widetilde{F}:I\times I\to\widetilde{X}$. The restrictions $\widetilde{F}|\{0\}\times I$ and $\widetilde{F}|\{1\}\times I$ are paths lifting constant paths, hence they must also be constant by the uniqueness part of (a). So $\widetilde{f}_t(s) = \widetilde{F}(s,t)$ is a homotopy of paths, and $\widetilde{f}_t$ lifts f_t since $p\widetilde{F} = F$.

To prove (c) we will first construct a lift $\widetilde{F}:N\times I\to\widetilde{X}$ for N some neighborhood in Y of a given point $y_0\in Y$. Since F is continuous, every point $(y_0,t)\in Y\times I$ has a product neighborhood $N_t\times(a_t,b_t)$ such that $F\big(N_t\times(a_t,b_t)\big)$ is contained in an evenly covered neighborhood of $F(y_0,t)$. By compactness of $\{y_0\}\times I$, finitely many such products $N_t\times(a_t,b_t)$ cover $\{y_0\}\times I$. This implies that we can choose a single neighborhood N of y_0 and a partition $0 = t_0 < t_1 < \cdots < t_m = 1$ of I so that for each i, $F(N\times[t_i,t_{i+1}])$ is contained in an evenly covered neighborhood U_i. Assume inductively that $\widetilde{F}$ has been constructed on $N\times[0,t_i]$, starting with the given $\widetilde{F}$ on $N\times\{0\}$. We have $F(N\times[t_i,t_{i+1}])\subset U_i$, so since U_i is evenly covered there is an open set $\widetilde{U}_i\subset\widetilde{X}$ projecting homeomorphically onto U_i by p and containing the point $\widetilde{F}(y_0,t_i)$. After replacing N by a smaller neighborhood of y_0 we may assume that $\widetilde{F}(N\times\{t_i\})$ is contained in $\widetilde{U}_i$, namely, replace $N\times\{t_i\}$ by its intersection with $(\widetilde{F}|N\times\{t_i\})^{-1}(\widetilde{U}_i)$. Now we can define $\widetilde{F}$ on $N\times[t_i,t_{i+1}]$ to be the composition of F with the homeomorphism $p^{-1}:U_i\to\widetilde{U}_i$. After a finite number of steps we eventually get a lift $\widetilde{F}:N\times I\to\widetilde{X}$ for some neighborhood N of y_0.

Next we show the uniqueness part of (c) in the special case that Y is a point. In this case we can omit Y from the notation. So suppose $\widetilde{F}$ and $\widetilde{F}'$ are two lifts of $F:I\to X$

such that $\widetilde{F}(0) = \widetilde{F}'(0)$. As before, choose a partition $0 = t_0 < t_1 < \cdots < t_m = 1$ of I so that for each i, $F([t_i, t_{i+1}])$ is contained in some evenly covered neighborhood U_i. Assume inductively that $\widetilde{F} = \widetilde{F}'$ on $[0, t_i]$. Since $[t_i, t_{i+1}]$ is connected, so is $\widetilde{F}([t_i, t_{i+1}])$, which must therefore lie in a single one of the disjoint open sets $\widetilde{U}_i$ projecting homeomorphically to U_i as in $(*)$. By the same token, $\widetilde{F}'([t_i, t_{i+1}])$ lies in a single $\widetilde{U}_i$, in fact in the same one that contains $\widetilde{F}([t_i, t_{i+1}])$ since $\widetilde{F}'(t_i) = \widetilde{F}(t_i)$. Because p is injective on $\widetilde{U}_i$ and $p\widetilde{F} = p\widetilde{F}'$, it follows that $\widetilde{F} = \widetilde{F}'$ on $[t_i, t_{i+1}]$, and the induction step is finished.

The last step in the proof of (c) is to observe that since the $\widetilde{F}$'s constructed above on sets of the form $N \times I$ are unique when restricted to each segment $\{y\} \times I$, they must agree whenever two such sets $N \times I$ overlap. So we obtain a well-defined lift $\widetilde{F}$ on all of $Y \times I$. This $\widetilde{F}$ is continuous since it is continuous on each $N \times I$. And $\widetilde{F}$ is unique since it is unique on each segment $\{y\} \times I$. □

Now we turn to some applications of the calculation of $\pi_1(S^1)$, beginning with a proof of the Fundamental Theorem of Algebra.

Theorem 1.8. *Every nonconstant polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$.*

Proof: We may assume the polynomial is of the form $p(z) = z^n + a_1 z^{n-1} + \cdots + a_n$. If $p(z)$ has no roots in $\mathbb{C}$, then for each real number $r \ge 0$ the formula

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

defines a loop in the unit circle $S^1 \subset \mathbb{C}$ based at 1. As r varies, f_r is a homotopy of loops based at 1. Since f_0 is the trivial loop, we deduce that the class $[f_r] \in \pi_1(S^1)$ is zero for all r. Now fix a large value of r, bigger than $|a_1| + \cdots + |a_n|$ and bigger than 1. Then for $|z| = r$ we have

$$|z^n| > (|a_1| + \cdots + |a_n|)|z^{n-1}| > |a_1 z^{n-1}| + \cdots + |a_n| \ge |a_1 z^{n-1} + \cdots + a_n|$$

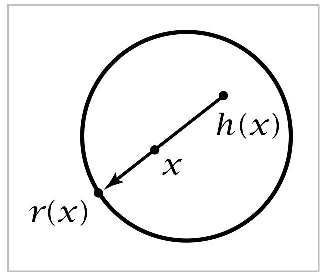
From the inequality $|z^n| > |a_1 z^{n-1} + \cdots + a_n|$ it follows that the polynomial $p_t(z) = z^n + t(a_1 z^{n-1} + \cdots + a_n)$ has no roots on the circle $|z| = r$ when $0 \le t \le 1$. Replacing p by p_t in the formula for f_r above and letting t go from 1 to 0, we obtain a homotopy from the loop f_r to the loop $\omega_n(s) = e^{2\pi ins}$. By Theorem 1.7, ω_n represents n times a generator of the infinite cyclic group $\pi_1(S^1)$. Since we have shown that $[\omega_n] = [f_r] = 0$, we conclude that $n = 0$. Thus the only polynomials without roots in $\mathbb{C}$ are constants. □

Our next application is the Brouwer fixed point theorem in dimension 2.

Theorem 1.9. *Every continuous map $h : D^2 \to D^2$ has a fixed point, that is, a point $x \in D^2$ with $h(x) = x$.*

Here we are using the standard notation D^n for the closed unit disk in $\mathbb{R}^n$, all vectors x of length $|x| \le 1$. Thus the boundary of D^n is the unit sphere S^{n-1}.

Proof: Suppose on the contrary that $h(x) \neq x$ for all $x \in D^2$. Then we can define a map $r : D^2 \to S^1$ by letting $r(x)$ be the point of S^1 where the ray in $\mathbb{R}^2$ starting at $h(x)$ and passing through x leaves D^2. Continuity of r is clear since small perturbations of x produce small perturbations of $h(x)$, hence also small perturbations of the ray through these two points.



The crucial property of r, besides continuity, is that $r(x) = x$ if $x \in S^1$. Thus r is a retraction of D^2 onto S^1. We will show that no such retraction can exist.

Let f_0 be any loop in S^1. In D^2 there is a homotopy of f_0 to a constant loop, for example the linear homotopy $f_t(s) = (1-t)f_0(s) + tx_0$ where x_0 is the basepoint of f_0. Since the retraction r is the identity on S^1, the composition rf_t is then a homotopy in S^1 from $rf_0 = f_0$ to the constant loop at x_0. But this contradicts the fact that $\pi_1(S^1)$ is nonzero. $\square$

This theorem was first proved by Brouwer around 1910, quite early in the history of topology. Brouwer in fact proved the corresponding result for D^n, and we shall obtain this generalization in Corollary 2.15 using homology groups in place of π_1. One could also use the higher homotopy group π_n. Brouwer's original proof used neither homology nor homotopy groups, which had not been invented at the time. Instead it used the notion of degree for maps $S^n \to S^n$, which we shall define in §2.2 using homology but which Brouwer defined directly in more geometric terms.

These proofs are all arguments by contradiction, and so they show just the existence of fixed points without giving any clue as to how to find one in explicit cases. Our proof of the Fundamental Theorem of Algebra was similar in this regard. There exist other proofs of the Brouwer fixed point theorem that are somewhat more constructive, for example the elegant and quite elementary proof by Sperner in 1928, which is explained very nicely in [Aigner-Ziegler 1999].

The techniques used to calculate $\pi_1(S^1)$ can be applied to prove the Borsuk–Ulam theorem in dimension two:

Theorem 1.10. *For every continuous map $f : S^2 \to \mathbb{R}^2$ there exists a pair of antipodal points x and $-x$ in S^2 with $f(x) = f(-x)$.*

It may be that there is only one such pair of antipodal points x, $-x$, for example if f is simply orthogonal projection of the standard sphere $S^2 \subset \mathbb{R}^3$ onto a plane.

The Borsuk–Ulam theorem holds more generally for maps $S^n \to \mathbb{R}^n$, as we will show in Corollary 2B.7. The proof for $n = 1$ is easy since the difference $f(x) - f(-x)$ changes sign as x goes halfway around the circle, hence this difference must be zero for some x. For $n \geq 2$ the theorem is certainly less obvious. Is it apparent, for example, that at every instant there must be a pair of antipodal points on the surface of the earth having the same temperature and the same barometric pressure?

The theorem says in particular that there is no one-to-one continuous map from S^2 to $\mathbb{R}^2$, so S^2 is not homeomorphic to a subspace of $\mathbb{R}^2$, an intuitively obvious fact that is not easy to prove directly.

Proof: If the conclusion is false for $f:S^2\to\mathbb{R}^2$, we can define a map $g:S^2\to S^1$ by $g(x) = (f(x)-f(-x))/|f(x)-f(-x)|$. Define a loop η circling the equator of $S^2\subset\mathbb{R}^3$ by $\eta(s)=(\cos 2\pi s,\sin 2\pi s,0)$, and let $h:I\to S^1$ be the composed loop $g\eta$. Since $g(-x)=-g(x)$, we have the relation $h(s+{}^1\!/_2)=-h(s)$ for all s in the interval $[0,{}^1\!/_2]$. As we showed in the calculation of $\pi_1(S^1)$, the loop h can be lifted to a path $\widetilde{h}:I\to\mathbb{R}$. The equation $h(s+{}^1\!/_2)=-h(s)$ implies that $\widetilde{h}(s+{}^1\!/_2)=\widetilde{h}(s)+{}^q\!/_2$ for some odd integer q that might conceivably depend on $s\in[0,{}^1\!/_2]$. But in fact q is independent of s since by solving the equation $\widetilde{h}(s+{}^1\!/_2)=\widetilde{h}(s)+{}^q\!/_2$ for q we see that q depends continuously on $s\in[0,{}^1\!/_2]$, so q must be a constant since it is constrained to integer values. In particular, we have $\widetilde{h}(1)=\widetilde{h}({}^1\!/_2)+{}^q\!/_2=\widetilde{h}(0)+q$. This means that h represents q times a generator of $\pi_1(S^1)$. Since q is odd, we conclude that h is not nullhomotopic. But h was the composition $g\eta:I\to S^2\to S^1$, and η is obviously nullhomotopic in S^2, so $g\eta$ is nullhomotopic in S^1 by composing a nullhomotopy of η with g. Thus we have arrived at a contradiction. $\square$

Corollary 1.11. *Whenever S^2 is expressed as the union of three closed sets A_1, A_2, and A_3, then at least one of these sets must contain a pair of antipodal points $\{x,-x\}$.*

Proof: Let $d_i:S^2\to\mathbb{R}$ measure distance to A_i, that is, $d_i(x)=\inf_{y\in A_i}|x-y|$. This is a continuous function, so we may apply the Borsuk–Ulam theorem to the map $S^2\to\mathbb{R}^2$, $x\mapsto(d_1(x),d_2(x))$, obtaining a pair of antipodal points x and $-x$ with $d_1(x)=d_1(-x)$ and $d_2(x)=d_2(-x)$. If either of these two distances is zero, then x and $-x$ both lie in the same set A_1 or A_2 since these are closed sets. On the other hand, if the distances from x and $-x$ to A_1 and A_2 are both strictly positive, then x and $-x$ lie in neither A_1 nor A_2 so they must lie in A_3. $\square$

To see that the number 'three' in this result is best possible, consider a sphere inscribed in a tetrahedron. Projecting the four faces of the tetrahedron radially onto the sphere, we obtain a cover of S^2 by four closed sets, none of which contains a pair of antipodal points.

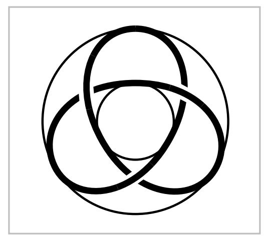
Assuming the higher-dimensional version of the Borsuk–Ulam theorem, the same arguments show that S^n cannot be covered by $n+1$ closed sets without antipodal pairs of points, though it can be covered by $n+2$ such sets, as the higher-dimensional analog of a tetrahedron shows. Even the case $n=1$ is somewhat interesting: If the circle is covered by two closed sets, one of them must contain a pair of antipodal points. This is of course false for nonclosed sets since the circle is the union of two disjoint half-open semicircles.

The relation between the fundamental group of a product space and the fundamental groups of its factors is as simple as one could wish:

Proposition 1.12. *$\pi_1(X\times Y)$ is isomorphic to $\pi_1(X)\times\pi_1(Y)$ if X and Y are path-connected.*

Proof: A basic property of the product topology is that a map $f:Z\to X\times Y$ is continuous iff the maps $g:Z\to X$ and $h:Z\to Y$ defined by $f(z)=(g(z),h(z))$ are both continuous. Hence a loop f in $X\times Y$ based at (x_0,y_0) is equivalent to a pair of loops g in X and h in Y based at x_0 and y_0 respectively. Similarly, a homotopy f_t of a loop in $X\times Y$ is equivalent to a pair of homotopies g_t and h_t of the corresponding loops in X and Y. Thus we obtain a bijection $\pi_1(X\times Y,(x_0,y_0))\approx\pi_1(X,x_0)\times\pi_1(Y,y_0)$, $[f]\mapsto([g],[h])$. This is obviously a group homomorphism, and hence an isomorphism. $\square$

Example 1.13: The Torus. By the proposition we have an isomorphism $\pi_1(S^1\times S^1)\approx\mathbb{Z}\times\mathbb{Z}$. Under this isomorphism a pair $(p,q)\in\mathbb{Z}\times\mathbb{Z}$ corresponds to a loop that winds p times around one S^1 factor of the torus and q times around the other S^1 factor, for example the loop $\omega_{pq}(s)=(\omega_p(s),\omega_q(s))$. Interestingly, this loop can be knotted, as the figure shows for the case $p=3$, $q=2$. The knots that arise in this fashion, the so-called *torus knots*, are studied in Example 1.24.



More generally, the n-dimensional torus, which is the product of n circles, has fundamental group isomorphic to the product of n copies of $\mathbb{Z}$. This follows by induction on n.

Induced Homomorphisms

Suppose $\varphi:X\to Y$ is a map taking the basepoint $x_0\in X$ to the basepoint $y_0\in Y$. For brevity we write $\varphi:(X,x_0)\to(Y,y_0)$ in this situation. Then φ induces a homomorphism $\varphi_*:\pi_1(X,x_0)\to\pi_1(Y,y_0)$, defined by composing loops $f:I\to X$ based at x_0 with φ, that is, $\varphi_*[f]=[\varphi f]$. This induced map φ_* is well-defined since a homotopy f_t of loops based at x_0 yields a composed homotopy φf_t of loops based at y_0, so $\varphi_*[f_0]=[\varphi f_0]=[\varphi f_1]=\varphi_*[f_1]$. Furthermore, φ_* is a homomorphism since $\varphi(f\cdot g)=(\varphi f)\cdot(\varphi g)$, both functions having the value $\varphi f(2s)$ for $0\le s\le 1/2$ and the value $\varphi g(2s-1)$ for $1/2\le s\le 1$.

Two basic properties of induced homomorphisms are:

- $(\varphi\psi)_*=\varphi_*\psi_*$ for a composition $(X,x_0)\xrightarrow{\psi}(Y,y_0)\xrightarrow{\varphi}(Z,z_0)$.
- $1\!\!1_*=1\!\!1$, which is a concise way of saying that the identity map $1\!\!1:X\to X$ induces the identity map $1\!\!1:\pi_1(X,x_0)\to\pi_1(X,x_0)$.

The first of these follows from the fact that composition of maps is associative, so $(\varphi\psi)f=\varphi(\psi f)$, and the second is obvious. These two properties of induced homomorphisms are what makes the fundamental group a functor. The formal definition of a functor requires the introduction of certain other preliminary concepts, however, so we postpone this until it is needed in §2.3.

As an application we can deduce easily that if φ is a homeomorphism with inverse ψ then φ_* is an isomorphism with inverse ψ_* since $\varphi_*\psi_* = (\varphi\psi)_* = \mathbb{1}_* = \mathbb{1}$ and similarly $\psi_*\varphi_* = \mathbb{1}$. We will use this fact in the following calculation of the fundamental groups of higher-dimensional spheres:

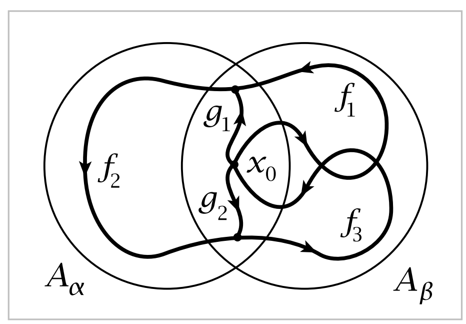
Proposition 1.14. *$\pi_1(S^n) = 0$ if $n \geq 2$.*

The main step in the proof will be a general fact that will also play a key role in the next section:

Lemma 1.15. *If a space X is the union of a collection of path-connected open sets A_α each containing the basepoint $x_0 \in X$ and if each intersection $A_\alpha \cap A_\beta$ is path-connected, then every loop in X at x_0 is homotopic to a product of loops each of which is contained in a single A_α.*

Proof: Given a loop $f: I \to X$ at the basepoint x_0, we claim there is a partition $0 = s_0 < s_1 < \cdots < s_m = 1$ of I such that each subinterval $[s_{i-1}, s_i]$ is mapped by f to a single A_α. Namely, since f is continuous, each $s \in I$ has an open neighborhood V_s in I mapped by f to some A_α. We may in fact take V_s to be an interval whose closure is mapped to a single A_α. Compactness of I implies that a finite number of these intervals cover I. The endpoints of this finite set of intervals then define the desired partition of I.

Denote the A_α containing $f([s_{i-1}, s_i])$ by A_i, and let f_i be the path obtained by restricting f to $[s_{i-1}, s_i]$. Then f is the composition $f_1 \cdot \cdots \cdot f_m$ with f_i a path in A_i. Since we assume $A_i \cap A_{i+1}$ is path-connected, we may choose a path g_i in $A_i \cap A_{i+1}$ from x_0 to the point $f(s_i) \in A_i \cap A_{i+1}$. Consider the loop



$$(f_1 \cdot \bar{g}_1) \cdot (g_1 \cdot f_2 \cdot \bar{g}_2) \cdot (g_2 \cdot f_3 \cdot \bar{g}_3) \cdot \cdots \cdot (g_{m-1} \cdot f_m)$$

which is homotopic to f. This loop is a composition of loops each lying in a single A_i, the loops indicated by the parentheses. $\square$

Proof of Proposition 1.14: We can express S^n as the union of two open sets A_1 and A_2 each homeomorphic to $\mathbb{R}^n$ such that $A_1 \cap A_2$ is homeomorphic to $S^{n-1} \times \mathbb{R}$, for example by taking A_1 and A_2 to be the complements of two antipodal points in S^n. Choose a basepoint x_0 in $A_1 \cap A_2$. If $n \geq 2$ then $A_1 \cap A_2$ is path-connected. The lemma then applies to say that every loop in S^n based at x_0 is homotopic to a product of loops in A_1 or A_2. Both $\pi_1(A_1)$ and $\pi_1(A_2)$ are zero since A_1 and A_2 are homeomorphic to $\mathbb{R}^n$. Hence every loop in S^n is nullhomotopic. $\square$

Corollary 1.16. *$\mathbb{R}^2$ is not homeomorphic to $\mathbb{R}^n$ for $n \neq 2$.*

Proof: Suppose $f:\mathbb{R}^2 \to \mathbb{R}^n$ is a homeomorphism. The case $n = 1$ is easily disposed of since $\mathbb{R}^2 - \{0\}$ is path-connected but the homeomorphic space $\mathbb{R}^n - \{f(0)\}$ is not path-connected when $n = 1$. When $n > 2$ we cannot distinguish $\mathbb{R}^2 - \{0\}$ from $\mathbb{R}^n - \{f(0)\}$ by the number of path-components, but we can distinguish them by their fundamental groups. Namely, for a point x in $\mathbb{R}^n$, the complement $\mathbb{R}^n - \{x\}$ is homeomorphic to $S^{n-1} \times \mathbb{R}$, so Proposition 1.12 implies that $\pi_1(\mathbb{R}^n - \{x\})$ is isomorphic to $\pi_1(S^{n-1}) \times \pi_1(\mathbb{R}) \approx \pi_1(S^{n-1})$. Hence $\pi_1(\mathbb{R}^n - \{x\})$ is $\mathbb{Z}$ for $n = 2$ and trivial for $n > 2$, using Proposition 1.14 in the latter case. $\square$

The more general statement that $\mathbb{R}^m$ is not homeomorphic to $\mathbb{R}^n$ if $m \neq n$ can be proved in the same way using either the higher homotopy groups or homology groups. In fact, nonempty open sets in $\mathbb{R}^m$ and $\mathbb{R}^n$ can be homeomorphic only if $m = n$, as we will show in Theorem 2.26 using homology.

Induced homomorphisms allow relations between spaces to be transformed into relations between their fundamental groups. Here is an illustration of this principle:

Proposition 1.17. *If a space X retracts onto a subspace A, then the homomorphism $i_*:\pi_1(A,x_0) \to \pi_1(X,x_0)$ induced by the inclusion $i:A \hookrightarrow X$ is injective. If A is a deformation retract of X, then i_* is an isomorphism.*

Proof: If $r:X \to A$ is a retraction, then $ri = \mathbb{1}$, hence $r_* i_* = \mathbb{1}$, which implies that i_* is injective. If $r_t:X \to X$ is a deformation retraction of X onto A, so $r_0 = \mathbb{1}$, $r_t|A = \mathbb{1}$, and $r_1(X) \subset A$, then for any loop $f:I \to X$ based at $x_0 \in A$ the composition $r_t f$ gives a homotopy of f to a loop in A, so i_* is also surjective. $\square$

This gives another way of seeing that S^1 is not a retract of D^2, a fact we showed earlier in the proof of the Brouwer fixed point theorem, since the inclusion-induced map $\pi_1(S^1) \to \pi_1(D^2)$ is a homomorphism $\mathbb{Z} \to 0$ that cannot be injective.

The exact group-theoretic analog of a retraction is a homomorphism ρ of a group G onto a subgroup H such that ρ restricts to the identity on H. In the notation above, if we identify $\pi_1(A)$ with its image under i_*, then r_* is such a homomorphism from $\pi_1(X)$ onto the subgroup $\pi_1(A)$. The existence of a retracting homomorphism $\rho:G \to H$ is quite a strong condition on H. If H is a normal subgroup, it implies that G is the direct product of H and the kernel of ρ. If H is not normal, then G is what is called in group theory the semi-direct product of H and the kernel of ρ.

Recall from Chapter 0 the general definition of a homotopy as a family $\varphi_t:X \to Y$, $t \in I$, such that the associated map $\Phi:X \times I \to Y$, $\Phi(x,t) = \varphi_t(x)$, is continuous. If φ_t takes a subspace $A \subset X$ to a subspace $B \subset Y$ for all t, then we speak of a homotopy of maps of pairs, $\varphi_t:(X,A) \to (Y,B)$. In particular, a **basepoint-preserving homotopy**

$\varphi_t : (X, x_0) \to (Y, y_0)$ is the case that $\varphi_t(x_0) = y_0$ for all t. Another basic property of induced homomorphisms is their invariance under such homotopies:

- If $\varphi_t : (X, x_0) \to (Y, y_0)$ is a basepoint-preserving homotopy, then $\varphi_{0*} = \varphi_{1*}$.

This holds since $\varphi_{0*}[f] = [\varphi_0 f] = [\varphi_1 f] = \varphi_{1*}[f]$, the middle equality coming from the homotopy $\varphi_t f$.

There is a notion of homotopy equivalence for spaces with basepoints. One says $(X, x_0) \simeq (Y, y_0)$ if there are maps $\varphi : (X, x_0) \to (Y, y_0)$ and $\psi : (Y, y_0) \to (X, x_0)$ with homotopies $\varphi\psi \simeq \mathbb{1}$ and $\psi\varphi \simeq \mathbb{1}$ through maps fixing the basepoints. In this case the induced maps on π_1 satisfy $\varphi_*\psi_* = (\varphi\psi)_* = \mathbb{1}_* = \mathbb{1}$ and likewise $\psi_*\varphi_* = \mathbb{1}$, so φ_* and ψ_* are inverse isomorphisms $\pi_1(X, x_0) \approx \pi_1(Y, y_0)$. This somewhat formal argument gives another proof that a deformation retraction induces an isomorphism on fundamental groups, since if X deformation retracts onto A then $(X, x_0) \simeq (A, x_0)$ for any choice of basepoint $x_0 \in A$.

Having to pay so much attention to basepoints when dealing with the fundamental group is something of a nuisance. For homotopy equivalences one does not have to be quite so careful, as the conditions on basepoints can actually be dropped:

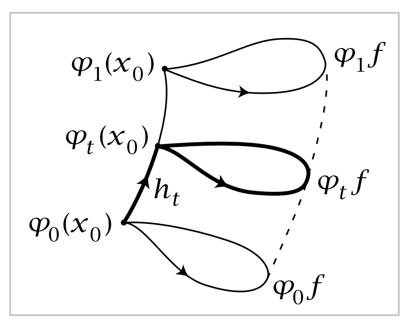
Proposition 1.18. *If $\varphi : X \to Y$ is a homotopy equivalence, then the induced homomorphism $\varphi_* : \pi_1(X, x_0) \to \pi_1(Y, \varphi(x_0))$ is an isomorphism for all $x_0 \in X$.*

The proof will use a simple fact about homotopies that do not fix the basepoint:

Lemma 1.19. *If $\varphi_t : X \to Y$ is a homotopy and h is the path $\varphi_t(x_0)$ formed by the images of a basepoint $x_0 \in X$, then the three maps in the diagram at the right satisfy $\varphi_{0*} = \beta_h \varphi_{1*}$.*

$$\begin{tikzcd} & \pi_1(Y, \varphi_1(x_0)) \arrow[dd, "\beta_h"] \\ \pi_1(X, x_0) \arrow[ur, "\varphi_{1*}"] \arrow[dr, "\varphi_{0*}"'] & \\ & \pi_1(Y, \varphi_0(x_0)) \end{tikzcd}$$

Proof: Let h_t be the restriction of h to the interval $[0, t]$, with a reparametrization so that the domain of h_t is still $[0, 1]$. Explicitly, we can take $h_t(s) = h(ts)$. Then if f is a loop in X at the basepoint x_0, the product $h_t \cdot (\varphi_t f) \cdot \overline{h}_t$ gives a homotopy of loops at $\varphi_0(x_0)$. Restricting this homotopy to $t = 0$ and $t = 1$, we see that $\varphi_{0*}([f]) = \beta_h(\varphi_{1*}([f]))$. $\square$



Proof of 1.18: Let $\psi : Y \to X$ be a homotopy-inverse for φ, so that $\varphi\psi \simeq \mathbb{1}$ and $\psi\varphi \simeq \mathbb{1}$. Consider the maps

$$\pi_1(X, x_0) \xrightarrow{\varphi_*} \pi_1(Y, \varphi(x_0)) \xrightarrow{\psi_*} \pi_1(X, \psi\varphi(x_0)) \xrightarrow{\varphi_*} \pi_1(Y, \varphi\psi\varphi(x_0))$$

The composition of the first two maps is an isomorphism since $\psi\varphi \simeq \mathbb{1}$ implies that $\psi_*\varphi_* = \beta_h$ for some h, by the lemma. In particular, since $\psi_*\varphi_*$ is an isomorphism,

φ_* is injective. The same reasoning with the second and third maps shows that ψ_* is injective. Thus the first two of the three maps are injections and their composition is an isomorphism, so the first map φ_* must be surjective as well as injective. $\square$

Exercises

1. Show that composition of paths satisfies the following cancellation property: If $f_0 \cdot g_0 \simeq f_1 \cdot g_1$ and $g_0 \simeq g_1$ then $f_0 \simeq f_1$.

2. Show that the change-of-basepoint homomorphism β_h depends only on the homotopy class of h.

3. For a path-connected space X, show that $\pi_1(X)$ is abelian iff all basepoint-change homomorphisms β_h depend only on the endpoints of the path h.

4. A subspace $X \subset \mathbb{R}^n$ is said to be *star-shaped* if there is a point $x_0 \in X$ such that, for each $x \in X$, the line segment from x_0 to x lies in X. Show that if a subspace $X \subset \mathbb{R}^n$ is locally star-shaped, in the sense that every point of X has a star-shaped neighborhood in X, then every path in X is homotopic in X to a piecewise linear path, that is, a path consisting of a finite number of straight line segments traversed at constant speed. Show this applies in particular when X is open or when X is a union of finitely many closed convex sets.

5. Show that for a space X, the following three conditions are equivalent:
(a) Every map $S^1 \to X$ is homotopic to a constant map, with image a point.
(b) Every map $S^1 \to X$ extends to a map $D^2 \to X$.
(c) $\pi_1(X, x_0) = 0$ for all $x_0 \in X$.
Deduce that a space X is simply-connected iff all maps $S^1 \to X$ are homotopic. [In this problem, 'homotopic' means 'homotopic without regard to basepoints'.]

6. We can regard $\pi_1(X, x_0)$ as the set of basepoint-preserving homotopy classes of maps $(S^1, s_0) \to (X, x_0)$. Let $[S^1, X]$ be the set of homotopy classes of maps $S^1 \to X$, with no conditions on basepoints. Thus there is a natural map $\Phi : \pi_1(X, x_0) \to [S^1, X]$ obtained by ignoring basepoints. Show that Φ is onto if X is path-connected, and that $\Phi([f]) = \Phi([g])$ iff $[f]$ and $[g]$ are conjugate in $\pi_1(X, x_0)$. Hence Φ induces a one-to-one correspondence between $[S^1, X]$ and the set of conjugacy classes in $\pi_1(X)$, when X is path-connected.

7. Define $f : S^1 \times I \to S^1 \times I$ by $f(\theta, s) = (\theta + 2\pi s, s)$, so f restricts to the identity on the two boundary circles of $S^1 \times I$. Show that f is homotopic to the identity by a homotopy f_t that is stationary on one of the boundary circles, but not by any homotopy f_t that is stationary on both boundary circles. [Consider what f does to the path $s \mapsto (\theta_0, s)$ for fixed $\theta_0 \in S^1$.]

8. Does the Borsuk–Ulam theorem hold for the torus? In other words, for every map $f : S^1 \times S^1 \to \mathbb{R}^2$ must there exist $(x, y) \in S^1 \times S^1$ such that $f(x, y) = f(-x, -y)$?

9. Let A_1, A_2, A_3 be compact sets in $\mathbb{R}^3$. Use the Borsuk–Ulam theorem to show that there is one plane $P \subset \mathbb{R}^3$ that simultaneously divides each A_i into two pieces of equal measure.

10. From the isomorphism $\pi_1(X\times Y,(x_0,y_0)) \approx \pi_1(X,x_0)\times\pi_1(Y,y_0)$ it follows that loops in $X\times\{y_0\}$ and $\{x_0\}\times Y$ represent commuting elements of $\pi_1(X\times Y,(x_0,y_0))$. Construct an explicit homotopy demonstrating this.

11. If X_0 is the path-component of a space X containing the basepoint x_0, show that the inclusion $X_0 \hookrightarrow X$ induces an isomorphism $\pi_1(X_0,x_0)\to\pi_1(X,x_0)$.

12. Show that every homomorphism $\pi_1(S^1)\to\pi_1(S^1)$ can be realized as the induced homomorphism φ_* of a map $\varphi : S^1\to S^1$.

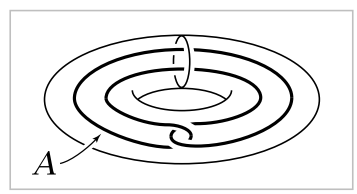
13. Given a space X and a path-connected subspace A containing the basepoint x_0, show that the map $\pi_1(A,x_0)\to\pi_1(X,x_0)$ induced by the inclusion $A\hookrightarrow X$ is surjective iff every path in X with endpoints in A is homotopic to a path in A.

14. Show that the isomorphism $\pi_1(X\times Y) \approx \pi_1(X)\times\pi_1(Y)$ in Proposition 1.12 is given by $[f] \mapsto (p_{1*}([f]), p_{2*}([f]))$ where p_1 and p_2 are the projections of $X\times Y$ onto its two factors.

15. Given a map $f : X\to Y$ and a path $h : I\to X$ from x_0 to x_1, show that $f_*\beta_h = \beta_{fh}f_*$ in the diagram at the right.

$$\begin{tikzcd}
\pi_1(X,x_1) \arrow[r, "\beta_h"] \arrow[d, "f_*"'] & \pi_1(X,x_0) \arrow[d, "f_*"] \\
\pi_1(Y,f(x_1)) \arrow[r, "\beta_{fh}"] & \pi_1(Y,f(x_0))
\end{tikzcd}$$

16. Show that there are no retractions $r : X\to A$ in the following cases:



(a) $X = \mathbb{R}^3$ with A any subspace homeomorphic to S^1.
(b) $X = S^1\times D^2$ with A its boundary torus $S^1\times S^1$.
(c) $X = S^1\times D^2$ and A the circle shown in the figure.
(d) $X = D^2 \vee D^2$ with A its boundary $S^1\vee S^1$.
(e) X a disk with two points on its boundary identified and A its boundary $S^1\vee S^1$.
(f) X the Möbius band and A its boundary circle.

17. Construct infinitely many nonhomotopic retractions $S^1\vee S^1\to S^1$.

18. Using Lemma 1.15, show that if a space X is obtained from a path-connected subspace A by attaching a cell e^n with $n\ge 2$, then the inclusion $A\hookrightarrow X$ induces a surjection on π_1. Apply this to show:
(a) The wedge sum $S^1\vee S^2$ has fundamental group $\mathbb{Z}$.
(b) For a path-connected CW complex X the inclusion map $X^1\hookrightarrow X$ of its 1-skeleton induces a surjection $\pi_1(X^1)\to\pi_1(X)$. [For the case that X has infinitely many cells, see Proposition A.1 in the Appendix.]

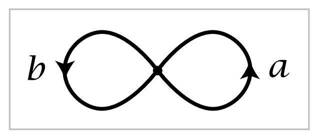
19. Show that if X is a path-connected 1-dimensional CW complex with basepoint x_0 a 0-cell, then every loop in X is homotopic to a loop consisting of a finite sequence of edges traversed monotonically. [See the proof of Lemma 1.15. This exercise gives an elementary proof that $\pi_1(S^1)$ is cyclic generated by the standard loop winding once around the circle. The more difficult part of the calculation of $\pi_1(S^1)$ is therefore the fact that no iterate of this loop is nullhomotopic.]

20. Suppose $f_t: X \to X$ is a homotopy such that f_0 and f_1 are each the identity map. Use Lemma 1.19 to show that for any $x_0 \in X$, the loop $f_t(x_0)$ represents an element of the center of $\pi_1(X, x_0)$. [One can interpret the result as saying that a loop represents an element of the center of $\pi_1(X)$ if it extends to a loop of maps $X \to X$.]

1.2 Van Kampen's Theorem

The van Kampen theorem gives a method for computing the fundamental groups of spaces that can be decomposed into simpler spaces whose fundamental groups are already known. By systematic use of this theorem one can compute the fundamental groups of a very large number of spaces. We shall see for example that for every group G there is a space X_G whose fundamental group is isomorphic to G.

To give some idea of how one might hope to compute fundamental groups by decomposing spaces into simpler pieces, let us look at an example. Consider the space X formed by two circles A and B intersecting in a single point, which we choose as the basepoint x_0. By our preceding calculations we know that $\pi_1(A)$ is infinite cyclic, generated by a loop a that goes once around A. Similarly, $\pi_1(B)$ is a copy of $\mathbb{Z}$ generated by a loop b going once around B. Each product of powers of a and b then gives an element of $\pi_1(X)$. For example, the product $a^5b^2a^{-3}ba^2$ is the loop that goes five times around A, then twice around B, then three times around A in the opposite direction, then once around B, then twice around A. The set of all words like this consisting of powers of a alternating with powers of b forms a group usually denoted $\mathbb{Z} * \mathbb{Z}$. Multiplication in this group is defined just as one would expect, for example $(b^4a^5b^2a^{-3})(a^4b^{-1}ab^3) = b^4a^5b^2ab^{-1}ab^3$. The identity element is the empty word, and inverses are what they have to be, for example $(ab^2a^{-3}b^{-4})^{-1} = b^4a^3b^{-2}a^{-1}$. It would be very nice if such words in a and b corresponded exactly to elements of $\pi_1(X)$, so that $\pi_1(X)$ was isomorphic to the group $\mathbb{Z} * \mathbb{Z}$. The van Kampen theorem will imply that this is indeed the case.



Similarly, if X is the union of three circles touching at a single point, the van Kampen theorem will imply that $\pi_1(X)$ is $\mathbb{Z} * \mathbb{Z} * \mathbb{Z}$, the group consisting of words in powers of three letters a, b, c. The generalization to a union of any number of circles touching at one point will also follow.

The group $\mathbb{Z} * \mathbb{Z}$ is an example of a general construction called the *free product* of groups. The statement of van Kampen's theorem will be in terms of free products, so before stating the theorem we will make an algebraic digression to describe the construction of free products in some detail.

Free Products of Groups

Suppose one is given a collection of groups G_α and one wishes to construct a single group containing all these groups as subgroups. One way to do this would be to take the product group $\prod_\alpha G_\alpha$, whose elements can be regarded as the functions $\alpha \mapsto g_\alpha \in G_\alpha$. Or one could restrict to functions taking on nonidentity values at most finitely often, forming the direct sum $\bigoplus_\alpha G_\alpha$. Both these constructions produce groups containing all the G_α's as subgroups, but with the property that elements of different subgroups G_α commute with each other. In the realm of nonabelian groups this commutativity is unnatural, and so one would like a 'nonabelian' version of $\prod_\alpha G_\alpha$ or $\bigoplus_\alpha G_\alpha$. Since the sum $\bigoplus_\alpha G_\alpha$ is smaller and presumably simpler than $\prod_\alpha G_\alpha$, it should be easier to construct a nonabelian version of $\bigoplus_\alpha G_\alpha$, and this is what the free product $*_\alpha G_\alpha$ achieves.

Here is the precise definition. As a set, the free product $*_\alpha G_\alpha$ consists of all words $g_1 g_2 \cdots g_m$ of arbitrary finite length $m \ge 0$, where each letter g_i belongs to a group G_{α_i} and is not the identity element of G_{α_i}, and adjacent letters g_i and g_{i+1} belong to different groups G_α, that is, $\alpha_i \neq \alpha_{i+1}$. Words satisfying these conditions are called *reduced*, the idea being that unreduced words can always be simplified to reduced words by writing adjacent letters that lie in the same G_{α_i} as a single letter and by canceling trivial letters. The empty word is allowed, and will be the identity element of $*_\alpha G_\alpha$. The group operation in $*_\alpha G_\alpha$ is juxtaposition, $(g_1 \cdots g_m)(h_1 \cdots h_n) = g_1 \cdots g_m h_1 \cdots h_n$. This product may not be reduced, however: If g_m and h_1 belong to the same G_α, they should be combined into a single letter $(g_m h_1)$ according to the multiplication in G_α, and if this new letter $g_m h_1$ happens to be the identity of G_α, it should be canceled from the product. This may allow g_{m-1} and h_2 to be combined, and possibly canceled too. Repetition of this process eventually produces a reduced word. For example, in the product $(g_1 \cdots g_m)(g_m^{-1} \cdots g_1^{-1})$ everything cancels and we get the identity element of $*_\alpha G_\alpha$, the empty word.

Verifying directly that this multiplication is associative would be rather tedious, but there is an indirect approach that avoids most of the work. Let W be the set of reduced words $g_1 \cdots g_m$ as above, including the empty word. To each $g \in G_\alpha$ we associate the function $L_g : W \to W$ given by multiplication on the left, $L_g(g_1 \cdots g_m) = g g_1 \cdots g_m$ where we combine g with g_1 if $g_1 \in G_\alpha$ to make $g g_1 \cdots g_m$ a reduced word. A key property of the association $g \mapsto L_g$ is the formula $L_{gg'} = L_g L_{g'}$ for $g, g' \in G_\alpha$, that is, $g(g'(g_1 \cdots g_m)) = (gg')(g_1 \cdots g_m)$. This special case of associativity follows rather trivially from associativity in G_α. The formula $L_{gg'} = L_g L_{g'}$ implies that L_g is invertible with inverse $L_{g^{-1}}$. Therefore the association $g \mapsto L_g$ defines a homomorphism from G_α to the group $P(W)$ of all permutations of W. More generally, we can define $L : W \to P(W)$ by $L(g_1 \cdots g_m) = L_{g_1} \cdots L_{g_m}$ for each reduced word $g_1 \cdots g_m$. This function L is injective since the permutation $L(g_1 \cdots g_m)$ sends the empty word to $g_1 \cdots g_m$. The product operation in W corresponds under L to composition in $P(W)$, because of the relation $L_{gg'} = L_g L_{g'}$. Since composition in $P(W)$ is associative, we conclude that the product in W is associative.

In particular, we have the free product $\mathbb{Z} * \mathbb{Z}$ as described earlier. This is an example of a *free group*, the free product of any number of copies of $\mathbb{Z}$, finite or infinite. The elements of a free group are uniquely representable as reduced words in powers of generators for the various copies of $\mathbb{Z}$, with one generator for each $\mathbb{Z}$, just as in the case of $\mathbb{Z} * \mathbb{Z}$. These generators are called a *basis* for the free group, and the number of basis elements is the *rank* of the free group. The abelianization of a free group is a free abelian group with basis the same set of generators, so since the rank of a free abelian group is well-defined, independent of the choice of basis, the same is true for the rank of a free group.

An interesting example of a free product that is not a free group is $\mathbb{Z}_2 * \mathbb{Z}_2$. This is like $\mathbb{Z} * \mathbb{Z}$ but simpler since $a^2 = e = b^2$, so powers of a and b are not needed, and $\mathbb{Z}_2 * \mathbb{Z}_2$ consists of just the alternating words in a and b: a, b, ab, ba, aba, bab, $abab$, $baba$, $ababa$, $\cdots$, together with the empty word. The structure of $\mathbb{Z}_2 * \mathbb{Z}_2$ can be elucidated by looking at the homomorphism $\varphi : \mathbb{Z}_2 * \mathbb{Z}_2 \to \mathbb{Z}_2$ associating to each word its length mod 2. Obviously φ is surjective, and its kernel consists of the words of even length. These form an infinite cyclic subgroup generated by ab since $ba = (ab)^{-1}$ in $\mathbb{Z}_2 * \mathbb{Z}_2$. In fact, $\mathbb{Z}_2 * \mathbb{Z}_2$ is the semi-direct product of the subgroups $\mathbb{Z}$ and $\mathbb{Z}_2$ generated by ab and a, with the conjugation relation $a(ab)a^{-1} = (ab)^{-1}$. This group is sometimes called the infinite dihedral group.

For a general free product $*_\alpha G_\alpha$, each group G_α is naturally identified with a subgroup of $*_\alpha G_\alpha$, the subgroup consisting of the empty word and the nonidentity one-letter words $g \in G_\alpha$. From this viewpoint the empty word is the common identity element of all the subgroups G_α, which are otherwise disjoint. A consequence of associativity is that any product $g_1 \cdots g_m$ of elements g_i in the groups G_α has a unique reduced form, the element of $*_\alpha G_\alpha$ obtained by performing the multiplications in any order. Any sequence of reduction operations on an unreduced product $g_1 \cdots g_m$, combining adjacent letters g_i and g_{i+1} that lie in the same G_α or canceling a g_i that is the identity, can be viewed as a way of inserting parentheses into $g_1 \cdots g_m$ and performing the resulting sequence of multiplications. Thus associativity implies that any two sequences of reduction operations performed on the same unreduced word always yield the same reduced word.

A basic property of the free product $*_\alpha G_\alpha$ is that any collection of homomorphisms $\varphi_\alpha : G_\alpha \to H$ extends uniquely to a homomorphism $\varphi : *_\alpha G_\alpha \to H$. Namely, the value of φ on a word $g_1 \cdots g_n$ with $g_i \in G_{\alpha_i}$ must be $\varphi_{\alpha_1}(g_1) \cdots \varphi_{\alpha_n}(g_n)$, and using this formula to define φ gives a well-defined homomorphism since the process of reducing an unreduced product in $*_\alpha G_\alpha$ does not affect its image under φ. For example, for a free product $G * H$ the inclusions $G \hookrightarrow G \times H$ and $H \hookrightarrow G \times H$ induce a surjective homomorphism $G * H \to G \times H$.

The van Kampen Theorem

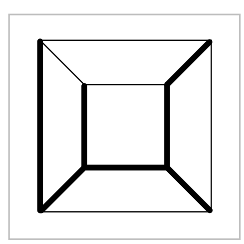
Suppose a space X is decomposed as the union of a collection of path-connected open subsets A_α, each of which contains the basepoint $x_0 \in X$. By the remarks in the preceding paragraph, the homomorphisms $j_\alpha : \pi_1(A_\alpha) \to \pi_1(X)$ induced by the inclusions $A_\alpha \hookrightarrow X$ extend to a homomorphism $\Phi : *_\alpha \pi_1(A_\alpha) \to \pi_1(X)$. The van Kampen theorem will say that Φ is very often surjective, but we can expect Φ to have a nontrivial kernel in general. For if $i_{\alpha\beta} : \pi_1(A_\alpha \cap A_\beta) \to \pi_1(A_\alpha)$ is the homomorphism induced by the inclusion $A_\alpha \cap A_\beta \hookrightarrow A_\alpha$ then $j_\alpha i_{\alpha\beta} = j_\beta i_{\beta\alpha}$, both these compositions being induced by the inclusion $A_\alpha \cap A_\beta \hookrightarrow X$, so the kernel of Φ contains all the elements of the form $i_{\alpha\beta}(\omega) i_{\beta\alpha}(\omega)^{-1}$ for $\omega \in \pi_1(A_\alpha \cap A_\beta)$. Van Kampen's theorem asserts that under fairly broad hypotheses this gives a full description of Φ:

Theorem 1.20. *If X is the union of path-connected open sets A_α each containing the basepoint $x_0 \in X$ and if each intersection $A_\alpha \cap A_\beta$ is path-connected, then the homomorphism $\Phi : *_\alpha \pi_1(A_\alpha) \to \pi_1(X)$ is surjective. If in addition each intersection $A_\alpha \cap A_\beta \cap A_\gamma$ is path-connected, then the kernel of Φ is the normal subgroup N generated by all elements of the form $i_{\alpha\beta}(\omega) i_{\beta\alpha}(\omega)^{-1}$ for $\omega \in \pi_1(A_\alpha \cap A_\beta)$, and hence Φ induces an isomorphism $\pi_1(X) \approx *_\alpha \pi_1(A_\alpha)/N$.*

Example 1.21: Wedge Sums. In Chapter 0 we defined the wedge sum $\bigvee_\alpha X_\alpha$ of a collection of spaces X_α with basepoints $x_\alpha \in X_\alpha$ to be the quotient space of the disjoint union $\coprod_\alpha X_\alpha$ in which all the basepoints x_α are identified to a single point. If each x_α is a deformation retract of an open neighborhood U_α in X_α, then X_α is a deformation retract of its open neighborhood $A_\alpha = X_\alpha \bigvee_{\beta \neq \alpha} U_\beta$. The intersection of two or more distinct A_α's is $\bigvee_\alpha U_\alpha$, which deformation retracts to a point. Van Kampen's theorem then implies that $\Phi : *_\alpha \pi_1(X_\alpha) \to \pi_1(\bigvee_\alpha X_\alpha)$ is an isomorphism, assuming that each X_α is path-connected, hence also each A_α.

Thus for a wedge sum $\bigvee_\alpha S^1_\alpha$ of circles, $\pi_1(\bigvee_\alpha S^1_\alpha)$ is a free group, the free product of copies of $\mathbb{Z}$, one for each circle S^1_α. In particular, $\pi_1(S^1 \vee S^1)$ is the free group $\mathbb{Z} * \mathbb{Z}$, as in the example at the beginning of this section.

It is true more generally that the fundamental group of any connected graph is free, as we show in §1.A. Here is an example illustrating the general technique.

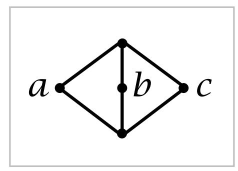


Example 1.22. Let X be the graph shown in the figure, consisting of the twelve edges of a cube. The seven heavily shaded edges form a maximal tree $T \subset X$, a contractible subgraph containing all the vertices of X. We claim that $\pi_1(X)$ is the free product of five copies of $\mathbb{Z}$, one for each edge not in T. To deduce this from van Kampen's theorem, choose for each edge e_α of $X - T$ an open neighborhood A_α of $T \cup e_\alpha$ in X that deformation retracts onto $T \cup e_\alpha$. The intersection of two or more A_α's deformation retracts onto T, hence is contractible. The A_α's form a cover of

X satisfying the hypotheses of van Kampen's theorem, and since the intersection of any two of them is simply-connected we obtain an isomorphism $\pi_1(X) \approx *_\alpha \pi_1(A_\alpha)$. Each A_α deformation retracts onto a circle, so $\pi_1(X)$ is free on five generators, as claimed. As explicit generators we can choose for each edge e_α of $X - T$ a loop f_α that starts at a basepoint in T, travels in T to one end of e_α, then across e_α, then back to the basepoint along a path in T.

Van Kampen's theorem is often applied when there are just two sets A_α and A_β in the cover of X, so the condition on triple intersections $A_\alpha \cap A_\beta \cap A_\gamma$ is superfluous and one obtains an isomorphism $\pi_1(X) \approx (\pi_1(A_\alpha) * \pi_1(A_\beta))/N$, under the assumption that $A_\alpha \cap A_\beta$ is path-connected. The proof in this special case is virtually identical with the proof in the general case, however.

One can see that the intersections $A_\alpha \cap A_\beta$ need to be path-connected by considering the example of S^1 decomposed as the union of two open arcs. In this case Φ is not surjective. For an example showing that triple intersections $A_\alpha \cap A_\beta \cap A_\gamma$ need to be path-connected, let X be the suspension of three points a, b, c, and let A_α, A_β, and A_γ be the complements of these three points. The theorem does apply to the covering $\{A_\alpha, A_\beta\}$, so there are isomorphisms $\pi_1(X) \approx \pi_1(A_\alpha) * \pi_1(A_\beta) \approx \mathbb{Z} * \mathbb{Z}$ since $A_\alpha \cap A_\beta$ is contractible. If we tried to use the covering $\{A_\alpha, A_\beta, A_\gamma\}$, which has each of the twofold intersections path-connected but not the triple intersection, then we would get $\pi_1(X) \approx \mathbb{Z} * \mathbb{Z} * \mathbb{Z}$, but this is not isomorphic to $\mathbb{Z} * \mathbb{Z}$ since it has a different abelianization.



Proof of van Kampen's theorem: We have already proved the first part of the theorem concerning surjectivity of Φ in Lemma 1.15. To prove the harder part of the theorem, that the kernel of Φ is N, we first introduce some terminology. By a *factorization* of an element $[f] \in \pi_1(X)$ we shall mean a formal product $[f_1] \cdots [f_k]$ where:

- Each f_i is a loop in some A_α at the basepoint x_0, and $[f_i] \in \pi_1(A_\alpha)$ is the homotopy class of f_i.
- The loop f is homotopic to $f_1 \cdot \cdots \cdot f_k$ in X.

A factorization of $[f]$ is thus a word in $*_\alpha \pi_1(A_\alpha)$, possibly unreduced, that is mapped to $[f]$ by Φ. Surjectivity of Φ is equivalent to saying that every $[f] \in \pi_1(X)$ has a factorization.

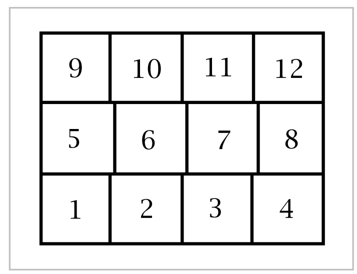
We will be concerned with the uniqueness of factorizations. Call two factorizations of $[f]$ *equivalent* if they are related by a sequence of the following two sorts of moves or their inverses:

- Combine adjacent terms $[f_i][f_{i+1}]$ into a single term $[f_i \cdot f_{i+1}]$ if $[f_i]$ and $[f_{i+1}]$ lie in the same group $\pi_1(A_\alpha)$.
- Regard the term $[f_i] \in \pi_1(A_\alpha)$ as lying in the group $\pi_1(A_\beta)$ rather than $\pi_1(A_\alpha)$ if f_i is a loop in $A_\alpha \cap A_\beta$.

The first move does not change the element of $*_\alpha \pi_1(A_\alpha)$ defined by the factorization. The second move does not change the image of this element in the quotient group $Q = *_\alpha \pi_1(A_\alpha)/N$, by the definition of N. So equivalent factorizations give the same element of Q.

If we can show that any two factorizations of $[f]$ are equivalent, this will say that the map $Q \to \pi_1(X)$ induced by Φ is injective, hence the kernel of Φ is exactly N, and the proof will be complete.

Let $[f_1]\cdots[f_k]$ and $[f_1']\cdots[f_\ell']$ be two factorizations of $[f]$. The composed paths $f_1 \boldsymbol{\cdot} \cdots \boldsymbol{\cdot} f_k$ and $f_1' \boldsymbol{\cdot} \cdots \boldsymbol{\cdot} f_\ell'$ are then homotopic, so let $F : I \times I \to X$ be a homotopy from $f_1 \boldsymbol{\cdot} \cdots \boldsymbol{\cdot} f_k$ to $f_1' \boldsymbol{\cdot} \cdots \boldsymbol{\cdot} f_\ell'$. There exist partitions $0 = s_0 < s_1 < \cdots < s_m = 1$ and $0 = t_0 < t_1 < \cdots < t_n = 1$ such that each rectangle $R_{ij} = [s_{i-1}, s_i] \times [t_{j-1}, t_j]$ is mapped by F into a single A_α, which we label A_{ij}. These partitions may be obtained by covering $I \times I$ by finitely many rectangles $[a,b] \times [c,d]$ each mapping to a single A_α, using a compactness argument, then partitioning $I \times I$ by the union of all the horizontal and vertical lines containing edges of these rectangles. We may assume the s-partition subdivides the partitions giving the products $f_1 \boldsymbol{\cdot} \cdots \boldsymbol{\cdot} f_k$ and $f_1' \boldsymbol{\cdot} \cdots \boldsymbol{\cdot} f_\ell'$. Since F maps a neighborhood of R_{ij} to A_{ij}, we may perturb the vertical sides of the rectangles R_{ij} so that each point of $I \times I$ lies in at most three R_{ij}'s. We may assume there are at least three rows of rectangles, so we can do this perturbation just on the rectangles in the intermediate rows, leaving the top and bottom rows unchanged. Let us relabel the new rectangles $R_1, R_2, \cdots, R_{mn}$, ordering them as in the figure.

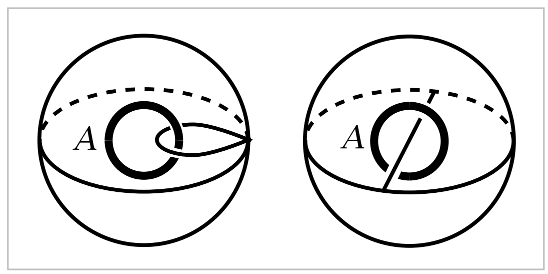


If γ is a path in $I \times I$ from the left edge to the right edge, then the restriction $F \,|\, \gamma$ is a loop at the basepoint x_0 since F maps both the left and right edges of $I \times I$ to x_0. Let γ_r be the path separating the first r rectangles $R_1, \cdots, R_r$ from the remaining rectangles. Thus γ_0 is the bottom edge of $I \times I$ and γ_{mn} is the top edge. We pass from γ_r to γ_{r+1} by pushing across the rectangle R_{r+1}.

Let us call the corners of the R_r's *vertices*. For each vertex v with $F(v) \neq x_0$ we can choose a path g_v from x_0 to $F(v)$ that lies in the intersection of the two or three A_{ij}'s corresponding to the R_r's containing v, since we assume the intersection of any two or three A_{ij}'s is path-connected. Then we obtain a factorization of $[F \,|\, \gamma_r]$ by inserting the appropriate paths $\overline{g}_v g_v$ into $F \,|\, \gamma_r$ at successive vertices, as in the proof of surjectivity of Φ in Lemma 1.15. This factorization depends on certain choices, since the loop corresponding to a segment between two successive vertices can lie in two different A_{ij}'s when there are two different rectangles R_{ij} containing this edge. Different choices of these A_{ij}'s change the factorization of $[F \,|\, \gamma_r]$ to an equivalent factorization, however. Furthermore, the factorizations associated to successive paths γ_r and γ_{r+1} are equivalent since pushing γ_r across R_{r+1} to γ_{r+1} changes $F \,|\, \gamma_r$ to $F \,|\, \gamma_{r+1}$ by a homotopy within the A_{ij} corresponding to R_{r+1}, and we can choose this A_{ij} for all the segments of γ_r and γ_{r+1} in R_{r+1}.

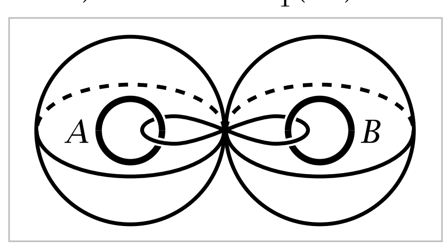
We can arrange that the factorization associated to γ_0 is equivalent to the factorization $[f_1]\cdots[f_k]$ by choosing the path g_v for each vertex v along the lower edge of $I\times I$ to lie not just in the two A_{ij}'s corresponding to the R_s's containing v, but also to lie in the A_α for the f_i containing v in its domain. In case v is the common endpoint of the domains of two consecutive f_i's we have $F(v)=x_0$, so there is no need to choose a g_v for such v's. In similar fashion we may assume that the factorization associated to the final γ_{mn} is equivalent to $[f_1']\cdots[f_\ell']$. Since the factorizations associated to all the γ_r's are equivalent, we conclude that the factorizations $[f_1]\cdots[f_k]$ and $[f_1']\cdots[f_\ell']$ are equivalent. □

Example 1.23: Linking of Circles. We can apply van Kampen's theorem to calculate the fundamental groups of three spaces discussed in the introduction to this chapter, the complements in $\mathbb{R}^3$ of a single circle, two unlinked circles, and two linked circles.

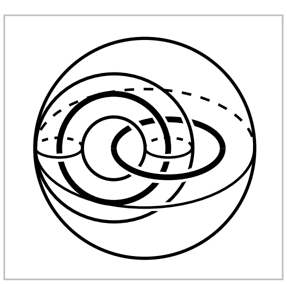


The complement $\mathbb{R}^3-A$ of a single circle A deformation retracts onto a wedge sum $S^1\vee S^2$ embedded in $\mathbb{R}^3-A$ as shown in the first of the two figures at the right. It may be easier to see that $\mathbb{R}^3-A$ deformation retracts onto the union of S^2 with a diameter, as in the second figure, where points outside S^2 deformation retract onto S^2, and points inside S^2 and not in A can be pushed away from A toward S^2 or the diameter. Having this deformation retraction in mind, one can then see how it must be modified if the two endpoints of the diameter are gradually moved toward each other along the equator until they coincide, forming the S^1 summand of $S^1\vee S^2$. Another way of seeing the deformation retraction of $\mathbb{R}^3-A$ onto $S^1\vee S^2$ is to note first that an open ε-neighborhood of $S^1\vee S^2$ obviously deformation retracts onto $S^1\vee S^2$ if ε is sufficiently small. Then observe that this neighborhood is homeomorphic to $\mathbb{R}^3-A$ by a homeomorphism that is the identity on $S^1\vee S^2$. In fact, the neighborhood can be gradually enlarged by homeomorphisms until it becomes all of $\mathbb{R}^3-A$.

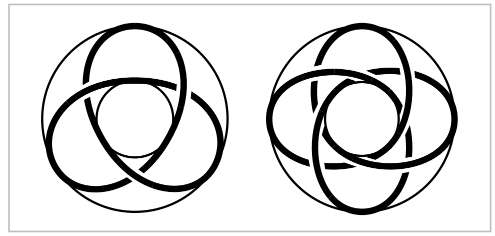
In any event, once we see that $\mathbb{R}^3-A$ deformation retracts to $S^1\vee S^2$, then we immediately obtain isomorphisms $\pi_1(\mathbb{R}^3-A)\approx\pi_1(S^1\vee S^2)\approx\mathbb{Z}$ since $\pi_1(S^2)=0$.



In similar fashion, the complement $\mathbb{R}^3-(A\cup B)$ of two unlinked circles A and B deformation retracts onto $S^1\vee S^1\vee S^2\vee S^2$, as in the figure to the right. From this we get $\pi_1(\mathbb{R}^3-(A\cup B))\approx\mathbb{Z}*\mathbb{Z}$. On the other hand, if A and B are linked, then $\mathbb{R}^3-(A\cup B)$ deformation retracts onto the wedge sum of S^2 and a torus $S^1\times S^1$ separating A and B, as shown in the figure to the left, hence $\pi_1(\mathbb{R}^3-(A\cup B))\approx\pi_1(S^1\times S^1)\approx\mathbb{Z}\times\mathbb{Z}$.



Example 1.24: Torus Knots. For relatively prime positive integers m and n, the **torus knot** $K = K_{m,n} \subset \mathbb{R}^3$ is the image of the embedding $f : S^1 \to S^1 \times S^1 \subset \mathbb{R}^3$, $f(z) = (z^m, z^n)$, where the torus $S^1 \times S^1$ is embedded in $\mathbb{R}^3$ in the standard way. The knot K winds around the torus a total of m times in the longitudinal direction and n times in the meridional direction, as shown in the figure for the cases $(m,n) = (2,3)$ and $(3,4)$. One needs to assume that m and n are relatively prime in order for the map f to be injective. Without this assumption f would be d–to–1 where d is the greatest common divisor of m and n, and the image of f would be the knot $K_{m/d,n/d}$. One could also allow negative values for m or n, but this would only change K to a mirror-image knot.

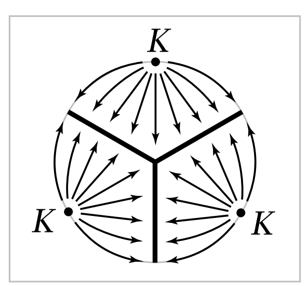


Let us compute $\pi_1(\mathbb{R}^3 - K)$. It is slightly easier to do the calculation with $\mathbb{R}^3$ replaced by its one-point compactification S^3. An application of van Kampen's theorem shows that this does not affect π_1. Namely, write $S^3 - K$ as the union of $\mathbb{R}^3 - K$ and an open ball B formed by the compactification point together with the complement of a large closed ball in $\mathbb{R}^3$ containing K. Both B and $B \cap (\mathbb{R}^3 - K)$ are simply-connected, the latter space being homeomorphic to $S^2 \times \mathbb{R}$. Hence van Kampen's theorem implies that the inclusion $\mathbb{R}^3 - K \hookrightarrow S^3 - K$ induces an isomorphism on π_1.

We compute $\pi_1(S^3 - K)$ by showing that it deformation retracts onto a 2-dimensional complex $X = X_{m,n}$ homeomorphic to the quotient space of a cylinder $S^1 \times I$ under the identifications $(z,0) \sim (e^{2\pi i/m}z, 0)$ and $(z,1) \sim (e^{2\pi i/n}z, 1)$. If we let X_m and X_n be the two halves of X formed by the quotients of $S^1 \times [0, 1/2]$ and $S^1 \times [1/2, 1]$, then X_m and X_n are the mapping cylinders of $z \mapsto z^m$ and $z \mapsto z^n$. The intersection $X_m \cap X_n$ is the circle $S^1 \times \{1/2\}$, the domain end of each mapping cylinder.

To obtain an embedding of X in $S^3 - K$ as a deformation retract we will use the standard decomposition of S^3 into two solid tori $S^1 \times D^2$ and $D^2 \times S^1$, the result of regarding S^3 as $\partial D^4 = \partial(D^2 \times D^2) = \partial D^2 \times D^2 \cup D^2 \times \partial D^2$. Geometrically, the first solid torus $S^1 \times D^2$ can be identified with the compact region in $\mathbb{R}^3$ bounded by the standard torus $S^1 \times S^1$ containing K, and the second solid torus $D^2 \times S^1$ is then the closure of the complement of the first solid torus, together with the compactification point at infinity. Notice that meridional circles in $S^1 \times S^1$ bound disks in the first solid torus, while it is longitudinal circles that bound disks in the second solid torus.

In the first solid torus, K intersects each of the meridian circles $\{x\} \times \partial D^2$ in m equally spaced points, as indicated in the figure at the right, which shows a meridian disk $\{x\} \times D^2$. These m points can be separated by a union of m radial line segments. Letting x vary, these radial segments then trace out a copy of the mapping cylinder X_m in the first solid torus. Symmetrically, there is a copy of the other mapping cylinder X_n in the second solid torus.



The complement of K in the first solid torus deformation retracts onto X_m by flowing within each meridian disk as shown. In similar fashion the complement of K in the second solid torus deformation retracts onto X_n. These two deformation retractions do not agree on their common domain of definition $S^1 \times S^1 - K$, but this is easy to correct by distorting the flows in the two solid tori so that in $S^1 \times S^1 - K$ both flows are orthogonal to K. After this modification we now have a well-defined deformation retraction of $S^3 - K$ onto X. Another way of describing the situation would be to say that for an open ε-neighborhood N of K bounded by a torus T, the complement $S^3 - N$ is the mapping cylinder of a map $T \to X$.

To compute $\pi_1(X)$ we apply van Kampen's theorem to the decomposition of X as the union of X_m and X_n, or more properly, open neighborhoods of these two sets that deformation retract onto them. Both X_m and X_n are mapping cylinders that deformation retract onto circles, and $X_m \cap X_n$ is a circle, so all three of these spaces have fundamental group $\mathbb{Z}$. A loop in $X_m \cap X_n$ representing a generator of $\pi_1(X_m \cap X_n)$ is homotopic in X_m to a loop representing m times a generator, and in X_n to a loop representing n times a generator. Van Kampen's theorem then says that $\pi_1(X)$ is the quotient of the free group on generators a and b obtained by factoring out the normal subgroup generated by the element $a^m b^{-n}$.

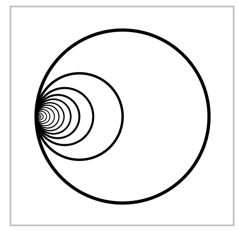
Let us denote by $G_{m,n}$ this group $\pi_1(X_{m,n})$ defined by two generators a and b and one relation $a^m = b^n$. If m or n is 1, then $G_{m,n}$ is infinite cyclic since in these cases the relation just expresses one generator as a power of the other. To describe the structure of $G_{m,n}$ when $m, n > 1$ let us first compute the center of $G_{m,n}$, the subgroup consisting of elements that commute with all elements of $G_{m,n}$. The element $a^m = b^n$ commutes with a and b, so the cyclic subgroup C generated by this element lies in the center. In particular, C is a normal subgroup, so we can pass to the quotient group $G_{m,n}/C$, which is the free product $\mathbb{Z}_m * \mathbb{Z}_n$. According to Exercise 1 at the end of this section, a free product of nontrivial groups has trivial center. From this it follows that C is exactly the center of $G_{m,n}$. As we will see in Example 1.44, the elements a and b have infinite order in $G_{m,n}$, so C is infinite cyclic, but we will not need this fact here.

We will show now that the integers m and n are uniquely determined by the group $\mathbb{Z}_m * \mathbb{Z}_n$, hence also by $G_{m,n}$. The abelianization of $\mathbb{Z}_m * \mathbb{Z}_n$ is $\mathbb{Z}_m \times \mathbb{Z}_n$, of order mn, so the product mn is uniquely determined by $\mathbb{Z}_m * \mathbb{Z}_n$. To determine m and n individually, we use another assertion from Exercise 1 at the end of the section, that all torsion elements of $\mathbb{Z}_m * \mathbb{Z}_n$ are conjugate to elements of one of the subgroups $\mathbb{Z}_m$ and $\mathbb{Z}_n$, hence have order dividing m or n. Thus the maximum order of torsion elements of $\mathbb{Z}_m * \mathbb{Z}_n$ is the larger of m and n. The larger of these two numbers is therefore uniquely determined by the group $\mathbb{Z}_m * \mathbb{Z}_n$, hence also the smaller since the product is uniquely determined.

The preceding analysis of $\pi_1(X_{m,n})$ did not need the assumption that m and n

are relatively prime, which was used only to relate $X_{m,n}$ to torus knots. An interesting fact is that $X_{m,n}$ can be embedded in $\mathbb{R}^3$ only when m and n are relatively prime. This is shown in the remarks following Corollary 3.46. For example, $X_{2,2}$ is the Klein bottle since it is the union of two copies of the Möbius band X_2 with their boundary circles identified, so this nonembeddability statement generalizes the fact that the Klein bottle cannot be embedded in $\mathbb{R}^3$.

An algorithm for computing a presentation for $\pi_1(\mathbb{R}^3 - K)$ for an arbitrary smooth or piecewise linear knot K is described in the exercises, but the problem of determining when two of these fundamental groups are isomorphic is generally much more difficult than in the special case of torus knots.



Example 1.25: The Shrinking Wedge of Circles. Consider the subspace $X \subset \mathbb{R}^2$ that is the union of the circles C_n of radius $1/n$ and center $(1/n, 0)$ for $n = 1, 2, \cdots$. At first glance one might confuse X with the wedge sum of an infinite sequence of circles, but we will show that X has a much larger fundamental group than the wedge sum. Consider the retractions $r_n : X \to C_n$ collapsing all C_i's except C_n to the origin. Each r_n induces a surjection $\rho_n : \pi_1(X) \to \pi_1(C_n) \approx \mathbb{Z}$, where we take the origin as the basepoint. The product of the ρ_n's is a homomorphism $\rho : \pi_1(X) \to \prod_\infty \mathbb{Z}$ to the direct product (not the direct sum) of infinitely many copies of $\mathbb{Z}$, and ρ is surjective since for every sequence of integers k_n we can construct a loop $f : I \to X$ that wraps k_n times around C_n in the time interval $[1 - 1/n, 1 - 1/n+1]$. This infinite composition of loops is certainly continuous at each time less than 1, and it is continuous at time 1 since every neighborhood of the basepoint in X contains all but finitely many of the circles C_n. Since $\pi_1(X)$ maps onto the uncountable group $\prod_\infty \mathbb{Z}$, it is uncountable. On the other hand, the fundamental group of a wedge sum of countably many circles is countably generated, hence countable.

The group $\pi_1(X)$ is actually far more complicated than $\prod_\infty \mathbb{Z}$. For one thing, it is nonabelian, since the retraction $X \to C_1 \cup \cdots \cup C_n$ that collapses all the circles smaller than C_n to the basepoint induces a surjection from $\pi_1(X)$ to a free group on n generators. For a complete description of $\pi_1(X)$ see [Cannon & Conner 2000].

It is a theorem of [Shelah 1988] that for a path-connected, locally path-connected compact metric space X, $\pi_1(X)$ is either finitely generated or uncountable.

Applications to Cell Complexes

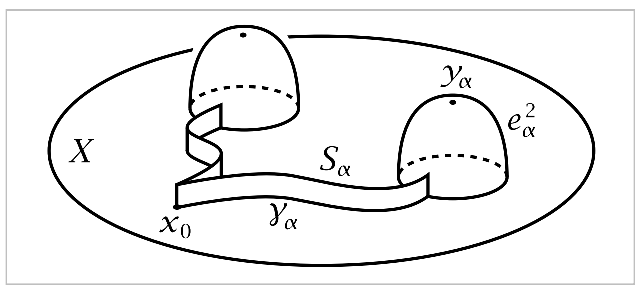
For the remainder of this section we shall be interested in cell complexes, and in particular in how the fundamental group is affected by attaching 2-cells.

Suppose we attach a collection of 2-cells e^2_α to a path-connected space X via maps $\varphi_\alpha : S^1 \to X$, producing a space Y. If s_0 is a basepoint of S^1 then φ_α determines a loop at $\varphi_\alpha(s_0)$ that we shall call φ_α, even though technically loops are maps $I \to X$ rather than $S^1 \to X$. For different α's the basepoints $\varphi_\alpha(s_0)$ of these loops φ_α may not all coincide. To remedy this, choose a basepoint $x_0 \in X$ and a path γ_α in X from x_0 to $\varphi_\alpha(s_0)$ for each α. Then $\gamma_\alpha \varphi_\alpha \overline{\gamma}_\alpha$ is a loop at x_0. This loop may not be nullhomotopic in X, but it will certainly be nullhomotopic after the cell e^2_α is attached. Thus the normal subgroup $N \subset \pi_1(X, x_0)$ generated by all the loops $\gamma_\alpha \varphi_\alpha \overline{\gamma}_\alpha$ for varying α lies in the kernel of the map $\pi_1(X, x_0) \to \pi_1(Y, x_0)$ induced by the inclusion $X \hookrightarrow Y$.

Proposition 1.26. (a) *If Y is obtained from X by attaching 2-cells as described above, then the inclusion $X \hookrightarrow Y$ induces a surjection $\pi_1(X, x_0) \to \pi_1(Y, x_0)$ whose kernel is N. Thus $\pi_1(Y) \approx \pi_1(X)/N$.*
(b) *If Y is obtained from X by attaching n-cells for a fixed $n > 2$, then the inclusion $X \hookrightarrow Y$ induces an isomorphism $\pi_1(X, x_0) \approx \pi_1(Y, x_0)$.*
(c) *For a path-connected cell complex X the inclusion of the 2-skeleton $X^2 \hookrightarrow X$ induces an isomorphism $\pi_1(X^2, x_0) \approx \pi_1(X, x_0)$.*

It follows from (a) that N is independent of the choice of the paths γ_α, but this can also be seen directly: If we replace γ_α by another path η_α having the same endpoints, then $\gamma_\alpha \varphi_\alpha \overline{\gamma}_\alpha$ changes to $\eta_\alpha \varphi_\alpha \overline{\eta}_\alpha = (\eta_\alpha \overline{\gamma}_\alpha)\gamma_\alpha \varphi_\alpha \overline{\gamma}_\alpha(\gamma_\alpha \overline{\eta}_\alpha)$, so $\gamma_\alpha \varphi_\alpha \overline{\gamma}_\alpha$ and $\eta_\alpha \varphi_\alpha \overline{\eta}_\alpha$ define conjugate elements of $\pi_1(X, x_0)$.

Proof: (a) Let us expand Y to a slightly larger space Z that deformation retracts onto Y and is more convenient for applying van Kampen's theorem. The space Z is obtained from Y by attaching rectangular strips $S_\alpha = I \times I$, with the lower edge $I \times \{0\}$ attached along γ_α, the right edge $\{1\} \times I$ attached along an arc that starts at $\varphi_\alpha(s_0)$ and goes radially into e^2_α, and all the left edges $\{0\} \times I$ of the different strips identified together. The top edges of the strips are not attached to anything, and this allows us to deformation retract Z onto Y.



In each cell e^2_α choose a point y_α not in the arc along which S_α is attached. Let $A = Z - \bigcup_\alpha \{y_\alpha\}$ and let $B = Z - X$. Then A deformation retracts onto X, and B is contractible. Since $\pi_1(B) = 0$, van Kampen's theorem applied to the cover $\{A, B\}$ says that $\pi_1(Z)$ is isomorphic to the quotient of $\pi_1(A)$ by the normal subgroup generated by the image of the map $\pi_1(A \cap B) \to \pi_1(A)$. More specifically, choose a basepoint $z_0 \in A \cap B$ near x_0 on the segment where all the strips S_α intersect, and choose loops δ_α in $A \cap B$ based at z_0 representing the elements of $\pi_1(A, z_0)$ corresponding to $[\gamma_\alpha \varphi_\alpha \overline{\gamma}_\alpha] \in \pi_1(A, x_0)$ under the basepoint-change isomorphism β_h for h the line segment connecting z_0 to x_0 in the intersection of the S_α's. To finish the proof of part (a) we just need to check that $\pi_1(A \cap B, z_0)$ is generated by the loops δ_α. This can be done by another application of van Kampen's theorem, this time to the cover of $A \cap B$ by the open sets $A_\alpha = A \cap B - \bigcup_{\beta \neq \alpha} e^2_\beta$. Since A_α deformation retracts onto a circle in $e^2_\alpha - \{y_\alpha\}$, we have $\pi_1(A_\alpha, z_0) \approx \mathbb{Z}$ generated by δ_α.

The proof of (b) follows the same plan with cells e^n_α instead of e^2_α. The only difference is that A_α deformation retracts onto a sphere S^{n-1} so $\pi_1(A_a) = 0$ if $n > 2$ by Proposition 1.14. Hence $\pi_1(A \cap B) = 0$ and the result follows.

Part (c) follows from (b) by induction when X is finite-dimensional, so $X = X^n$ for some n. When X is not finite-dimensional we argue as follows. Let $f: I \to X$ be a loop at the basepoint $x_0 \in X^2$. This has compact image, which must lie in X^n for some n by Proposition A.1 in the Appendix. Part (b) then implies that f is homotopic to a loop in X^2. Thus $\pi_1(X^2, x_0) \to \pi_1(X, x_0)$ is surjective. To see that it is also injective, suppose that f is a loop in X^2 which is nullhomotopic in X via a homotopy $F: I \times I \to X$. This has compact image lying in some X^n, and we can assume $n > 2$. Since $\pi_1(X^2, x_0) \to \pi_1(X^n, x_0)$ is injective by (b), we conclude that f is nullhomotopic in X^2. $\square$

As a first application we compute the fundamental group of the orientable surface M_g of genus g. This has a cell structure with one 0-cell, $2g$ 1-cells, and one 2-cell, as we saw in Chapter 0. The 1-skeleton is a wedge sum of $2g$ circles, with fundamental group free on $2g$ generators. The 2-cell is attached along the loop given by the product of the commutators of these generators, say $[a_1, b_1] \cdots [a_g, b_g]$. Therefore

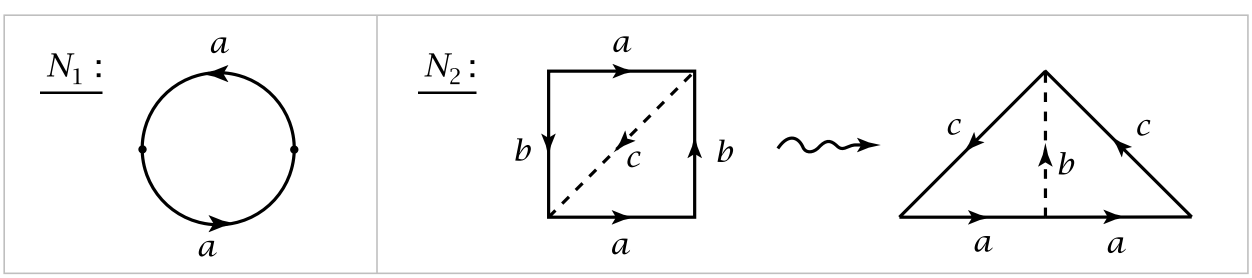
$$\pi_1(M_g) \approx \langle\, a_1, b_1, \cdots, a_g, b_g \mid [a_1, b_1] \cdots [a_g, b_g] \,\rangle$$

where $\langle g_\alpha \mid r_\beta \rangle$ denotes the group with generators g_α and relators r_β, in other words, the free group on the generators g_α modulo the normal subgroup generated by the words r_β in these generators.

Corollary 1.27. *The surface M_g is not homeomorphic, or even homotopy equivalent, to M_h if $g \neq h$.*

Proof: The abelianization of $\pi_1(M_g)$ is the direct sum of $2g$ copies of $\mathbb{Z}$. So if $M_g \simeq M_h$ then $\pi_1(M_g) \approx \pi_1(M_h)$, hence the abelianizations of these groups are isomorphic, which implies $g = h$. $\square$

Nonorientable surfaces can be treated in the same way. If we attach a 2-cell to the wedge sum of g circles by the word $a_1^2 \cdots a_g^2$ we obtain a nonorientable surface N_g. For example, N_1 is the projective plane $\mathbb{R}P^2$, the quotient of D^2 with antipodal points of ∂D^2 identified, and N_2 is the Klein bottle, though the more usual representation of the Klein bottle is as a square with opposite sides identified via the word $aba^{-1}b$.



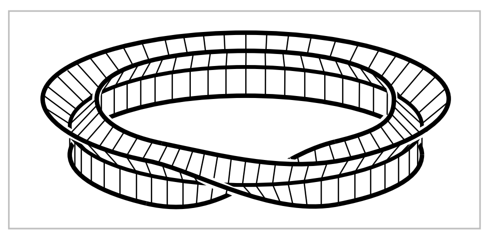
If one cuts the square along a diagonal and reassembles the resulting two triangles as shown in the figure, one obtains the other representation as a square with sides identified via the word a^2c^2. By the proposition, $\pi_1(N_g) \approx \langle\, a_1, \cdots, a_g \mid a_1^2 \cdots a_g^2 \,\rangle$. This abelianizes to the direct sum of $\mathbb{Z}_2$ with $g-1$ copies of $\mathbb{Z}$ since in the abelianization we can rechoose the generators to be $a_1, \cdots, a_{g-1}$ and $a_1 + \cdots + a_g$, with $2(a_1 + \cdots + a_g) = 0$. Hence N_g is not homotopy equivalent to N_h if $g \neq h$, nor is N_g homotopy equivalent to any orientable surface M_h.

Here is another application of the preceding proposition:

Corollary 1.28. *For every group G there is a 2-dimensional cell complex X_G with $\pi_1(X_G) \approx G$.*

Proof: Choose a presentation $G = \langle\, g_\alpha \mid r_\beta \,\rangle$. This exists since every group is a quotient of a free group, so the g_α's can be taken to be the generators of this free group with the r_β's generators of the kernel of the map from the free group to G. Now construct X_G from $\bigvee_\alpha S^1_\alpha$ by attaching 2-cells e^2_β by the loops specified by the words r_β. $\square$

Example 1.29. If $G = \langle\, a \mid a^n \,\rangle = \mathbb{Z}_n$ then X_G is S^1 with a cell e^2 attached by the map $z \mapsto z^n$, thinking of S^1 as the unit circle in $\mathbb{C}$. When $n = 2$ we get $X_G = \mathbb{R}\mathrm{P}^2$, but for $n > 2$ the space X_G is not a surface since there are n 'sheets' of e^2 attached at each point of the circle $S^1 \subset X_G$. For example, when $n = 3$ one can construct a neighborhood N of S^1 in X_G by taking the product of the graph $\curlyvee$ with the interval I, and then identifying the two ends of this product via a one-third twist as shown in the figure. The boundary of N consists of a single circle, formed by the three endpoints of each $\curlyvee$ cross section of N. To complete the construction of X_G from N one attaches a disk along the boundary circle of N. This cannot be done in $\mathbb{R}^3$, though it can in $\mathbb{R}^4$. For $n = 4$ one would use the graph $\times$ instead of $\curlyvee$, with a one-quarter twist instead of a one-third twist. For larger n one would use an n-pointed 'asterisk' and a ${}^1\!/_n$ twist.



Exercises

1. Show that the free product $G * H$ of nontrivial groups G and H has trivial center, and that the only elements of $G * H$ of finite order are the conjugates of finite-order elements of G and H.

2. Let $X \subset \mathbb{R}^m$ be the union of convex open sets $X_1, \cdots, X_n$ such that $X_i \cap X_j \cap X_k \neq \varnothing$ for all i, j, k. Show that X is simply-connected.

3. Show that the complement of a finite set of points in $\mathbb{R}^n$ is simply-connected if $n \geq 3$.

4. Let $X \subset \mathbb{R}^3$ be the union of n lines through the origin. Compute $\pi_1(\mathbb{R}^3 - X)$.

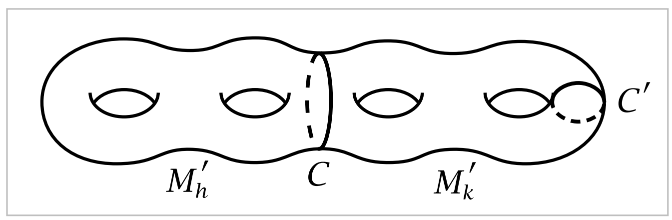
5. Let $X \subset \mathbb{R}^2$ be a connected graph that is the union of a finite number of straight line segments. Show that $\pi_1(X)$ is free with a basis consisting of loops formed by the boundaries of the bounded complementary regions of X, joined to a basepoint by suitably chosen paths in X. [Assume the Jordan curve theorem for polygonal simple closed curves, which is equivalent to the case that X is homeomorphic to S^1.]

6. Use Proposition 1.26 to show that the complement of a closed discrete subspace of $\mathbb{R}^n$ is simply-connected if $n \geq 3$.

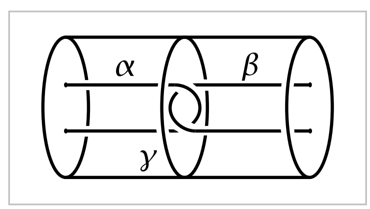
7. Let X be the quotient space of S^2 obtained by identifying the north and south poles to a single point. Put a cell complex structure on X and use this to compute $\pi_1(X)$.

8. Compute the fundamental group of the space obtained from two tori $S^1 \times S^1$ by identifying a circle $S^1 \times \{x_0\}$ in one torus with the corresponding circle $S^1 \times \{x_0\}$ in the other torus.

9. In the surface M_g of genus g, let C be a circle that separates M_g into two compact subsurfaces M'_h and M'_k obtained from the closed surfaces M_h and M_k by deleting an open disk from each. Show that M'_h does not retract onto its boundary circle C, and hence M_g does not retract onto C. [Hint: abelianize π_1.] But show that M_g does retract onto the nonseparating circle C' in the figure.

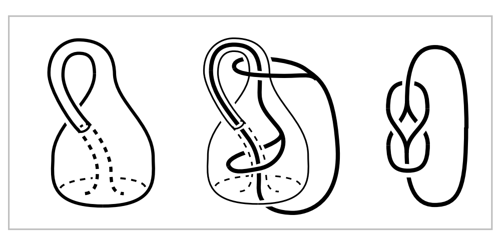


10. Consider two arcs α and β embedded in $D^2 \times I$ as shown in the figure. The loop γ is obviously nullhomotopic in $D^2 \times I$, but show that there is no nullhomotopy of γ in the complement of $\alpha \cup \beta$.



11. The **mapping torus** T_f of a map $f : X \to X$ is the quotient of $X \times I$ obtained by identifying each point $(x, 0)$ with $(f(x), 1)$. In the case $X = S^1 \vee S^1$ with f basepoint-preserving, compute a presentation for $\pi_1(T_f)$ in terms of the induced map $f_* : \pi_1(X) \to \pi_1(X)$. Do the same when $X = S^1 \times S^1$. [One way to do this is to regard T_f as built from $X \vee S^1$ by attaching cells.]

12. The Klein bottle is usually pictured as a subspace of $\mathbb{R}^3$ like the subspace $X \subset \mathbb{R}^3$ shown in the first figure at the right. If one wanted a model that could actually function as a bottle, one would delete the open disk bounded by the circle of self-intersection of X, producing a subspace $Y \subset X$. Show that $\pi_1(X) \approx \mathbb{Z} * \mathbb{Z}$ and that



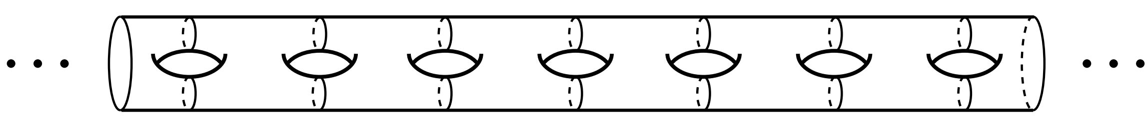
$\pi_1(Y)$ has the presentation $\langle\, a,b,c \mid aba^{-1}b^{-1}cb^{\varepsilon}c^{-1} \,\rangle$ for $\varepsilon = \pm 1$. (Changing the sign of ε gives an isomorphic group, as it happens.) Show also that $\pi_1(Y)$ is isomorphic to $\pi_1(\mathbb{R}^3 - Z)$ for Z the graph shown in the figure. The groups $\pi_1(X)$ and $\pi_1(Y)$ are not isomorphic, but this is not easy to prove; see the discussion in Example 1B.13.

13. The space Y in the preceding exercise can be obtained from a disk with two holes by identifying its three boundary circles. There are only two essentially different ways of identifying the three boundary circles. Show that the other way yields a space Z with $\pi_1(Z)$ not isomorphic to $\pi_1(Y)$. [Abelianize the fundamental groups to show they are not isomorphic.]

14. Consider the quotient space of a cube I^3 obtained by identifying each square face with the opposite square face via the right-handed screw motion consisting of a translation by one unit in the direction perpendicular to the face combined with a one-quarter twist of the face about its center point. Show this quotient space X is a cell complex with two 0-cells, four 1-cells, three 2-cells, and one 3-cell. Using this structure, show that $\pi_1(X)$ is the quaternion group $\{\pm 1, \pm i, \pm j, \pm k\}$, of order eight.

15. Given a space X with basepoint $x_0 \in X$, we may construct a CW complex $L(X)$ having a single 0-cell, a 1-cell e^1_γ for each loop γ in X based at x_0, and a 2-cell e^2_τ for each map τ of a standard triangle PQR into X taking the three vertices P, Q, and R of the triangle to x_0. The 2-cell e^2_τ is attached to the three 1-cells that are the loops obtained by restricting τ to the three oriented edges PQ, PR, and QR. Show that the natural map $L(X) \to X$ induces an isomorphism $\pi_1(L(X)) \approx \pi_1(X, x_0)$.

16. Show that the fundamental group of the surface of infinite genus shown below is free on an infinite number of generators.



17. Show that $\pi_1(\mathbb{R}^2 - \mathbb{Q}^2)$ is uncountable.

18. In this problem we use the notions of suspension, reduced suspension, cone, and mapping cone defined in Chapter 0. Let X be the subspace of $\mathbb{R}$ consisting of the sequence $1, {}^1\!/_2, {}^1\!/_3, {}^1\!/_4, \cdots$ together with its limit point 0.

(a) For the suspension SX, show that $\pi_1(SX)$ is free on a countably infinite set of generators, and deduce that $\pi_1(SX)$ is countable. In contrast to this, the reduced suspension ΣX, obtained from SX by collapsing the segment $\{0\} \times I$ to a point, is the shrinking wedge of circles in Example 1.25, with an uncountable fundamental group.

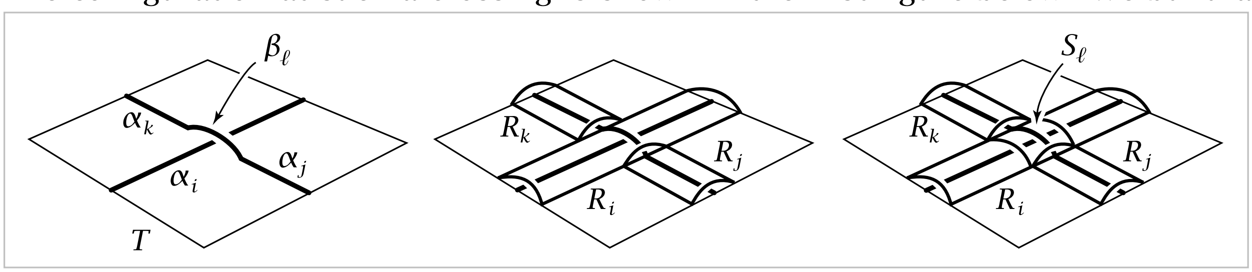
(b) Let C be the mapping cone of the quotient map $SX \to \Sigma X$. Show that $\pi_1(C)$ is uncountable by constructing a homomorphism from $\pi_1(C)$ onto $\prod_\infty \mathbb{Z} / \bigoplus_\infty \mathbb{Z}$. Note that C is the reduced suspension of the cone CX. Thus the reduced suspension of a contractible space need not be contractible, unlike the unreduced suspension.

19. Show that the subspace of $\mathbb{R}^3$ that is the union of the spheres of radius ${}^1\!/_n$ and center $({}^1\!/_n, 0, 0)$ for $n = 1, 2, \cdots$ is simply-connected.

20. Let X be the subspace of $\mathbb{R}^2$ that is the union of the circles C_n of radius n and center $(n, 0)$ for $n = 1, 2, \cdots$. Show that $\pi_1(X)$ is the free group $*_n\,\pi_1(C_n)$, the same as for the infinite wedge sum $\bigvee_\infty S^1$. Show that X and $\bigvee_\infty S^1$ are in fact homotopy equivalent, but not homeomorphic.

21. Show that the join $X * Y$ of two nonempty spaces X and Y is simply-connected if X is path-connected.

22. In this exercise we describe an algorithm for computing a presentation of the fundamental group of the complement of a smooth or piecewise linear knot K in $\mathbb{R}^3$, called the *Wirtinger presentation*. To begin, we position the knot to lie almost flat on a table, so that K consists of finitely many disjoint arcs α_i where it intersects the table top together with finitely many disjoint arcs β_ℓ where K crosses over itself. The configuration at such a crossing is shown in the first figure below. We build a



2-dimensional complex X that is a deformation retract of $\mathbb{R}^3 - K$ by the following three steps. First, start with the rectangle T formed by the table top. Next, just above each arc α_i place a long, thin rectangular strip R_i, curved to run parallel to α_i along the full length of α_i and arched so that the two long edges of R_i are identified with points of T, as in the second figure. Any arcs β_ℓ that cross over α_i are positioned to lie in R_i. Finally, over each arc β_ℓ put a square S_ℓ, bent downward along its four edges so that these edges are identified with points of three strips R_i, R_j, and R_k as in the third figure; namely, two opposite edges of S_ℓ are identified with short edges of R_j and R_k and the other two opposite edges of S_ℓ are identified with two arcs crossing the interior of R_i. The knot K is now a subspace of X, but after we lift K up slightly into the complement of X, it becomes evident that X is a deformation retract of $\mathbb{R}^3 - K$.

(a) Assuming this bit of geometry, show that $\pi_1(\mathbb{R}^3 - K)$ has a presentation with one generator x_i for each strip R_i and one relation of the form $x_i x_j x_i^{-1} = x_k$ for each square S_ℓ, where the indices are as in the figures above. [To get the correct signs it is helpful to use an orientation of K.]

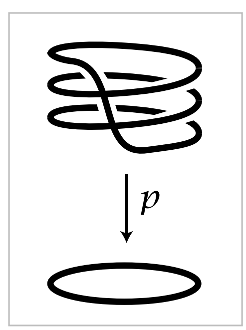
(b) Use this presentation to show that the abelianization of $\pi_1(\mathbb{R}^3 - K)$ is $\mathbb{Z}$.

1.3 Covering Spaces

We come now to the second main topic of this chapter, covering spaces. We have already encountered these briefly in our calculation of $\pi_1(S^1)$ which used the example of the projection $\mathbb{R}\to S^1$ of a helix onto a circle. As we will see, covering spaces can be used to calculate fundamental groups of other spaces as well. But the connection between the fundamental group and covering spaces runs much deeper than this, and in many ways they can be regarded as two viewpoints toward the same thing. Algebraic aspects of the fundamental group can often be translated into the geometric language of covering spaces. This is exemplified in one of the main results in this section, an exact correspondence between connected covering spaces of a given space X and subgroups of $\pi_1(X)$. This is strikingly reminiscent of Galois theory, with its correspondence between field extensions and subgroups of the Galois group.

Let us recall the definition. A **covering space** of a space X is a space $\widetilde{X}$ together with a map $p:\widetilde{X}\to X$ satisfying the following condition: Each point $x\in X$ has an open neighborhood U in X such that $p^{-1}(U)$ is a union of disjoint open sets in $\widetilde{X}$, each of which is mapped homeomorphically onto U by p. Such a U is called **evenly covered** and the disjoint open sets in $\widetilde{X}$ that project homeomorphically to U by p are called **sheets** of $\widetilde{X}$ over U. If U is connected these sheets are the connected components of $p^{-1}(U)$ so in this case they are uniquely determined by U, but when U is not connected the decomposition of $p^{-1}(U)$ into sheets may not be unique. We allow $p^{-1}(U)$ to be empty, the union of an empty collection of sheets over U, so p need not be surjective. The number of sheets over U is the cardinality of $p^{-1}(x)$ for $x\in U$. As x varies over X this number is locally constant, so it is constant if X is connected.

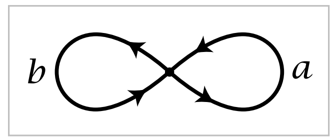
An example related to the helix example is the helicoid surface $S\subset\mathbb{R}^3$ consisting of points of the form $(s\cos 2\pi t, s\sin 2\pi t, t)$ for $(s,t)\in(0,\infty)\times\mathbb{R}$. This projects onto $\mathbb{R}^2-\{0\}$ via the map $(x,y,z)\mapsto(x,y)$, and this projection defines a covering space $p:S\to\mathbb{R}^2-\{0\}$ since each point of $\mathbb{R}^2-\{0\}$ is contained in an open disk U in $\mathbb{R}^2-\{0\}$ with $p^{-1}(U)$ consisting of countably many disjoint open disks in S projecting homeomorphically onto U.



Another example is the map $p:S^1\to S^1$, $p(z)=z^n$ where we view z as a complex number with $|z|=1$ and n is any positive integer. The closest one can come to realizing this covering space as a linear projection in 3-space analogous to the projection of the helix is to draw a circle wrapping around a cylinder n times and intersecting itself in $n-1$ points that one has to imagine are not really intersections. For an alternative picture without this defect, embed S^1 in the boundary torus of a solid torus $S^1\times D^2$ so that it winds n times monotonically around the S^1 factor without self-intersections, then restrict the projection $S^1 \times D^2 \to S^1 \times \{0\}$ to this embedded circle. The figure for Example 1.29 in the preceding section illustrates the case $n = 3$.

These n-sheeted covering spaces $S^1 \to S^1$ for $n \geq 1$ together with the infinite-sheeted helix example exhaust all the connected coverings spaces of S^1, as our general theory will show. There are many other disconnected covering spaces of S^1, such as n disjoint circles each mapped homeomorphically onto S^1, but these disconnected covering spaces are just disjoint unions of connected ones. We will usually restrict our attention to connected covering spaces as these contain most of the interesting features of covering spaces.

The covering spaces of $S^1 \vee S^1$ form a remarkably rich family illustrating most of the general theory very concretely, so let us look at a few of these covering spaces to get an idea of what is going on. To abbreviate notation, set $X = S^1 \vee S^1$. We view this as a graph with one vertex and two edges. We label the edges a and b and we choose orientations for a and b. Now let $\widetilde{X}$ be any other graph with four ends of edges at each vertex, as in X, and suppose each edge of $\widetilde{X}$ has been assigned a label a or b and an orientation in such a way that the local picture near each vertex is the same as in X, so there is an a-edge end oriented toward the vertex, an a-edge end oriented away from the vertex, a b-edge end oriented toward the vertex, and a b-edge end oriented away from the vertex. To give a name to this structure, let us call $\widetilde{X}$ a 2-*oriented* graph.



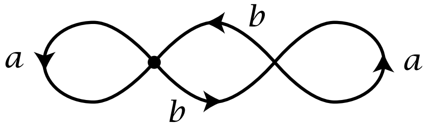
The table on the next page shows just a small sample of the infinite variety of possible examples.

Given a 2-oriented graph $\widetilde{X}$ we can construct a map $p : \widetilde{X} \to X$ sending all vertices of $\widetilde{X}$ to the vertex of X and sending each edge of $\widetilde{X}$ to the edge of X with the same label by a map that is a homeomorphism on the interior of the edge and preserves orientation. It is clear that the covering space condition is satisfied for p. Conversely, every covering space of X is a graph that inherits a 2-orientation from X.

As the reader will discover by experimentation, it seems that every graph having four edge ends at each vertex can be 2-oriented. This can be proved for finite graphs as follows. A very classical and easily shown fact is that every finite connected graph with an even number of edge ends at each vertex has an Eulerian circuit, a loop traversing each edge exactly once. If there are four edge ends at each vertex, then labeling the edges of an Eulerian circuit alternately a and b produces a labeling with two a-edge ends and two b-edge ends at each vertex. The union of the a edges is then a collection of disjoint circles, as is the union of the b edges. Choosing orientations for all these circles gives a 2-orientation. It is a theorem in graph theory that infinite graphs with four edge ends at each vertex can also be 2-oriented; see Chapter 13 of [König 1990] for a proof. There is also a generalization to n-oriented graphs, which are covering spaces of the wedge sum of n circles.

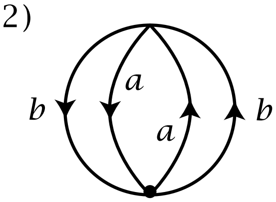
Some Covering Spaces of $S^1 \vee S^1$

(1)



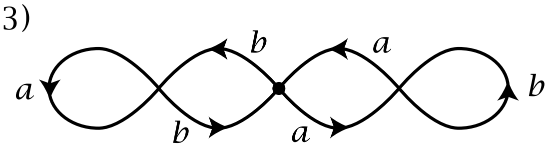
$\langle a, b^2, bab^{-1} \rangle$

(2)



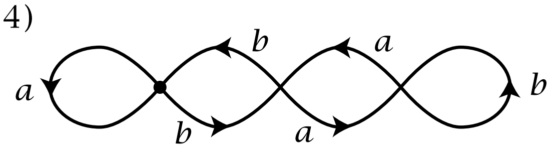
$\langle a^2, b^2, ab \rangle$

(3)



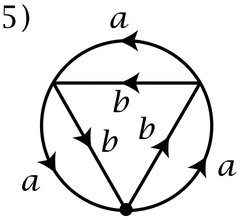
$\langle a^2, b^2, aba^{-1}, bab^{-1} \rangle$

(4)



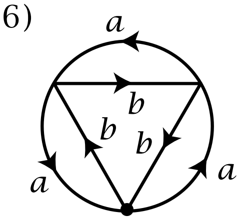
$\langle a, b^2, ba^2b^{-1}, baba^{-1}b^{-1} \rangle$

(5)



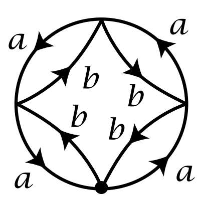
$\langle a^3, b^3, ab^{-1}, b^{-1}a \rangle$

(6)



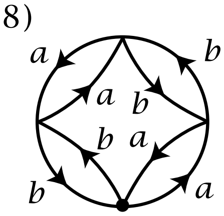
$\langle a^3, b^3, ab, ba \rangle$

(7)



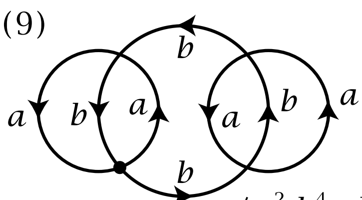
$\langle a^4, b^4, ab, ba, a^2b^2 \rangle$

(8)



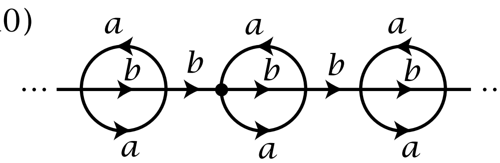
$\langle a^2, b^2, (ab)^2, (ba)^2, ab^2a \rangle$

(9)



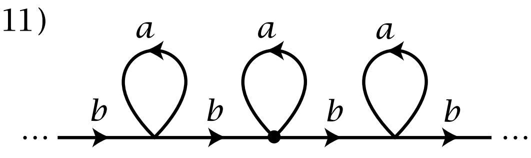
$\langle a^2, b^4, ab, ba^2b^{-1}, bab^{-2} \rangle$

(10)



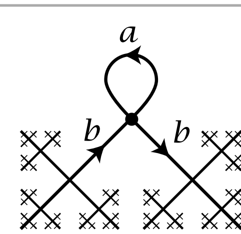
$\langle b^{2n}ab^{-2n-1}, b^{2n+1}ab^{-2n} \mid n \in \mathbb{Z} \rangle$

(11)



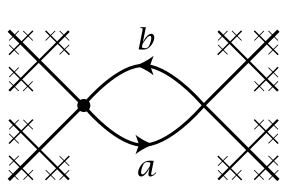
$\langle b^nab^{-n} \mid n \in \mathbb{Z} \rangle$

(12)



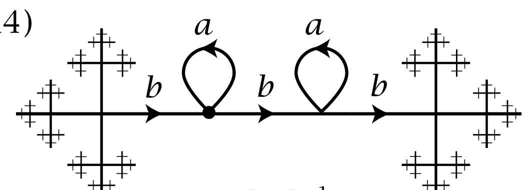
$\langle a \rangle$

(13)



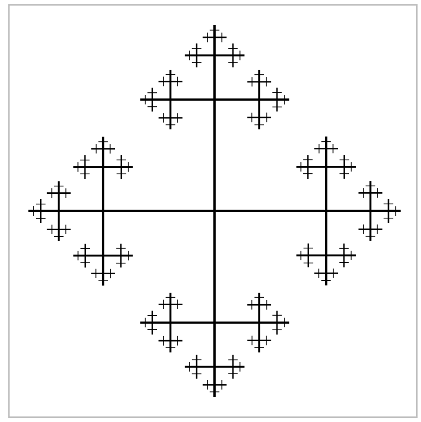
$\langle ab \rangle$

(14)



$\langle a, bab^{-1} \rangle$

A simply-connected covering space of $X = S^1 \vee S^1$ can be constructed in the following way. Start with the open intervals $(-1,1)$ in the coordinate axes of $\mathbb{R}^2$. Next, for a fixed number λ, $0 < \lambda < 1/2$, for example $\lambda = 1/3$, adjoin four open segments of length 2λ, at distance λ from the ends of the previous segments and perpendicular to them, the new shorter segments being bisected by the older ones. For the third stage, add perpendicular open segments of length $2\lambda^2$ at distance λ^2 from the endpoints of all the previous segments and bisected by them. The process is now repeated indefinitely, at the n^{th} stage adding open segments of length $2\lambda^{n-1}$ at distance λ^{n-1} from all the previous endpoints. The union of all these open segments is a graph, with vertices the intersection points of horizontal and vertical segments, and edges the subsegments between adjacent vertices. We label all the horizontal edges a, oriented to the right, and all the vertical edges b, oriented upward.



This covering space is called the *universal cover* of X because, as our general theory will show, it is a covering space of every other connected covering space of X.

The covering spaces (1)–(14) in the table are all nonsimply-connected. Their fundamental groups are free with bases represented by the loops specified by the listed words in a and b, starting at the basepoint $\widetilde{x}_0$ indicated by the heavily shaded vertex. This can be proved in each case by applying van Kampen's theorem. One can also interpret the list of words as generators of the image subgroup $p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$ in $\pi_1(X,x_0) = \langle\, a,b \,\rangle$. A general fact we shall prove about covering spaces is that the induced map $p_*:\pi_1(\widetilde{X},\widetilde{x}_0)\to\pi_1(X,x_0)$ is always injective. Thus we have the at-first-glance paradoxical fact that the free group on two generators can contain as a subgroup a free group on any finite number of generators, or even on a countably infinite set of generators as in examples (10) and (11).

Changing the basepoint vertex changes the subgroup $p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$ to a conjugate subgroup in $\pi_1(X,x_0)$. The conjugating element of $\pi_1(X,x_0)$ is represented by any loop that is the projection of a path in $\widetilde{X}$ joining one basepoint to the other. For example, the covering spaces (3) and (4) differ only in the choice of basepoints, and the corresponding subgroups of $\pi_1(X,x_0)$ differ by conjugation by b.

The main classification theorem for covering spaces says that by associating the subgroup $p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$ to the covering space $p:\widetilde{X}\to X$, we obtain a one-to-one correspondence between all the different connected covering spaces of X and the conjugacy classes of subgroups of $\pi_1(X,x_0)$. If one keeps track of the basepoint vertex $\widetilde{x}_0 \in \widetilde{X}$, then this is a one-to-one correspondence between covering spaces $p:(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$ and actual subgroups of $\pi_1(X,x_0)$, not just conjugacy classes. Of course, for these statements to make sense one has to have a precise notion of when two covering spaces are the same, or 'isomorphic'. In the case at hand, an isomorphism between covering spaces of X is just a graph isomorphism that preserves the labeling and orientations of edges. Thus the covering spaces in (3) and (4) are isomorphic, but not by an isomorphism preserving basepoints, so the two subgroups of $\pi_1(X,x_0)$ corresponding to these covering spaces are distinct but conjugate. On the other hand, the two covering spaces in (5) and (6) are not isomorphic, though the graphs are homeomorphic, so the corresponding subgroups of $\pi_1(X,x_0)$ are isomorphic but not conjugate.

Some of the covering spaces (1)–(14) are more symmetric than others, where by a 'symmetry' we mean an automorphism of the graph preserving the labeling and orientations. The most symmetric covering spaces are those having symmetries taking any one vertex onto any other. The examples (1), (2), (5)–(8), and (11) are the ones with this property. We shall see that a covering space of X has maximal symmetry exactly when the corresponding subgroup of $\pi_1(X,x_0)$ is a normal subgroup, and in this case the symmetries form a group isomorphic to the quotient group of $\pi_1(X,x_0)$ by the normal subgroup. Since every group generated by two elements is a quotient group of $\mathbb{Z}*\mathbb{Z}$, this implies that every two-generator group is the symmetry group of some covering space of X.

Lifting Properties

Covering spaces are defined in fairly geometric terms, as maps $p:\widetilde{X}\to X$ that are local homeomorphisms in a rather strong sense. But from the viewpoint of algebraic topology, the distinctive feature of covering spaces is their behavior with respect to lifting of maps. Recall the terminology from the proof of Theorem 1.7: A **lift** of a map $f:Y\to X$ is a map $\widetilde{f}:Y\to\widetilde{X}$ such that $p\widetilde{f}=f$. We will describe three special lifting properties of covering spaces and derive a few applications of these.

First we have the **homotopy lifting property**, also known as the **covering homotopy property**:

Proposition 1.30. *Given a covering space $p:\widetilde{X}\to X$, a homotopy $f_t:Y\to X$, and a map $\widetilde{f}_0:Y\to\widetilde{X}$ lifting f_0, then there exists a unique homotopy $\widetilde{f}_t:Y\to\widetilde{X}$ of $\widetilde{f}_0$ that lifts f_t.*

Proof: This was proved as property (c) in the proof of Theorem 1.7. □

Taking Y to be a point gives the **path lifting property** for a covering space $p:\widetilde{X}\to X$, which says that for each path $f:I\to X$ and each lift $\widetilde{x}_0$ of the starting point $f(0)=x_0$ there is a unique path $\widetilde{f}:I\to\widetilde{X}$ lifting f starting at $\widetilde{x}_0$. In particular, the uniqueness of lifts implies that every lift of a constant path is constant, but this could be deduced more simply from the fact that $p^{-1}(x_0)$ has the discrete topology, by the definition of a covering space.

Taking Y to be I, we see that every homotopy f_t of a path f_0 in X lifts to a homotopy $\widetilde{f}_t$ of each lift $\widetilde{f}_0$ of f_0. The lifted homotopy $\widetilde{f}_t$ is a homotopy of paths, fixing the endpoints, since as t varies each endpoint of $\widetilde{f}_t$ traces out a path lifting a constant path, which must therefore be constant.

Here is a simple application:

Proposition 1.31. *The map $p_*:\pi_1(\widetilde{X},\widetilde{x}_0)\to\pi_1(X,x_0)$ induced by a covering space $p:(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$ is injective. The image subgroup $p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$ in $\pi_1(X,x_0)$ consists of the homotopy classes of loops in X based at x_0 whose lifts to $\widetilde{X}$ starting at $\widetilde{x}_0$ are loops.*

Proof: An element of the kernel of p_* is represented by a loop $\widetilde{f}_0:I\to\widetilde{X}$ with a homotopy $f_t:I\to X$ of $f_0=p\widetilde{f}_0$ to the trivial loop f_1. By the remarks preceding the proposition, there is a lifted homotopy of loops $\widetilde{f}_t$ starting with $\widetilde{f}_0$ and ending with a constant loop. Hence $[\widetilde{f}_0]=0$ in $\pi_1(\widetilde{X},\widetilde{x}_0)$ and p_* is injective.

For the second statement of the proposition, loops at x_0 lifting to loops at $\widetilde{x}_0$ certainly represent elements of the image of $p_*:\pi_1(\widetilde{X},\widetilde{x}_0)\to\pi_1(X,x_0)$. Conversely, a loop representing an element of the image of p_* is homotopic to a loop having such a lift, so by homotopy lifting, the loop itself must have such a lift. □

Proposition 1.32. *The number of sheets of a covering space $p:(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$ with X and $\widetilde{X}$ path-connected equals the index of $p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$ in $\pi_1(X,x_0)$.*

Proof: For a loop g in X based at x_0, let $\widetilde{g}$ be its lift to $\widetilde{X}$ starting at $\widetilde{x}_0$. A product $h\cdot g$ with $[h]\in H=p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$ has the lift $\widetilde{h}\cdot\widetilde{g}$ ending at the same point as $\widetilde{g}$ since $\widetilde{h}$ is a loop. Thus we may define a function Φ from cosets $H[g]$ to $p^{-1}(x_0)$ by sending $H[g]$ to $\widetilde{g}(1)$. The path-connectedness of $\widetilde{X}$ implies that Φ is surjective since $\widetilde{x}_0$ can be joined to any point in $p^{-1}(x_0)$ by a path $\widetilde{g}$ projecting to a loop g at x_0. To see that Φ is injective, observe that $\Phi(H[g_1])=\Phi(H[g_2])$ implies that $g_1\cdot\overline{g}_2$ lifts to a loop in $\widetilde{X}$ based at $\widetilde{x}_0$, so $[g_1][g_2]^{-1}\in H$ and hence $H[g_1]=H[g_2]$. □

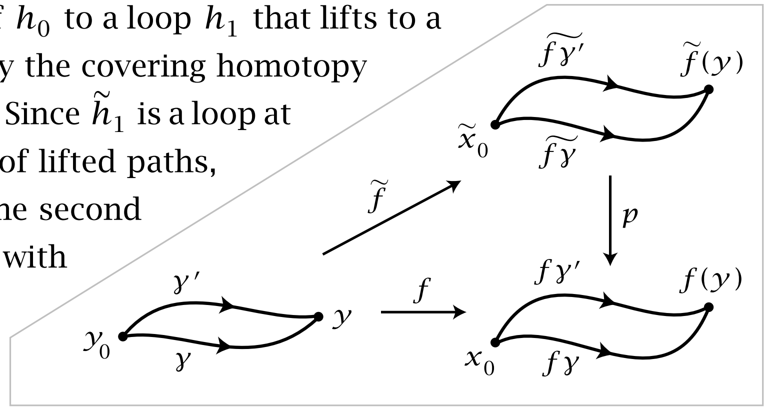
It is important also to know about the existence and uniqueness of lifts of general maps, not just lifts of homotopies. For the existence question an answer is provided by the following **lifting criterion**:

Proposition 1.33. *Suppose given a covering space $p:(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$ and a map $f:(Y,y_0)\to(X,x_0)$ with Y path-connected and locally path-connected. Then a lift $\widetilde{f}:(Y,y_0)\to(\widetilde{X},\widetilde{x}_0)$ of f exists iff $f_*(\pi_1(Y,y_0))\subset p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$.*

When we say a space has a certain property locally, such as being locally path-connected, we usually mean that each point has arbitrarily small open neighborhoods with this property. Thus for Y to be locally path-connected means that for each point

$y \in Y$ and each neighborhood U of y there is an open neighborhood $V \subset U$ of y that is path-connected.

Proof: The 'only if' statement is obvious since $f_* = p_*\widetilde{f}_*$. For the converse, let $y \in Y$ and let γ be a path in Y from y_0 to y. The path $f\gamma$ in X starting at x_0 has a unique lift $\widetilde{f\gamma}$ starting at $\widetilde{x}_0$. Define $\widetilde{f}(y) = \widetilde{f\gamma}(1)$. To show this is well-defined, independent of the choice of γ, let γ' be another path from y_0 to y. Then $(f\gamma')\cdot(\overline{f\gamma})$ is a loop h_0 at x_0 with $[h_0] \in f_*(\pi_1(Y,y_0)) \subset p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$. This means there is a homotopy h_t of h_0 to a loop h_1 that lifts to a loop $\widetilde{h}_1$ in $\widetilde{X}$ based at $\widetilde{x}_0$. Apply the covering homotopy property to h_t to get a lifting $\widetilde{h}_t$. Since $\widetilde{h}_1$ is a loop at $\widetilde{x}_0$, so is $\widetilde{h}_0$. By the uniqueness of lifted paths, the first half of $\widetilde{h}_0$ is $\widetilde{f\gamma'}$ and the second half is $\widetilde{f\gamma}$ traversed backwards, with the common midpoint $\widetilde{f\gamma}(1) = \widetilde{f\gamma'}(1)$. This shows that $\widetilde{f}$ is well-defined.



To see that $\widetilde{f}$ is continuous, let $U \subset X$ be an open neighborhood of $f(y)$ having a lift $\widetilde{U} \subset \widetilde{X}$ containing $\widetilde{f}(y)$ such that $p : \widetilde{U} \to U$ is a homeomorphism. Choose a path-connected open neighborhood V of y with $f(V) \subset U$. For paths from y_0 to points $y' \in V$ we can take a fixed path γ from y_0 to y followed by paths η in V from y to the points y'. Then the paths $(f\gamma)\cdot(f\eta)$ in X have lifts $(\widetilde{f\gamma})\cdot(\widetilde{f\eta})$ where $\widetilde{f\eta} = p^{-1}f\eta$ and $p^{-1} : U \to \widetilde{U}$ is the inverse of $p : \widetilde{U} \to U$. Thus $\widetilde{f}(V) \subset \widetilde{U}$ and $\widetilde{f}|V = p^{-1}f$, hence $\widetilde{f}$ is continuous at y. ☐

Without the local path-connectedness assumption on Y the lifting criterion in the preceding proposition can fail, as shown by an example in Exercise 7 at the end of this section.

Next we have the **unique lifting property**:

Proposition 1.34. *Given a covering space $p : \widetilde{X} \to X$ and a map $f : Y \to X$, if two lifts $\widetilde{f}_1, \widetilde{f}_2 : Y \to \widetilde{X}$ of f agree at one point of Y and Y is connected, then $\widetilde{f}_1$ and $\widetilde{f}_2$ agree on all of Y.*

Proof: For a point $y \in Y$, let U be an evenly covered open neighborhood of $f(y)$ in X, so $p^{-1}(U)$ is decomposed into disjoint sheets each mapped homeomorphically onto U by p. Let $\widetilde{U}_1$ and $\widetilde{U}_2$ be the sheets containing $\widetilde{f}_1(y)$ and $\widetilde{f}_2(y)$, respectively. By continuity of $\widetilde{f}_1$ and $\widetilde{f}_2$ there is a neighborhood N of y mapped into $\widetilde{U}_1$ by $\widetilde{f}_1$ and into $\widetilde{U}_2$ by $\widetilde{f}_2$. If $\widetilde{f}_1(y) \neq \widetilde{f}_2(y)$ then $\widetilde{U}_1 \neq \widetilde{U}_2$, hence $\widetilde{U}_1$ and $\widetilde{U}_2$ are disjoint and $\widetilde{f}_1 \neq \widetilde{f}_2$ throughout the neighborhood N. On the other hand, if $\widetilde{f}_1(y) = \widetilde{f}_2(y)$ then $\widetilde{U}_1 = \widetilde{U}_2$ so $\widetilde{f}_1 = \widetilde{f}_2$ on N since $p\widetilde{f}_1 = p\widetilde{f}_2$ and p is injective on $\widetilde{U}_1 = \widetilde{U}_2$. Thus the set of points where $\widetilde{f}_1$ and $\widetilde{f}_2$ agree is both open and closed in Y. ☐

The Classification of Covering Spaces

We consider next the problem of classifying all the different covering spaces of a fixed space X. Since the whole chapter is about paths, it should not be surprising that we will restrict attention to spaces X that are at least locally path-connected. Path-components of X are then the same as components, and for the purpose of classifying the covering spaces of X there is no loss in assuming that X is connected, or equivalently, path-connected. Local path-connectedness is inherited by covering spaces, so connected covering spaces of X are the same as path-connected covering spaces. The main thrust of the classification will be a correspondence between connected covering spaces of X and subgroups of $\pi_1(X)$. This is often called the Galois correspondence because of its surprising similarity to another basic correspondence in the purely algebraic subject of Galois theory. We will also describe a different method of classification that includes disconnected covering spaces as well.

The Galois correspondence arises from the function that assigns to each covering space $p:(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$ the subgroup $p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$ of $\pi_1(X,x_0)$. First we consider whether this function is surjective. That is, we ask whether every subgroup of $\pi_1(X,x_0)$ is realized as $p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$ for some covering space $p:(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$. In particular we can ask whether the trivial subgroup is realized. Since p_* is always injective, this amounts to asking whether X has a simply-connected covering space. Answering this will take some work.

A necessary condition for X to have a simply-connected covering space is the following: Each point $x\in X$ has a neighborhood U such that the inclusion-induced map $\pi_1(U,x)\to\pi_1(X,x)$ is trivial; one says X is **semilocally simply-connected** if this holds. To see the necessity of this condition, suppose $p:\widetilde{X}\to X$ is a covering space with $\widetilde{X}$ simply-connected. Every point $x\in X$ has a neighborhood U having a lift $\widetilde{U}\subset\widetilde{X}$ projecting homeomorphically to U by p. Each loop in U lifts to a loop in $\widetilde{U}$, and the lifted loop is nullhomotopic in $\widetilde{X}$ since $\pi_1(\widetilde{X})=0$. So, composing this nullhomotopy with p, the original loop in U is nullhomotopic in X.

A locally simply-connected space is certainly semilocally simply-connected. For example, CW complexes have the much stronger property of being locally contractible, as we show in the Appendix. An example of a space that is not semilocally simply-connected is the shrinking wedge of circles, the subspace $X\subset\mathbb{R}^2$ consisting of the circles of radius $1/n$ centered at the point $(1/n,0)$ for $n=1,2,\cdots$, introduced in Example 1.25. On the other hand, if we take the cone $CX=(X\times I)/(X\times\{0\})$ on the shrinking wedge of circles, this is semilocally simply-connected since it is contractible, but it is not locally simply-connected.

We shall now show how to construct a simply-connected covering space of X if X is path-connected, locally path-connected, and semilocally simply-connected. To motivate the construction, suppose $p:(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$ is a simply-connected covering space. Each point $\widetilde{x}\in\widetilde{X}$ can then be joined to $\widetilde{x}_0$ by a unique homotopy class of paths, by Proposition 1.6, so we can view points of $\widetilde{X}$ as homotopy classes of paths starting at $\widetilde{x}_0$. The advantage of this is that, by the homotopy lifting property, homotopy classes of paths in $\widetilde{X}$ starting at $\widetilde{x}_0$ are the same as homotopy classes of paths in X starting at x_0. This gives a way of describing $\widetilde{X}$ purely in terms of X.

Given a path-connected, locally path-connected, semilocally simply-connected space X with a basepoint $x_0 \in X$, we are therefore led to define

$$\widetilde{X} = \{\, [\gamma] \mid \gamma \text{ is a path in } X \text{ starting at } x_0 \,\}$$

where, as usual, $[\gamma]$ denotes the homotopy class of γ with respect to homotopies that fix the endpoints $\gamma(0)$ and $\gamma(1)$. The function $p:\widetilde{X} \to X$ sending $[\gamma]$ to $\gamma(1)$ is then well-defined. Since X is path-connected, the endpoint $\gamma(1)$ can be any point of X, so p is surjective.

Before we define a topology on $\widetilde{X}$ we make a few preliminary observations. Let $\mathcal{U}$ be the collection of path-connected open sets $U \subset X$ such that $\pi_1(U) \to \pi_1(X)$ is trivial. Note that if the map $\pi_1(U) \to \pi_1(X)$ is trivial for one choice of basepoint in U, it is trivial for all choices of basepoint since U is path-connected. A path-connected open subset $V \subset U \in \mathcal{U}$ is also in $\mathcal{U}$ since the composition $\pi_1(V) \to \pi_1(U) \to \pi_1(X)$ will also be trivial. It follows that $\mathcal{U}$ is a basis for the topology on X if X is locally path-connected and semilocally simply-connected.

Given a set $U \in \mathcal{U}$ and a path γ in X from x_0 to a point in U, let

$$U_{[\gamma]} = \{\, [\gamma \cdot \eta] \mid \eta \text{ is a path in } U \text{ with } \eta(0) = \gamma(1) \,\}$$

As the notation indicates, $U_{[\gamma]}$ depends only on the homotopy class $[\gamma]$. Observe that $p:U_{[\gamma]} \to U$ is surjective since U is path-connected and injective since different choices of η joining $\gamma(1)$ to a fixed $x \in U$ are all homotopic in X, the map $\pi_1(U) \to \pi_1(X)$ being trivial. Another property is

$(*)$ $U_{[\gamma]} = U_{[\gamma']}$ if $[\gamma'] \in U_{[\gamma]}$. For if $\gamma' = \gamma \cdot \eta$ then elements of $U_{[\gamma']}$ have the form $[\gamma \cdot \eta \cdot \mu]$ and hence lie in $U_{[\gamma]}$, while elements of $U_{[\gamma]}$ have the form $[\gamma \cdot \mu] = [\gamma \cdot \eta \cdot \overline{\eta} \cdot \mu] = [\gamma' \cdot \overline{\eta} \cdot \mu]$ and hence lie in $U_{[\gamma']}$.

This can be used to show that the sets $U_{[\gamma]}$ form a basis for a topology on $\widetilde{X}$. For if we are given two such sets $U_{[\gamma]}$, $V_{[\gamma']}$ and an element $[\gamma''] \in U_{[\gamma]} \cap V_{[\gamma']}$, we have $U_{[\gamma]} = U_{[\gamma'']}$ and $V_{[\gamma']} = V_{[\gamma'']}$ by $(*)$. So if $W \in \mathcal{U}$ is contained in $U \cap V$ and contains $\gamma''(1)$ then $W_{[\gamma'']} \subset U_{[\gamma'']} \cap V_{[\gamma'']}$ and $[\gamma''] \in W_{[\gamma'']}$.

The bijection $p:U_{[\gamma]} \to U$ is a homeomorphism since it gives a bijection between the subsets $V_{[\gamma']} \subset U_{[\gamma]}$ and the sets $V \in \mathcal{U}$ contained in U. Namely, in one direction we have $p(V_{[\gamma']}) = V$ and in the other direction we have $p^{-1}(V) \cap U_{[\gamma]} = V_{[\gamma']}$ for any $[\gamma'] \in U_{[\gamma]}$ with endpoint in V, since $V_{[\gamma']} \subset U_{[\gamma']} = U_{[\gamma]}$ and $V_{[\gamma']}$ maps onto V by the bijection p.

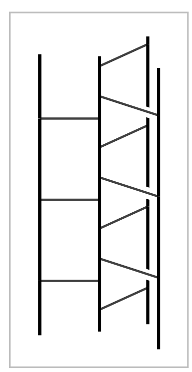
The preceding paragraph implies that $p:\widetilde{X} \to X$ is continuous. We can also deduce that this is a covering space since for fixed $U \in \mathcal{U}$, the sets $U_{[\gamma]}$ for varying $[\gamma]$ partition $p^{-1}(U)$ because if $[\gamma''] \in U_{[\gamma]} \cap U_{[\gamma']}$ then $U_{[\gamma]} = U_{[\gamma'']} = U_{[\gamma']}$ by $(*)$.

It remains only to show that $\widetilde{X}$ is simply-connected. For a point $[\gamma] \in \widetilde{X}$ let γ_t be the path in X that equals γ on $[0,t]$ and is stationary at $\gamma(t)$ on $[t,1]$. Then the function $t \mapsto [\gamma_t]$ is a path in $\widetilde{X}$ lifting γ that starts at $[x_0]$, the homotopy class of the constant path at x_0, and ends at $[\gamma]$. Since $[\gamma]$ was an arbitrary point in $\widetilde{X}$, this shows that $\widetilde{X}$ is path-connected. To show that $\pi_1(\widetilde{X},[x_0]) = 0$ it suffices to show that the image of this group under p_* is trivial since p_* is injective. Elements in the image of p_* are represented by loops γ at x_0 that lift to loops in $\widetilde{X}$ at $[x_0]$. We have observed that the path $t \mapsto [\gamma_t]$ lifts γ starting at $[x_0]$, and for this lifted path to be a loop means that $[\gamma_1] = [x_0]$. Since $\gamma_1 = \gamma$, this says that $[\gamma] = [x_0]$, so γ is nullhomotopic and the image of p_* is trivial.

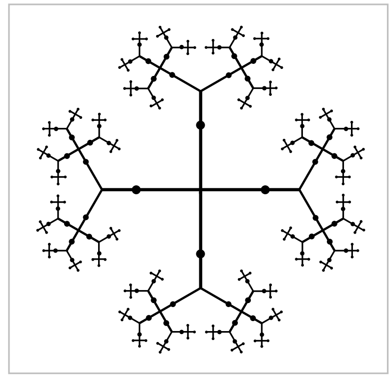
This completes the construction of a simply-connected covering space $\widetilde{X} \to X$.

In concrete cases one usually constructs a simply-connected covering space by more direct methods. For example, suppose X is the union of subspaces A and B for which simply-connected covering spaces $\widetilde{A} \to A$ and $\widetilde{B} \to B$ are already known. Then one can attempt to build a simply-connected covering space $\widetilde{X} \to X$ by assembling copies of $\widetilde{A}$ and $\widetilde{B}$. For example, for $X = S^1 \vee S^1$, if we take A and B to be the two circles, then $\widetilde{A}$ and $\widetilde{B}$ are each $\mathbb{R}$, and we can build the simply-connected cover $\widetilde{X}$ described earlier in this section by glueing together infinitely many copies of $\widetilde{A}$ and $\widetilde{B}$, the horizontal and vertical lines in $\widetilde{X}$. Here is another illustration of this method:

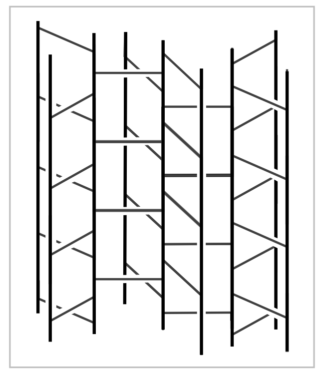
Example 1.35. For integers $m,n \geq 2$, let $X_{m,n}$ be the quotient space of a cylinder $S^1 \times I$ under the identifications $(z,0) \sim (e^{2\pi i/m}z,0)$ and $(z,1) \sim (e^{2\pi i/n}z,1)$. Let $A \subset X$ and $B \subset X$ be the quotients of $S^1 \times [0,1/2]$ and $S^1 \times [1/2,1]$, so A and B are the mapping cylinders of $z \mapsto z^m$ and $z \mapsto z^n$, with $A \cap B = S^1$. The simplest case is $m = n = 2$, when A and B are Möbius bands and $X_{2,2}$ is the Klein bottle. We encountered the complexes $X_{m,n}$ previously in analyzing torus knot complements in Example 1.24.



The figure for Example 1.29 at the end of the preceding section shows what A looks like in the typical case $m = 3$. We have $\pi_1(A) \approx \mathbb{Z}$, and the universal cover $\widetilde{A}$ is homeomorphic to a product $C_m \times \mathbb{R}$ where C_m is the graph that is a cone on m points, as shown in the figure to the right. The situation for B is similar, and $\widetilde{B}$ is homeomorphic to $C_n \times \mathbb{R}$. Now we attempt to build the universal cover $\widetilde{X}_{m,n}$ from copies of $\widetilde{A}$ and $\widetilde{B}$. Start with a copy of $\widetilde{A}$. Its boundary, the outer edges of its fins, consists of m copies of $\mathbb{R}$. Along each of these m boundary lines we attach a copy of $\widetilde{B}$. Each of these copies of $\widetilde{B}$ has one of its boundary lines attached to the initial copy of $\widetilde{A}$, leaving $n-1$ boundary lines free, and we attach a new copy of $\widetilde{A}$ to each of these free boundary lines. Thus we now have $m(n-1)+1$ copies of $\widetilde{A}$. Each of the newly attached copies of $\widetilde{A}$ has $m-1$ free boundary lines, and to each of these lines we attach a new copy of $\widetilde{B}$. The process is now repeated ad infinitum in the evident way. Let $\widetilde{X}_{m,n}$ be the resulting space.



The product structures $\widetilde{A} = C_m \times \mathbb{R}$ and $\widetilde{B} = C_n \times \mathbb{R}$ give $\widetilde{X}_{m,n}$ the structure of a product $T_{m,n} \times \mathbb{R}$ where $T_{m,n}$ is an infinite graph constructed by an inductive scheme just like the construction of $\widetilde{X}_{m,n}$. Thus $T_{m,n}$ is the union of a sequence of finite subgraphs, each obtained from the preceding by attaching new copies of C_m or C_n. Each of these finite subgraphs deformation retracts onto the preceding one. The infinite concatenation of these deformation retractions, with the k^{th} graph deformation retracting to the previous one during the time interval $[1/2^k, 1/2^{k-1}]$, gives a deformation retraction of $T_{m,n}$ onto the initial stage C_m. Since C_m is contractible, this means $T_{m,n}$ is contractible, hence also $\widetilde{X}_{m,n}$, which is the product $T_{m,n} \times \mathbb{R}$. In particular, $\widetilde{X}_{m,n}$ is simply-connected.



The map that projects each copy of $\widetilde{A}$ in $\widetilde{X}_{m,n}$ to A and each copy of $\widetilde{B}$ to B is a covering space. To define this map precisely, choose a point $x_0 \in S^1$, and then the image of the line segment $\{x_0\} \times I$ in $X_{m,n}$ meets A in a line segment whose preimage in $\widetilde{A}$ consists of an infinite number of line segments, appearing in the earlier figure as the horizontal segments spiraling around the central vertical axis. The picture in $\widetilde{B}$ is similar, and when we glue together all the copies of $\widetilde{A}$ and $\widetilde{B}$ to form $\widetilde{X}_{m,n}$, we do so in such a way that these horizontal segments always line up exactly. This decomposes $\widetilde{X}_{m,n}$ into infinitely many rectangles, each formed from a rectangle in an $\widetilde{A}$ and a rectangle in a $\widetilde{B}$. The covering projection $\widetilde{X}_{m,n} \to X_{m,n}$ is the quotient map that identifies all these rectangles.

Now we return to the general theory. The hypotheses for constructing a simply-connected covering space of X in fact suffice for constructing covering spaces realizing arbitrary subgroups of $\pi_1(X)$:

Proposition 1.36. *Suppose X is path-connected, locally path-connected, and semilocally simply-connected. Then for every subgroup $H \subset \pi_1(X, x_0)$ there is a covering space $p: X_H \to X$ such that $p_*(\pi_1(X_H, \widetilde{x}_0)) = H$ for a suitably chosen basepoint $\widetilde{x}_0 \in X_H$.*

Proof: For points $[\gamma]$, $[\gamma']$ in the simply-connected covering space $\widetilde{X}$ constructed above, define $[\gamma] \sim [\gamma']$ to mean $\gamma(1) = \gamma'(1)$ and $[\gamma \cdot \overline{\gamma'}] \in H$. It is easy to see that this is an equivalence relation since H is a subgroup: it is reflexive since H contains the identity element, symmetric since H is closed under inverses, and transitive since H is closed under multiplication. Let X_H be the quotient space of $\widetilde{X}$ obtained by identifying $[\gamma]$ with $[\gamma']$ if $[\gamma] \sim [\gamma']$. Note that if $\gamma(1) = \gamma'(1)$, then $[\gamma] \sim [\gamma']$ iff $[\gamma \cdot \eta] \sim [\gamma' \cdot \eta]$. This means that if any two points in basic neighborhoods $U_{[\gamma]}$

and $U_{[\gamma']}$ are identified in X_H then the whole neighborhoods are identified. Hence the natural projection $X_H \to X$ induced by $[\gamma] \mapsto \gamma(1)$ is a covering space.

If we choose for the basepoint $\widetilde{x}_0 \in X_H$ the equivalence class of the constant path c at x_0, then the image of $p_*:\pi_1(X_H,\widetilde{x}_0)\to\pi_1(X,x_0)$ is exactly H. This is because for a loop γ in X based at x_0, its lift to $\widetilde{X}$ starting at $[c]$ ends at $[\gamma]$, so the image of this lifted path in X_H is a loop iff $[\gamma] \sim [c]$, or equivalently, $[\gamma] \in H$. $\square$

Having taken care of the existence of covering spaces of X corresponding to all subgroups of $\pi_1(X)$, we turn now to the question of uniqueness. More specifically, we are interested in uniqueness up to isomorphism, where an **isomorphism** between covering spaces $p_1:\widetilde{X}_1\to X$ and $p_2:\widetilde{X}_2\to X$ is a homeomorphism $f:\widetilde{X}_1\to\widetilde{X}_2$ such that $p_1 = p_2 f$. This condition means exactly that f preserves the covering space structures, taking $p_1^{-1}(x)$ to $p_2^{-1}(x)$ for each $x \in X$. The inverse f^{-1} is then also an isomorphism, and the composition of two isomorphisms is an isomorphism, so we have an equivalence relation.

Proposition 1.37. *If X is path-connected and locally path-connected, then two path-connected covering spaces $p_1:\widetilde{X}_1\to X$ and $p_2:\widetilde{X}_2\to X$ are isomorphic via an isomorphism $f:\widetilde{X}_1\to\widetilde{X}_2$ taking a basepoint $\widetilde{x}_1 \in p_1^{-1}(x_0)$ to a basepoint $\widetilde{x}_2 \in p_2^{-1}(x_0)$ iff $p_{1*}(\pi_1(\widetilde{X}_1,\widetilde{x}_1)) = p_{2*}(\pi_1(\widetilde{X}_2,\widetilde{x}_2))$.*

Proof: If there is an isomorphism $f:(\widetilde{X}_1,\widetilde{x}_1)\to(\widetilde{X}_2,\widetilde{x}_2)$, then from the two relations $p_1 = p_2 f$ and $p_2 = p_1 f^{-1}$ it follows that $p_{1*}(\pi_1(\widetilde{X}_1,\widetilde{x}_1)) = p_{2*}(\pi_1(\widetilde{X}_2,\widetilde{x}_2))$. Conversely, suppose that $p_{1*}(\pi_1(\widetilde{X}_1,\widetilde{x}_1)) = p_{2*}(\pi_1(\widetilde{X}_2,\widetilde{x}_2))$. By the lifting criterion, we may lift p_1 to a map $\widetilde{p}_1:(\widetilde{X}_1,\widetilde{x}_1)\to(\widetilde{X}_2,\widetilde{x}_2)$ with $p_2\widetilde{p}_1 = p_1$. Symmetrically, we obtain $\widetilde{p}_2:(\widetilde{X}_2,\widetilde{x}_2)\to(\widetilde{X}_1,\widetilde{x}_1)$ with $p_1\widetilde{p}_2 = p_2$. Then by the unique lifting property, $\widetilde{p}_1\widetilde{p}_2 = \mathbb{1}$ and $\widetilde{p}_2\widetilde{p}_1 = \mathbb{1}$ since these composed lifts fix the basepoints. Thus $\widetilde{p}_1$ and $\widetilde{p}_2$ are inverse isomorphisms. $\square$

We have proved the first half of the following classification theorem:

Theorem 1.38. *Let X be path-connected, locally path-connected, and semilocally simply-connected. Then there is a bijection between the set of basepoint-preserving isomorphism classes of path-connected covering spaces $p:(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$ and the set of subgroups of $\pi_1(X,x_0)$, obtained by associating the subgroup $p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$ to the covering space $(\widetilde{X},\widetilde{x}_0)$. If basepoints are ignored, this correspondence gives a bijection between isomorphism classes of path-connected covering spaces $p:\widetilde{X}\to X$ and conjugacy classes of subgroups of $\pi_1(X,x_0)$.*

Proof: It remains only to prove the last statement. We show that for a covering space $p:(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$, changing the basepoint $\widetilde{x}_0$ within $p^{-1}(x_0)$ corresponds exactly to changing $p_*(\pi_1(\widetilde{X},\widetilde{x}_0))$ to a conjugate subgroup of $\pi_1(X,x_0)$. Suppose that $\widetilde{x}_1$ is another basepoint in $p^{-1}(x_0)$, and let $\widetilde{\gamma}$ be a path from $\widetilde{x}_0$ to $\widetilde{x}_1$. Then $\widetilde{\gamma}$ projects to a loop γ in X representing some element $g \in \pi_1(X,x_0)$. Set $H_i = p_*(\pi_1(\widetilde{X},\widetilde{x}_i))$ for $i = 0,1$. We have an inclusion $g^{-1}H_0 g \subset H_1$ since for $\widetilde{f}$ a loop at $\widetilde{x}_0$, $\overline{\widetilde{\gamma}} \cdot \widetilde{f} \cdot \widetilde{\gamma}$ is a loop at $\widetilde{x}_1$. Similarly we have $gH_1g^{-1} \subset H_0$. Conjugating the latter relation by g^{-1} gives $H_1 \subset g^{-1}H_0 g$, so $g^{-1}H_0 g = H_1$. Thus, changing the basepoint from $\widetilde{x}_0$ to $\widetilde{x}_1$ changes H_0 to the conjugate subgroup $H_1 = g^{-1}H_0 g$.

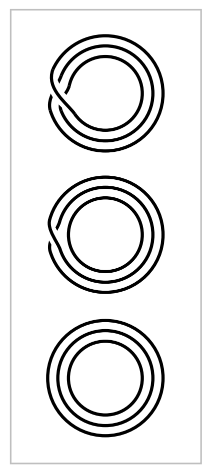
Conversely, to change H_0 to a conjugate subgroup $H_1 = g^{-1}H_0 g$, choose a loop γ representing g, lift this to a path $\widetilde{\gamma}$ starting at $\widetilde{x}_0$, and let $\widetilde{x}_1 = \widetilde{\gamma}(1)$. The preceding argument then shows that we have the desired relation $H_1 = g^{-1}H_0 g$. □

A consequence of the lifting criterion is that a simply-connected covering space of a path-connected, locally path-connected space X is a covering space of every other path-connected covering space of X. A simply-connected covering space of X is therefore called a **universal cover**. It is unique up to isomorphism, so one is justified in calling it *the* universal cover.

More generally, there is a partial ordering on the various path-connected covering spaces of X, according to which ones cover which others. This corresponds to the partial ordering by inclusion of the corresponding subgroups of $\pi_1(X)$, or conjugacy classes of subgroups if basepoints are ignored.

Representing Covering Spaces by Permutations

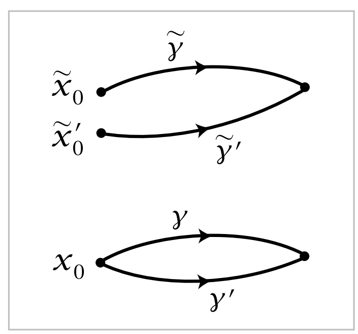
We wish to describe now another way of classifying the different covering spaces of a connected, locally path-connected, semilocally simply-connected space X, without restricting just to connected covering spaces. To give the idea, consider the 3-sheeted covering spaces of S^1. There are three of these, $\widetilde{X}_1$, $\widetilde{X}_2$, and $\widetilde{X}_3$, with the subscript indicating the number of components. For each of these covering spaces $p : \widetilde{X}_i \to S^1$ the three different lifts of a loop in S^1 generating $\pi_1(S^1,x_0)$ determine a permutation of $p^{-1}(x_0)$ sending the starting point of the lift to the ending point of the lift. For $\widetilde{X}_1$ this is a cyclic permutation, for $\widetilde{X}_2$ it is a transposition of two points fixing the third point, and for $\widetilde{X}_3$ it is the identity permutation. These permutations obviously determine the covering spaces uniquely, up to isomorphism. The same would be true for n-sheeted covering spaces of S^1 for arbitrary n, even for n infinite.



The covering spaces of $S^1 \vee S^1$ can be encoded using the same idea. Referring back to the large table of examples near the beginning of this section, we see in the covering space (1) that the loop a lifts to the identity permutation of the two vertices and b lifts to the permutation that transposes the two vertices. In (2), both a and b lift to transpositions of the two vertices. In (3) and (4), a and b lift to transpositions of different pairs of the three vertices, while in (5) and (6) they lift to cyclic permutations of the vertices. In (11) the vertices can be labeled by $\mathbb{Z}$, with a lifting to the identity permutation and b lifting to the shift $n \mapsto n+1$. Indeed, one can see from these examples that a covering space of $S^1 \vee S^1$ is nothing more than an efficient graphical representation of a pair of permutations of a given set.

This idea of lifting loops to permutations generalizes to arbitrary covering spaces. For a covering space $p:\widetilde{X}\to X$, a path γ in X has a unique lift $\widetilde{\gamma}$ starting at a given point of $p^{-1}(\gamma(0))$, so we obtain a well-defined map $L_\gamma : p^{-1}(\gamma(0))\to p^{-1}(\gamma(1))$ by sending the starting point $\widetilde{\gamma}(0)$ of each lift $\widetilde{\gamma}$ to its ending point $\widetilde{\gamma}(1)$. It is evident that L_γ is a bijection since $L_{\overline{\gamma}}$ is its inverse. For a composition of paths $\gamma\cdot\eta$ we have $L_{\gamma\cdot\eta} = L_\eta L_\gamma$, rather than $L_\gamma L_\eta$, since composition of paths is written from left to right while composition of functions is written from right to left. To compensate for this, let us modify the definition by replacing L_γ by its inverse. Thus the new L_γ is a bijection $p^{-1}(\gamma(1))\to p^{-1}(\gamma(0))$, and $L_{\gamma\cdot\eta} = L_\gamma L_\eta$. Since L_γ depends only on the homotopy class of γ, this means that if we restrict attention to loops at a basepoint $x_0\in X$, then the association $\gamma\mapsto L_\gamma$ gives a homomorphism from $\pi_1(X,x_0)$ to the group of permutations of $p^{-1}(x_0)$. This is called *the action of* $\pi_1(X,x_0)$ *on the fiber* $p^{-1}(x_0)$.

Let us see how the covering space $p:\widetilde{X}\to X$ can be reconstructed from the associated action of $\pi_1(X,x_0)$ on the fiber $F = p^{-1}(x_0)$, assuming that X is path-connected, locally path-connected, and semilocally simply-connected, so it has a universal cover $\widetilde{X}_0\to X$. We can take the points of $\widetilde{X}_0$ to be homotopy classes of paths in X starting at x_0, as in the general construction of a universal cover. Define a map $h:\widetilde{X}_0\times F\to\widetilde{X}$ sending a pair $([\gamma],\widetilde{x}_0)$ to $\widetilde{\gamma}(1)$ where $\widetilde{\gamma}$ is the lift of γ to $\widetilde{X}$ starting at $\widetilde{x}_0$. Then h is continuous, and in fact a local homeomorphism, since a neighborhood of $([\gamma],\widetilde{x}_0)$ in $\widetilde{X}_0\times F$ consists of the pairs $([\gamma\cdot\eta],\widetilde{x}_0)$ with η a path in a suitable neighborhood of $\gamma(1)$. It is obvious that h is surjective since X is path-connected. If h were injective as well, it would be a homeomorphism, which is unlikely since $\widetilde{X}$ is probably not homeomorphic to $\widetilde{X}_0\times F$. Even if h is not injective, it will induce a homeomorphism from some quotient space of $\widetilde{X}_0\times F$ onto $\widetilde{X}$. To see what this quotient space is, suppose $h([\gamma],\widetilde{x}_0) = h([\gamma'],\widetilde{x}_0')$. Then γ and γ' are both paths from x_0 to the same endpoint, and from the figure we see that $\widetilde{x}_0' = L_{\gamma'\cdot\overline{\gamma}}(\widetilde{x}_0)$. Letting λ be the loop $\gamma'\cdot\overline{\gamma}$, this means that $h([\gamma],\widetilde{x}_0) = h([\lambda\cdot\gamma],L_\lambda(\widetilde{x}_0))$. Conversely, for any loop λ we have $h([\gamma],\widetilde{x}_0) = h([\lambda\cdot\gamma],L_\lambda(\widetilde{x}_0))$. Thus h induces a well-defined map to $\widetilde{X}$ from the quotient space of $\widetilde{X}_0\times F$ obtained by identifying $([\gamma],\widetilde{x}_0)$ with $([\lambda\cdot\gamma],L_\lambda(\widetilde{x}_0))$ for each $[\lambda]\in\pi_1(X,x_0)$. Let this quotient space be denoted $\widetilde{X}_\rho$ where ρ is the homomorphism from $\pi_1(X,x_0)$ to the permutation group of F specified by the action.



Notice that the definition of $\widetilde{X}_\rho$ makes sense whenever we are given an action ρ of $\pi_1(X,x_0)$ on a set F regarded as a space with the discrete topology. There is a natural projection $\widetilde{X}_\rho\to X$ sending $([\gamma],\widetilde{x}_0)$ to $\gamma(1)$, and this is a covering space since if $U\subset X$ is an open set over which the universal cover $\widetilde{X}_0$ is a product $U\times\pi_1(X,x_0)$, then the identifications defining $\widetilde{X}_\rho$ simply collapse $U \times \pi_1(X, x_0) \times F$ to $U \times F$.

Returning to our given covering space $\widetilde{X} \to X$ with associated action ρ, the map $\widetilde{X}_\rho \to \widetilde{X}$ induced by h is a bijection and therefore a homeomorphism since h was a local homeomorphism. Since this homeomorphism $\widetilde{X}_\rho \to \widetilde{X}$ takes each fiber of $\widetilde{X}_\rho$ to the corresponding fiber of $\widetilde{X}$, it is an isomorphism of covering spaces.

If two covering spaces $p_1 : \widetilde{X}_1 \to X$ and $p_2 : \widetilde{X}_2 \to X$ are isomorphic, one may ask how the corresponding actions of $\pi_1(X, x_0)$ on the fibers F_1 and F_2 over x_0 are related. An isomorphism $h : \widetilde{X}_1 \to \widetilde{X}_2$ restricts to a bijection $F_1 \to F_2$, and evidently $L_\gamma(h(\tilde{x}_0)) = h(L_\gamma(\tilde{x}_0))$. Using the less cumbersome notation $\gamma\tilde{x}_0$ for $L_\gamma(\tilde{x}_0)$, this relation can be written more concisely as $\gamma h(\tilde{x}_0) = h(\gamma\tilde{x}_0)$. A bijection $F_1 \to F_2$ with this property is what one would naturally call an *isomorphism of sets with $\pi_1(X, x_0)$ action*. Thus isomorphic covering spaces have isomorphic actions on fibers. The converse is also true, and easy to prove. One just observes that for isomorphic actions ρ_1 and ρ_2, an isomorphism $h : F_1 \to F_2$ induces a map $\widetilde{X}_{\rho_1} \to \widetilde{X}_{\rho_2}$ and h^{-1} induces a similar map in the opposite direction, such that the compositions of these two maps, in either order, are the identity.

This shows that n-sheeted covering spaces of X are classified by equivalence classes of homomorphisms $\pi_1(X, x_0) \to \Sigma_n$, where Σ_n is the symmetric group on n symbols and the equivalence relation identifies a homomorphism ρ with each of its conjugates $h^{-1}\rho h$ by elements $h \in \Sigma_n$. The study of the various homomorphisms from a given group to Σ_n is a very classical topic in group theory, so we see that this algebraic question has a nice geometric interpretation.

Deck Transformations and Group Actions

For a covering space $p : \widetilde{X} \to X$ the isomorphisms $\widetilde{X} \to \widetilde{X}$ are called **deck transformations** or **covering transformations**. These form a group $G(\widetilde{X})$ under composition. For example, for the covering space $p : \mathbb{R} \to S^1$ projecting a vertical helix onto a circle, the deck transformations are the vertical translations taking the helix onto itself, so $G(\widetilde{X}) \approx \mathbb{Z}$ in this case. For the n-sheeted covering space $S^1 \to S^1$, $z \mapsto z^n$, the deck transformations are the rotations of S^1 through angles that are multiples of $2\pi/n$, so $G(\widetilde{X}) = \mathbb{Z}_n$.

By the unique lifting property, a deck transformation is completely determined by where it sends a single point, assuming $\widetilde{X}$ is path-connected. In particular, only the identity deck transformation can fix a point of $\widetilde{X}$.

A covering space $p : \widetilde{X} \to X$ is called **normal** if for each $x \in X$ and each pair of lifts $\tilde{x}, \tilde{x}'$ of x there is a deck transformation taking $\tilde{x}$ to $\tilde{x}'$. For example, the covering space $\mathbb{R} \to S^1$ and the n-sheeted covering spaces $S^1 \to S^1$ are normal. Intuitively, a normal covering space is one with maximal symmetry. This can be seen in the covering spaces of $S^1 \vee S^1$ shown in the table earlier in this section, where the normal covering spaces are (1), (2), (5)–(8), and (11). Note that in (7) the group of deck transformations is $\mathbb{Z}_4$ while in (8) it is $\mathbb{Z}_2 \times \mathbb{Z}_2$.

Sometimes normal covering spaces are called regular covering spaces. The term ‘normal’ is motivated by the following result.

Proposition 1.39. *Let $p:(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$ be a path-connected covering space of the path-connected, locally path-connected space X, and let H be the subgroup $p_*(\pi_1(\widetilde{X},\widetilde{x}_0)) \subset \pi_1(X,x_0)$. Then:*

(a) *This covering space is normal iff H is a normal subgroup of $\pi_1(X,x_0)$.*

(b) *$G(\widetilde{X})$ is isomorphic to the quotient $N(H)/H$ where $N(H)$ is the normalizer of H in $\pi_1(X,x_0)$.*

In particular, $G(\widetilde{X})$ is isomorphic to $\pi_1(X,x_0)/H$ if $\widetilde{X}$ is a normal covering. Hence for the universal cover $\widetilde{X}\to X$ we have $G(\widetilde{X}) \approx \pi_1(X)$.

Proof: We observed earlier in the proof of the classification theorem that changing the basepoint $\widetilde{x}_0 \in p^{-1}(x_0)$ to $\widetilde{x}_1 \in p^{-1}(x_0)$ corresponds precisely to conjugating H by an element $[\gamma] \in \pi_1(X,x_0)$ where γ lifts to a path $\widetilde{\gamma}$ from $\widetilde{x}_0$ to $\widetilde{x}_1$. Thus $[\gamma]$ is in the normalizer $N(H)$ iff $p_*(\pi_1(\widetilde{X},\widetilde{x}_0)) = p_*(\pi_1(\widetilde{X},\widetilde{x}_1))$, which by the lifting criterion is equivalent to the existence of a deck transformation taking $\widetilde{x}_0$ to $\widetilde{x}_1$. Hence the covering space is normal iff $N(H) = \pi_1(X,x_0)$, that is, iff H is a normal subgroup of $\pi_1(X,x_0)$.

Define $\varphi: N(H) \to G(\widetilde{X})$ sending $[\gamma]$ to the deck transformation τ taking $\widetilde{x}_0$ to $\widetilde{x}_1$, in the notation above. Then φ is a homomorphism, for if γ' is another loop corresponding to the deck transformation τ' taking $\widetilde{x}_0$ to $\widetilde{x}_1'$ then $\gamma \cdot \gamma'$ lifts to $\widetilde{\gamma} \cdot (\tau(\widetilde{\gamma}'))$, a path from $\widetilde{x}_0$ to $\tau(\widetilde{x}_1') = \tau\tau'(\widetilde{x}_0)$, so $\tau\tau'$ is the deck transformation corresponding to $[\gamma][\gamma']$. By the preceding paragraph φ is surjective. Its kernel consists of classes $[\gamma]$ lifting to loops in $\widetilde{X}$. These are exactly the elements of $p_*(\pi_1(\widetilde{X},\widetilde{x}_0)) = H$. ☐

The group of deck transformations is a special case of the general notion of ‘groups acting on spaces’. Given a group G and a space Y, then an **action** of G on Y is a homomorphism ρ from G to the group $\mathrm{Homeo}(Y)$ of all homeomorphisms from Y to itself. Thus to each $g \in G$ is associated a homeomorphism $\rho(g): Y \to Y$, which for notational simplicity we write simply as $g: Y \to Y$. For ρ to be a homomorphism amounts to requiring that $g_1(g_2(y)) = (g_1 g_2)(y)$ for all $g_1, g_2 \in G$ and $y \in Y$. If ρ is injective then it identifies G with a subgroup of $\mathrm{Homeo}(Y)$, and in practice not much is lost in assuming ρ is an inclusion $G \hookrightarrow \mathrm{Homeo}(Y)$ since in any case the subgroup $\rho(G) \subset \mathrm{Homeo}(Y)$ contains all the topological information about the action.

We shall be interested in actions satisfying the following condition:

(∗) Each $y \in Y$ has a neighborhood U such that all the images $g(U)$ for varying $g \in G$ are disjoint. In other words, $g_1(U) \cap g_2(U) \neq \varnothing$ implies $g_1 = g_2$.

The action of the deck transformation group $G(\widetilde{X})$ on $\widetilde{X}$ satisfies $(*)$. To see this, let $\widetilde{U} \subset \widetilde{X}$ project homeomorphically to $U \subset X$. If $g_1(\widetilde{U}) \cap g_2(\widetilde{U}) \neq \varnothing$ for some $g_1, g_2 \in G(\widetilde{X})$, then $g_1(\tilde{x}_1) = g_2(\tilde{x}_2)$ for some $\tilde{x}_1, \tilde{x}_2 \in \widetilde{U}$. Since $\tilde{x}_1$ and $\tilde{x}_2$ must lie in the same set $p^{-1}(x)$, which intersects $\widetilde{U}$ in only one point, we must have $\tilde{x}_1 = \tilde{x}_2$. Then $g_1^{-1}g_2$ fixes this point, so $g_1^{-1}g_2 = \mathbb{1}$ and $g_1 = g_2$.

Note that in $(*)$ it suffices to take g_1 to be the identity since $g_1(U) \cap g_2(U) \neq \varnothing$ is equivalent to $U \cap g_1^{-1}g_2(U) \neq \varnothing$. Thus we have the equivalent condition that $U \cap g(U) \neq \varnothing$ only when g is the identity.

Given an action of a group G on a space Y, we can form a space Y/G, the quotient space of Y in which each point y is identified with all its images $g(y)$ as g ranges over G. The points of Y/G are thus the **orbits** $Gy = \{\, g(y) \mid g \in G \,\}$ in Y, and Y/G is called the **orbit space** of the action. For example, for a normal covering space $\widetilde{X} \to X$, the orbit space $\widetilde{X}/G(\widetilde{X})$ is just X.

Proposition 1.40. *If an action of a group G on a space Y satisfies $(*)$, then:*

(a) *The quotient map $p: Y \to Y/G$, $p(y) = Gy$, is a normal covering space.*

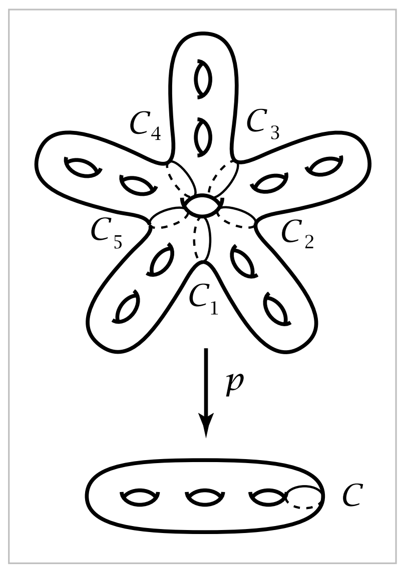
(b) *G is the group of deck transformations of this covering space $Y \to Y/G$ if Y is path-connected.*

(c) *G is isomorphic to $\pi_1(Y/G)/p_*(\pi_1(Y))$ if Y is path-connected and locally path-connected.*

Proof: Given an open set $U \subset Y$ as in condition $(*)$, the quotient map p simply identifies all the disjoint homeomorphic sets $\{\, g(U) \mid g \in G \,\}$ to a single open set $p(U)$ in Y/G. By the definition of the quotient topology on Y/G, p restricts to a homeomorphism from $g(U)$ onto $p(U)$ for each $g \in G$ so we have a covering space. Each element of G acts as a deck transformation, and the covering space is normal since $g_2 g_1^{-1}$ takes $g_1(U)$ to $g_2(U)$. The deck transformation group contains G as a subgroup, and equals this subgroup if Y is path-connected, since if f is any deck transformation, then for an arbitrarily chosen point $y \in Y$, y and $f(y)$ are in the same orbit and there is a $g \in G$ with $g(y) = f(y)$, hence $f = g$ since deck transformations of a path-connected covering space are uniquely determined by where they send a point. The final statement of the proposition is immediate from part (b) of Proposition 1.39. ☐

In view of the preceding proposition, we shall call an action satisfying $(*)$ a **covering space action**. This is not standard terminology, but there does not seem to be a universally accepted name for actions satisfying $(*)$. Sometimes these are called ‘properly discontinuous’ actions, but more often this rather unattractive term means something weaker: Every point $x \in X$ has a neighborhood U such that $U \cap g(U)$ is nonempty for only finitely many $g \in G$. Many symmetry groups have this proper discontinuity property without satisfying $(*)$, for example the group of symmetries of the familiar tiling of $\mathbb{R}^2$ by regular hexagons. The reason why the action of this group on $\mathbb{R}^2$ fails to satisfy $(*)$ is that there are **fixed points**: points y for which there is a nontrivial element $g \in G$ with $g(y) = y$. For example, the vertices of the hexagons are fixed by the 120 degree rotations about these points, and the midpoints of edges are fixed by 180 degree rotations. An action without fixed points is called a **free** action. Thus for a free action of G on Y, only the identity element of G fixes any point of Y. This is equivalent to requiring that all the images $g(y)$ of each $y \in Y$ are distinct, or in other words $g_1(y) = g_2(y)$ only when $g_1 = g_2$, since $g_1(y) = g_2(y)$ is equivalent to $g_1^{-1}g_2(y) = y$. Though condition $(*)$ implies freeness, the converse is not always true. An example is the action of $\mathbb{Z}$ on S^1 in which a generator of $\mathbb{Z}$ acts by rotation through an angle α that is an irrational multiple of 2π. In this case each orbit $\mathbb{Z}y$ is dense in S^1, so condition $(*)$ cannot hold since it implies that orbits are discrete subspaces. An exercise at the end of the section is to show that for actions on Hausdorff spaces, freeness plus proper discontinuity implies condition $(*)$. Note that proper discontinuity is automatic for actions by a finite group.

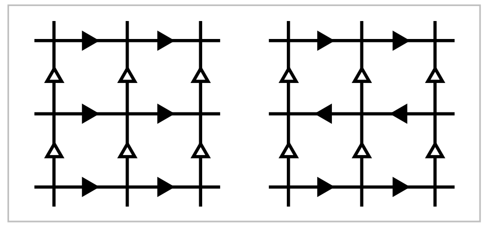
Example 1.41. Let Y be the closed orientable surface of genus 11, an '11-hole torus' as shown in the figure. This has a 5-fold rotational symmetry, generated by a rotation of angle $2\pi/5$. Thus we have the cyclic group $\mathbb{Z}_5$ acting on Y, and the condition $(*)$ is obviously satisfied. The quotient space $Y/\mathbb{Z}_5$ is a surface of genus 3, obtained from one of the five subsurfaces of Y cut off by the circles $C_1, \cdots, C_5$ by identifying its two boundary circles C_i and C_{i+1} to form the circle C as shown. Thus we have a covering space $M_{11} \to M_3$ where M_g denotes the closed orientable surface of genus g. In particular, we see that $\pi_1(M_3)$ contains the 'larger' group $\pi_1(M_{11})$ as a normal subgroup of index 5, with quotient $\mathbb{Z}_5$. This example obviously generalizes by replacing the two holes in each 'arm' of M_{11} by m holes and the 5-fold symmetry by n-fold symmetry. This gives a covering space $M_{mn+1} \to M_{m+1}$. An exercise in §2.2 is to show by an Euler characteristic argument that if there is a covering space $M_g \to M_h$ then $g = mn + 1$ and $h = m + 1$ for some m and n.



As a special case of the final statement of the preceding proposition we see that for a covering space action of a group G on a simply-connected locally path-connected space Y, the orbit space Y/G has fundamental group isomorphic to G. Under this isomorphism an element $g \in G$ corresponds to a loop in Y/G that is the projection of a path in Y from a chosen basepoint y_0 to $g(y_0)$. Any two such paths are homotopic since Y is simply-connected, so we get a well-defined element of $\pi_1(Y/G)$ associated to g.

This method for computing fundamental groups via group actions on simply-connected spaces is essentially how we computed $\pi_1(S^1)$ in §1.1, via the covering space $\mathbb{R}\to S^1$ arising from the action of $\mathbb{Z}$ on $\mathbb{R}$ by translations. This is a useful general technique for computing fundamental groups, in fact. Here are some examples illustrating this idea.

Example 1.42. Consider the grid in $\mathbb{R}^2$ formed by the horizontal and vertical lines through points in $\mathbb{Z}^2$. Let us decorate this grid with arrows in either of the two ways shown in the figure, the difference between the two cases being that in the second case the horizontal arrows in adjacent lines point in opposite directions. The group G consisting of all symmetries of the first decorated grid is isomorphic to $\mathbb{Z}\times\mathbb{Z}$ since it consists of all translations $(x,y)\mapsto(x+m,y+n)$ for $m,n\in\mathbb{Z}$. For the second grid the symmetry group G contains a subgroup of translations of the form $(x,y)\mapsto(x+m,y+2n)$ for $m,n\in\mathbb{Z}$, but there are also glide-reflection symmetries consisting of vertical translation by an odd integer distance followed by reflection across a vertical line, either a vertical line of the grid or a vertical line halfway between two adjacent grid lines. For both decorated grids there are elements of G taking any square to any other, but only the identity element of G takes a square to itself. The minimum distance any point is moved by a nontrivial element of G is 1, which easily implies the covering space condition $(*)$. The orbit space $\mathbb{R}^2/G$ is the quotient space of a square in the grid with opposite edges identified according to the arrows. Thus we see that the fundamental groups of the torus and the Klein bottle are the symmetry groups G in the two cases. In the second case the subgroup of G formed by the translations has index two, and the orbit space for this subgroup is a torus forming a two-sheeted covering space of the Klein bottle.



Example 1.43: $\mathbb{R}\mathrm{P}^n$. The antipodal map of S^n, $x\mapsto -x$, generates an action of $\mathbb{Z}_2$ on S^n with orbit space $\mathbb{R}\mathrm{P}^n$, real projective n-space, as defined in Example 0.4. The action is a covering space action since each open hemisphere in S^n is disjoint from its antipodal image. As we saw in Proposition 1.14, S^n is simply-connected if $n\geq 2$, so from the covering space $S^n\to\mathbb{R}\mathrm{P}^n$ we deduce that $\pi_1(\mathbb{R}\mathrm{P}^n)\approx\mathbb{Z}_2$ for $n\geq 2$. A generator for $\pi_1(\mathbb{R}\mathrm{P}^n)$ is any loop obtained by projecting a path in S^n connecting two antipodal points. One can see explicitly that such a loop γ has order two in $\pi_1(\mathbb{R}\mathrm{P}^n)$ if $n\geq 2$ since the composition $\gamma\cdot\gamma$ lifts to a loop in S^n, and this can be homotoped to the trivial loop since $\pi_1(S^n)=0$, so the projection of this homotopy into $\mathbb{R}\mathrm{P}^n$ gives a nullhomotopy of $\gamma\cdot\gamma$.

One may ask whether there are other finite groups that act freely on S^n, defining covering spaces $S^n \to S^n/G$. We will show in Proposition 2.29 that $\mathbb{Z}_2$ is the only possibility when n is even, but for odd n the question is much more difficult. It is easy to construct a free action of any cyclic group $\mathbb{Z}_m$ on S^{2k-1}, the action generated by the rotation $v \mapsto e^{2\pi i/m}v$ of the unit sphere S^{2k-1} in $\mathbb{C}^k = \mathbb{R}^{2k}$. This action is free since an equation $v = e^{2\pi i\ell/m}v$ with $0 < \ell < m$ implies $v = 0$, but 0 is not a point of S^{2k-1}. The orbit space $S^{2k-1}/\mathbb{Z}_m$ is one of a family of spaces called *lens spaces* defined in Example 2.43.

There are also noncyclic finite groups that act freely as rotations of S^n for odd $n > 1$. These actions are classified quite explicitly in [Wolf 1984]. Examples in the simplest case $n = 3$ can be produced as follows. View $\mathbb{R}^4$ as the quaternion algebra $\mathbb{H}$. Multiplication of quaternions satisfies $|ab| = |a||b|$ where $|a|$ denotes the usual Euclidean length of a vector $a \in \mathbb{R}^4$. Thus if a and b are unit vectors, so is ab, and hence quaternion multiplication defines a map $S^3 \times S^3 \to S^3$. This in fact makes S^3 into a group, though associativity is all we need now since associativity implies that any subgroup G of S^3 acts on S^3 by left-multiplication, $g(x) = gx$. This action is free since an equation $x = gx$ in the division algebra $\mathbb{H}$ implies $g = 1$ or $x = 0$. As a concrete example, G could be the familiar quaternion group $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$ from group theory. More generally, for a positive integer m, let Q_{4m} be the subgroup of S^3 generated by the two quaternions $a = e^{\pi i/m}$ and $b = j$. Thus a has order $2m$ and b has order 4. The easily verified relations $a^m = b^2 = -1$ and $bab^{-1} = a^{-1}$ imply that the subgroup $\mathbb{Z}_{2m}$ generated by a is normal and of index 2 in Q_{4m}. Hence Q_{4m} is a group of order $4m$, called the *generalized quaternion group*. Another common name for this group is the *binary dihedral group* D^*_{4m} since its quotient by the subgroup $\{\pm 1\}$ is the ordinary dihedral group D_{2m} of order $2m$.

Besides the groups $Q_{4m} = D^*_{4m}$ there are just three other noncyclic finite subgroups of S^3: the binary tetrahedral, octahedral, and icosahedral groups T^*_{24}, O^*_{48}, and I^*_{120}, of orders indicated by the subscripts. These project two-to-one onto the groups of rotational symmetries of a regular tetrahedron, octahedron (or cube), and icosahedron (or dodecahedron). In fact, it is not hard to see that the homomorphism $S^3 \to SO(3)$ sending $u \in S^3 \subset \mathbb{H}$ to the isometry $v \to u^{-1}vu$ of $\mathbb{R}^3$, viewing $\mathbb{R}^3$ as the 'pure imaginary' quaternions $v = ai + bj + ck$, is surjective with kernel $\{\pm 1\}$. Then the groups D^*_{4m}, T^*_{24}, O^*_{48}, I^*_{120} are the preimages in S^3 of the groups of rotational symmetries of a regular polygon or polyhedron in $\mathbb{R}^3$.

There are two conditions that a finite group G acting freely on S^n must satisfy:

(a) Every abelian subgroup of G is cyclic. This is equivalent to saying that G contains no subgroup $\mathbb{Z}_p \times \mathbb{Z}_p$ with p prime.

(b) G contains at most one element of order 2.

A proof of (a) is sketched in an exercise for §4.2. For a proof of (b) the original source [Milnor 1957] is recommended reading. The groups satisfying (a) have been completely classified; see [Brown 1982], section VI.9, for details. An example of a group satisfying (a) but not (b) is the dihedral group D_{2m} for odd $m > 1$.

There is also a much more difficult converse: A finite group satisfying (a) and (b) acts freely on S^n for some n. References for this are [Madsen, Thomas, & Wall 1976] and [Davis & Milgram 1985]. There is also almost complete information about which n's are possible for a given group.

Example 1.44. In Example 1.35 we constructed a contractible 2-complex $\widetilde{X}_{m,n} = T_{m,n}\times\mathbb{R}$ as the universal cover of a finite 2-complex $X_{m,n}$ that was the union of the mapping cylinders of the two maps $S^1\to S^1$, $z\mapsto z^m$ and $z\mapsto z^n$. The group of deck transformations of this covering space is therefore the fundamental group $\pi_1(X_{m,n})$. From van Kampen's theorem applied to the decomposition of $X_{m,n}$ into the two mapping cylinders we have the presentation $\langle\, a,b \mid a^m b^{-n}\,\rangle$ for this group $G_{m,n} = \pi_1(X_{m,n})$. It is interesting to look at the action of $G_{m,n}$ on $\widetilde{X}_{m,n}$ more closely. We described a decomposition of $\widetilde{X}_{m,n}$ into rectangles, with $X_{m,n}$ the quotient of one rectangle. These rectangles in fact define a cell structure on $\widetilde{X}_{m,n}$ lifting a cell structure on $X_{m,n}$ with two vertices, three edges, and one 2-cell. The group $G_{m,n}$ is thus a group of symmetries of this cell structure on $\widetilde{X}_{m,n}$. If we orient the three edges of $X_{m,n}$ and lift these orientations to the edges of $\widetilde{X}_{m,n}$, then $G_{m,n}$ is the group of all symmetries of $\widetilde{X}_{m,n}$ preserving the orientations of edges. For example, the element a acts as a 'screw motion' about an axis that is a vertical line $\{v_a\}\times\mathbb{R}$ with v_a a vertex of $T_{m,n}$, and b acts similarly for a vertex v_b.

Since the action of $G_{m,n}$ on $\widetilde{X}_{m,n}$ preserves the cell structure, it also preserves the product structure $T_{m,n}\times\mathbb{R}$. This means that there are actions of $G_{m,n}$ on $T_{m,n}$ and $\mathbb{R}$ such that the action on the product $X_{m,n} = T_{m,n}\times\mathbb{R}$ is the diagonal action $g(x,y) = \big(g(x),g(y)\big)$ for $g\in G_{m,n}$. If we make the rectangles of unit height in the $\mathbb{R}$ coordinate, then the element $a^m = b^n$ acts on $\mathbb{R}$ as unit translation, while a acts by $^1\!/_m$ translation and b by $^1\!/_n$ translation. The translation actions of a and b on $\mathbb{R}$ generate a group of translations of $\mathbb{R}$ that is infinite cyclic, generated by translation by the reciprocal of the least common multiple of m and n.

The action of $G_{m,n}$ on $T_{m,n}$ has kernel consisting of the powers of the element $a^m = b^n$. This infinite cyclic subgroup is precisely the center of $G_{m,n}$, as we saw in Example 1.24. There is an induced action of the quotient group $\mathbb{Z}_m * \mathbb{Z}_n$ on $T_{m,n}$, but this is not a free action since the elements a and b and all their conjugates fix vertices of $T_{m,n}$. On the other hand, if we restrict the action of $G_{m,n}$ on $T_{m,n}$ to the kernel K of the map $G_{m,n}\to\mathbb{Z}$ given by the action of $G_{m,n}$ on the $\mathbb{R}$ factor of $X_{m,n}$, then we do obtain a free action of K on $T_{m,n}$. Since this action takes vertices to vertices and edges to edges, it is a covering space action, so K is a free group, the fundamental group of the graph $T_{m,n}/K$. An exercise at the end of the section is to determine $T_{m,n}/K$ explicitly and compute the number of generators of K.

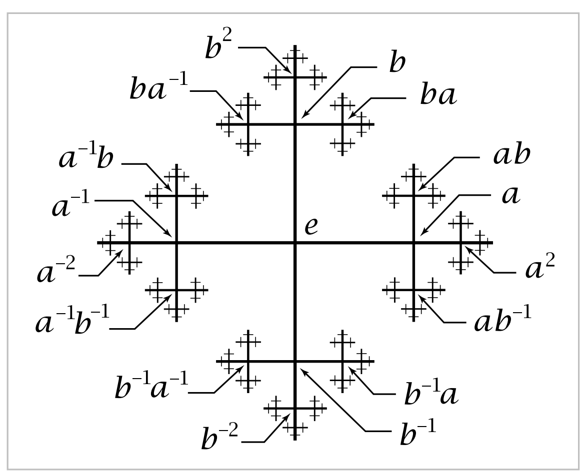
Cayley Complexes

Covering spaces can be used to describe a very classical method for viewing groups geometrically as graphs. Recall from Corollary 1.28 how we associated to each group presentation $G = \langle\, g_\alpha \mid r_\beta \,\rangle$ a 2-dimensional cell complex X_G with $\pi_1(X_G) \approx G$ by taking a wedge-sum of circles, one for each generator g_α, and then attaching a 2-cell for each relator r_β. We can construct a cell complex $\widetilde{X}_G$ with a covering space action of G such that $\widetilde{X}_G/G = X_G$ in the following way. Let the vertices of $\widetilde{X}_G$ be the elements of G themselves. Then, at each vertex $g \in G$, insert an edge joining g to gg_α for each of the chosen generators g_α. The resulting graph is known as the **Cayley graph** of G with respect to the generators g_α. This graph is connected since every element of G is a product of g_α's, so there is a path in the graph joining each vertex to the identity vertex e. Each relation r_β determines a loop in the graph, starting at any vertex g, and we attach a 2-cell for each such loop. The resulting cell complex $\widetilde{X}_G$ is the **Cayley complex** of G. The group G acts on $\widetilde{X}_G$ by multiplication on the left. Thus, an element $g \in G$ sends a vertex $g' \in G$ to the vertex gg', and the edge from g' to $g'g_\alpha$ is sent to the edge from gg' to $gg'g_\alpha$. The action extends to 2-cells in the obvious way. This is clearly a covering space action, and the orbit space is just X_G.

In fact $\widetilde{X}_G$ is the universal cover of X_G since it is simply-connected. This can be seen by considering the homomorphism $\varphi:\pi_1(X_G)\to G$ defined in the proof of Proposition 1.39. For an edge e_α in X_G corresponding to a generator g_α of G, it is clear from the definition of φ that $\varphi([e_\alpha]) = g_\alpha$, so φ is an isomorphism. In particular the kernel of φ, $p_*(\pi_1(\widetilde{X}_G))$, is zero, hence also $\pi_1(\widetilde{X}_G)$ since p_* is injective.

Let us look at some examples of Cayley complexes.

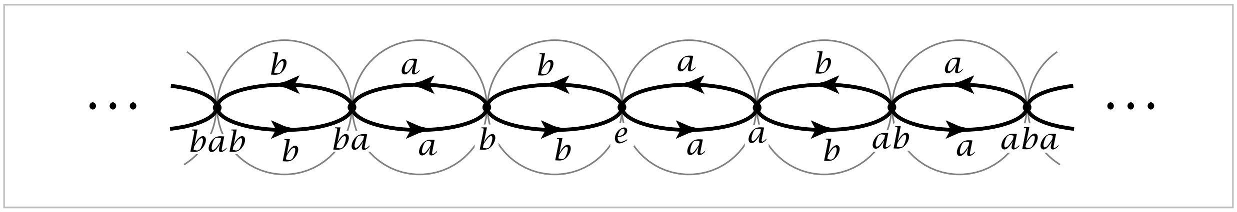
Example 1.45. When G is the free group on two generators a and b, X_G is $S^1 \vee S^1$ and $\widetilde{X}_G$ is the Cayley graph of $\mathbb{Z} * \mathbb{Z}$ pictured at the right. The action of a on this graph is a rightward shift along the central horizontal axis, while b acts by an upward shift along the central vertical axis. The composition ab of these two shifts then takes the vertex e to the vertex ab. Similarly, the action of any $w \in \mathbb{Z} * \mathbb{Z}$ takes e to the vertex w.



Example 1.46. The group $G = \mathbb{Z}\times\mathbb{Z}$ with presentation $\langle\, x,y \mid xyx^{-1}y^{-1} \,\rangle$ has X_G the torus $S^1\times S^1$, and $\widetilde{X}_G$ is $\mathbb{R}^2$ with vertices the integer lattice $\mathbb{Z}^2 \subset \mathbb{R}^2$ and edges the horizontal and vertical segments between these lattice points. The action of G is by translations $(x,y) \mapsto (x+m, y+n)$.

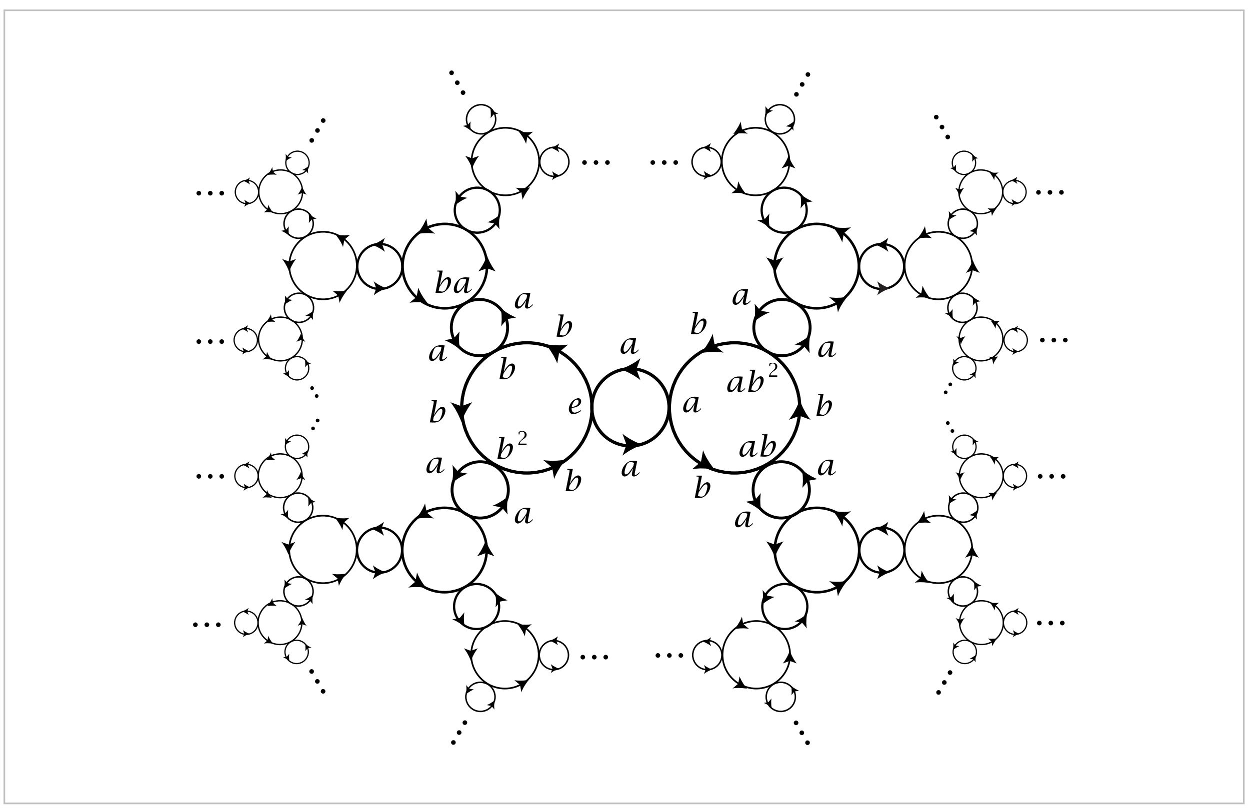
Example 1.47. For $G = \mathbb{Z}_2 = \langle\, x \mid x^2 \,\rangle$, X_G is $\mathbb{R}\mathrm{P}^2$ and $\widetilde{X}_G = S^2$. More generally, for $\mathbb{Z}_n = \langle\, x \mid x^n \,\rangle$, X_G is S^1 with a disk attached by the map $z \mapsto z^n$ and $\widetilde{X}_G$ consists of n disks $D_1, \cdots, D_n$ with their boundary circles identified. A generator of $\mathbb{Z}_n$ acts on this union of disks by sending D_i to D_{i+1} via a $2\pi/n$ rotation, the subscript i being taken mod n. The common boundary circle of the disks is rotated by $2\pi/n$.

Example 1.48. If $G = \mathbb{Z}_2 * \mathbb{Z}_2 = \langle\, a, b \mid a^2, b^2 \,\rangle$ then the Cayley graph is a union of an infinite sequence of circles each tangent to its two neighbors.



We obtain $\widetilde{X}_G$ from this graph by making each circle the equator of a 2-sphere, yielding an infinite sequence of tangent 2-spheres. Elements of the index-two normal subgroup $\mathbb{Z} \subset \mathbb{Z}_2 * \mathbb{Z}_2$ generated by ab act on $\widetilde{X}_G$ as translations by an even number of units, while each of the remaining elements of $\mathbb{Z}_2 * \mathbb{Z}_2$ acts as the antipodal map on one of the spheres and flips the whole chain of spheres end-for-end about this sphere. The orbit space X_G is $\mathbb{R}\mathrm{P}^2 \vee \mathbb{R}\mathrm{P}^2$.

It is not hard to see the generalization of this example to $\mathbb{Z}_m * \mathbb{Z}_n$ with the presentation $\langle\, a, b \mid a^m, b^n \,\rangle$, so that $\widetilde{X}_G$ consists of an infinite union of copies of the Cayley complexes for $\mathbb{Z}_m$ and $\mathbb{Z}_n$ constructed in Example 1.47, arranged in a tree-like pattern. The case of $\mathbb{Z}_2 * \mathbb{Z}_3$ is pictured below.



Exercises

1. For a covering space $p:\widetilde{X}\to X$ and a subspace $A\subset X$, let $\widetilde{A}=p^{-1}(A)$. Show that the restriction $p:\widetilde{A}\to A$ is a covering space.

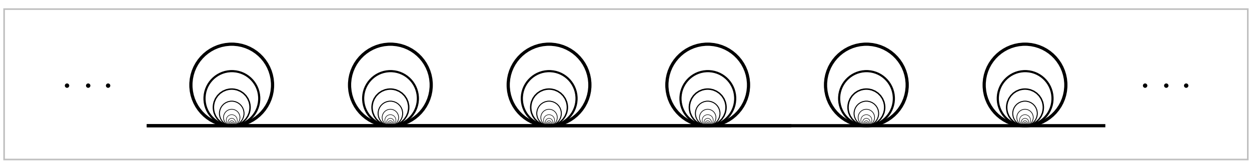
2. Show that if $p_1:\widetilde{X}_1\to X_1$ and $p_2:\widetilde{X}_2\to X_2$ are covering spaces, so is their product $p_1\times p_2:\widetilde{X}_1\times\widetilde{X}_2\to X_1\times X_2$.

3. Let $p:\widetilde{X}\to X$ be a covering space with $p^{-1}(x)$ finite and nonempty for all $x\in X$. Show that $\widetilde{X}$ is compact Hausdorff iff X is compact Hausdorff.

4. Construct a simply-connected covering space of the space $X\subset\mathbb{R}^3$ that is the union of a sphere and a diameter. Do the same when X is the union of a sphere and a circle intersecting it in two points.

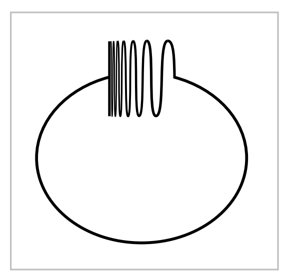
5. Let X be the subspace of $\mathbb{R}^2$ consisting of the four sides of the square $[0,1]\times[0,1]$ together with the segments of the vertical lines $x={}^1\!/_2,{}^1\!/_3,{}^1\!/_4,\cdots$ inside the square. Show that for every covering space $\widetilde{X}\to X$ there is some neighborhood of the left edge of X that lifts homeomorphically to $\widetilde{X}$. Deduce that X has no simply-connected covering space.

6. Let X be the shrinking wedge of circles in Example 1.25, and let $\widetilde{X}$ be its covering space shown in the figure below.



Construct a two-sheeted covering space $Y\to\widetilde{X}$ such that the composition $Y\to\widetilde{X}\to X$ of the two covering spaces is not a covering space. Note that a composition of two covering spaces does have the unique path lifting property, however.

7. Let Y be the *quasi-circle* shown in the figure, a closed subspace of $\mathbb{R}^2$ consisting of a portion of the graph of $y=\sin(1/x)$, the segment $[-1,1]$ in the y-axis, and an arc connecting these two pieces. Collapsing the segment of Y in the y-axis to a point gives a quotient map $f:Y\to S^1$. Show that f does not lift to the covering space $\mathbb{R}\to S^1$, even though $\pi_1(Y)=0$. Thus local path-connectedness of Y is a necessary hypothesis in the lifting criterion.



8. Let $\widetilde{X}$ and $\widetilde{Y}$ be simply-connected covering spaces of the path-connected, locally path-connected spaces X and Y. Show that if $X\simeq Y$ then $\widetilde{X}\simeq\widetilde{Y}$. [Exercise 11 in Chapter 0 may be helpful.]

9. Show that if a path-connected, locally path-connected space X has $\pi_1(X)$ finite, then every map $X\to S^1$ is nullhomotopic. [Use the covering space $\mathbb{R}\to S^1$.]

10. Find all the connected 2-sheeted and 3-sheeted covering spaces of $S^1\vee S^1$, up to isomorphism of covering spaces without basepoints.

11. Construct finite graphs X_1 and X_2 having a common finite-sheeted covering space $\widetilde{X}_1 = \widetilde{X}_2$, but such that there is no space having both X_1 and X_2 as covering spaces.

12. Let a and b be the generators of $\pi_1(S^1 \vee S^1)$ corresponding to the two S^1 summands. Draw a picture of the covering space of $S^1 \vee S^1$ corresponding to the normal subgroup generated by a^2, b^2, and $(ab)^4$, and prove that this covering space is indeed the correct one.

13. Determine the covering space of $S^1 \vee S^1$ corresponding to the subgroup of $\pi_1(S^1 \vee S^1)$ generated by the cubes of all elements. The covering space is 27-sheeted and can be drawn on a torus so that the complementary regions are nine triangles with edges labeled aaa, nine triangles with edges labeled bbb, and nine hexagons with edges labeled $ababab$. [For the analogous problem with sixth powers instead of cubes, the resulting covering space would have $2^{28}3^{25}$ sheets! And for k^{th} powers with k sufficiently large, the covering space would have infinitely many sheets. The underlying group theory question here, whether the quotient of $\mathbb{Z} * \mathbb{Z}$ obtained by factoring out all k^{th} powers is finite, is known as Burnside's problem. It can also be asked for a free group on n generators.]

14. Find all the connected covering spaces of $\mathbb{R}\mathrm{P}^2 \vee \mathbb{R}\mathrm{P}^2$.

15. Let $p : \widetilde{X} \to X$ be a simply-connected covering space of X and let $A \subset X$ be a path-connected, locally path-connected subspace, with $\widetilde{A} \subset \widetilde{X}$ a path-component of $p^{-1}(A)$. Show that $p : \widetilde{A} \to A$ is the covering space corresponding to the kernel of the map $\pi_1(A) \to \pi_1(X)$.

16. Given maps $X \to Y \to Z$ such that both $Y \to Z$ and the composition $X \to Z$ are covering spaces, show that $X \to Y$ is a covering space if Z is locally path-connected, and show that this covering space is normal if $X \to Z$ is a normal covering space.

17. Given a group G and a normal subgroup N, show that there exists a normal covering space $\widetilde{X} \to X$ with $\pi_1(X) \approx G$, $\pi_1(\widetilde{X}) \approx N$, and deck transformation group $G(\widetilde{X}) \approx G/N$.

18. For a path-connected, locally path-connected, and semilocally simply-connected space X, call a path-connected covering space $\widetilde{X} \to X$ *abelian* if it is normal and has abelian deck transformation group. Show that X has an abelian covering space that is a covering space of every other abelian covering space of X, and that such a 'universal' abelian covering space is unique up to isomorphism. Describe this covering space explicitly for $X = S^1 \vee S^1$ and $X = S^1 \vee S^1 \vee S^1$.

19. Use the preceding problem to show that a closed orientable surface M_g of genus g has a connected normal covering space with deck transformation group isomorphic to $\mathbb{Z}^n$ (the product of n copies of $\mathbb{Z}$) iff $n \le 2g$. For $n = 3$ and $g \ge 3$, describe such a covering space explicitly as a subspace of $\mathbb{R}^3$ with translations of $\mathbb{R}^3$ as deck transformations. Show that such a covering space in $\mathbb{R}^3$ exists iff there is an embedding of M_g in the 3-torus $T^3 = S^1 \times S^1 \times S^1$ such that the induced map $\pi_1(M_g) \to \pi_1(T^3)$ is surjective.

20. Construct nonnormal covering spaces of the Klein bottle by a Klein bottle and by a torus.

21. Let X be the space obtained from a torus $S^1 \times S^1$ by attaching a Möbius band via a homeomorphism from the boundary circle of the Möbius band to the circle $S^1 \times \{x_0\}$ in the torus. Compute $\pi_1(X)$, describe the universal cover of X, and describe the action of $\pi_1(X)$ on the universal cover. Do the same for the space Y obtained by attaching a Möbius band to $\mathbb{R}\mathrm{P}^2$ via a homeomorphism from its boundary circle to the circle in $\mathbb{R}\mathrm{P}^2$ formed by the 1-skeleton of the usual CW structure on $\mathbb{R}\mathrm{P}^2$.

22. Given covering space actions of groups G_1 on X_1 and G_2 on X_2, show that the action of $G_1 \times G_2$ on $X_1 \times X_2$ defined by $(g_1, g_2)(x_1, x_2) = (g_1(x_1), g_2(x_2))$ is a covering space action, and that $(X_1 \times X_2)/(G_1 \times G_2)$ is homeomorphic to $X_1/G_1 \times X_2/G_2$.

23. Show that if a group G acts freely and properly discontinuously on a Hausdorff space X, then the action is a covering space action. (Here 'properly discontinuously' means that each $x \in X$ has a neighborhood U such that $\{\, g \in G \mid U \cap g(U) \neq \varnothing \,\}$ is finite.) In particular, a free action of a finite group on a Hausdorff space is a covering space action.

24. Given a covering space action of a group G on a path-connected, locally path-connected space X, then each subgroup $H \subset G$ determines a composition of covering spaces $X \to X/H \to X/G$. Show:

(a) Every path-connected covering space between X and X/G is isomorphic to X/H for some subgroup $H \subset G$.

(b) Two such covering spaces X/H_1 and X/H_2 of X/G are isomorphic iff H_1 and H_2 are conjugate subgroups of G.

(c) The covering space $X/H \to X/G$ is normal iff H is a normal subgroup of G, in which case the group of deck transformations of this cover is G/H.

25. Let $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation $\varphi(x, y) = (2x, y/2)$. This generates an action of $\mathbb{Z}$ on $X = \mathbb{R}^2 - \{0\}$. Show this action is a covering space action and compute $\pi_1(X/\mathbb{Z})$. Show the orbit space $X/\mathbb{Z}$ is non-Hausdorff, and describe how it is a union of four subspaces homeomorphic to $S^1 \times \mathbb{R}$, coming from the complementary components of the x-axis and the y-axis.

26. For a covering space $p : \widetilde{X} \to X$ with X connected, locally path-connected, and semilocally simply-connected, show:

(a) The components of $\widetilde{X}$ are in one-to-one correspondence with the orbits of the action of $\pi_1(X, x_0)$ on the fiber $p^{-1}(x_0)$.

(b) Under the Galois correspondence between connected covering spaces of X and subgroups of $\pi_1(X, x_0)$, the subgroup corresponding to the component of $\widetilde{X}$

containing a given lift $\widetilde{x}_0$ of x_0 is the *stabilizer* of $\widetilde{x}_0$, the subgroup consisting of elements whose action on the fiber leaves $\widetilde{x}_0$ fixed.

27. For a universal cover $p:\widetilde{X}\to X$ there are two actions of $\pi_1(X,x_0)$ on the fiber $p^{-1}(x_0)$. The first is the action defined on page 69 in which the element of $\pi_1(X,x_0)$ determined by a loop γ sends $\widetilde{\gamma}(1)$ to $\widetilde{\gamma}(0)$ for each lift $\widetilde{\gamma}$ of γ to $\widetilde{X}$, and the second is the action given by restricting deck transformations to the fiber (see Proposition 1.39). Show that these two actions are different when $X = S^1 \vee S^1$ and when $X = S^1\times S^1$ and determine when the two actions are the same. [This is a revised version of the original form of this exercise.]

28. Show that for a covering space action of a group G on a simply-connected space Y, $\pi_1(Y/G)$ is isomorphic to G. [If Y is locally path-connected, this is a special case of part (c) of Proposition 1.40.]

29. Let Y be path-connected, locally path-connected, and simply-connected, and let G_1 and G_2 be subgroups of $\mathrm{Homeo}(Y)$ defining covering space actions on Y. Show that the orbit spaces Y/G_1 and Y/G_2 are homeomorphic iff G_1 and G_2 are conjugate subgroups of $\mathrm{Homeo}(Y)$.

30. Draw the Cayley graph of the group $\mathbb{Z} * \mathbb{Z}_2 = \langle\, a, b \mid b^2 \,\rangle$.

31. Show that the normal covering spaces of $S^1 \vee S^1$ are precisely the graphs that are Cayley graphs of groups with two generators. More generally, the normal covering spaces of the wedge sum of n circles are the Cayley graphs of groups with n generators.

32. Consider covering spaces $p:\widetilde{X}\to X$ with $\widetilde{X}$ and X connected CW complexes, the cells of $\widetilde{X}$ projecting homeomorphically onto cells of X. Restricting p to the 1-skeleton then gives a covering space $\widetilde{X}^1\to X^1$ over the 1-skeleton of X. Show:

(a) Two such covering spaces $\widetilde{X}_1\to X$ and $\widetilde{X}_2\to X$ are isomorphic iff the restrictions $\widetilde{X}_1^1\to X^1$ and $\widetilde{X}_2^1\to X^1$ are isomorphic.

(b) $\widetilde{X}\to X$ is a normal covering space iff $\widetilde{X}^1\to X^1$ is normal.

(c) The groups of deck transformations of the coverings $\widetilde{X}\to X$ and $\widetilde{X}^1\to X^1$ are isomorphic, via the restriction map.

33. In Example 1.44 let d be the greatest common divisor of m and n, and let $m' = m/d$ and $n' = n/d$. Show that the graph $T_{m,n}/K$ consists of m' vertices labeled a, n' vertices labeled b, together with d edges joining each a vertex to each b vertex. Deduce that the subgroup $K \subset G_{m,n}$ is free on $dm'n' - m' - n' + 1$ generators.

Additional Topics

1.A Graphs and Free Groups

Since all groups can be realized as fundamental groups of spaces, this opens the way for using topology to study algebraic properties of groups. The topics in this section and the next give some illustrations of this principle, mainly using covering space theory.

We remind the reader that the Additional Topics which form the remainder of this chapter are not to be regarded as an essential part of the basic core of the book. Readers who are eager to move on to new topics should feel free to skip ahead.

By definition, a **graph** is a 1-dimensional CW complex, in other words, a space X obtained from a discrete set X^0 by attaching a collection of 1-cells e_α. Thus X is obtained from the disjoint union of X^0 with closed intervals I_α by identifying the two endpoints of each I_α with points of X^0. The points of X^0 are the **vertices** and the 1-cells the **edges** of X. Note that with this definition an edge does not include its endpoints, so an edge is an open subset of X. The two endpoints of an edge can be the same vertex, so the closure $\overline{e}_\alpha$ of an edge e_α is homeomorphic either to I or S^1.

Since X has the quotient topology from the disjoint union $X^0 \coprod_\alpha I_\alpha$, a subset of X is open (or closed) iff it intersects the closure $\overline{e}_\alpha$ of each edge e_α in an open (or closed) set in $\overline{e}_\alpha$. One says that X has the **weak topology** with respect to the subspaces $\overline{e}_\alpha$. In this topology a sequence of points in the interiors of distinct edges forms a closed subset, hence never converges. This is true in particular if the edges containing the sequence all have a common vertex and one tries to choose the sequence so that it gets 'closer and closer' to the vertex. Thus if there is a vertex that is the endpoint of infinitely many edges, then the weak topology cannot be a metric topology. An exercise at the end of this section is to show the converse, that the weak topology is a metric topology if each vertex is an endpoint of only finitely many edges.

A basis for the topology of X consists of the open intervals in the edges together with the path-connected neighborhoods of the vertices. A neighborhood of the latter sort about a vertex v is the union of connected open neighborhoods U_α of v in $\overline{e}_\alpha$ for all $\overline{e}_\alpha$ containing v. In particular, we see that X is locally path-connected. Hence a graph is connected iff it is path-connected.

If X has only finitely many vertices and edges, then X is compact, being the continuous image of the compact space $X^0 \coprod_\alpha I_\alpha$. The converse is also true, and more generally, a compact subset C of a graph X can meet only finitely many vertices and edges of X. To see this, let the subspace $D \subset C$ consist of the vertices in C together with one point in each edge that C meets. Then D is a closed subset of X since it meets each $\overline{e}_\alpha$ in a closed set. For the same reason, any subset of D is closed, so D has the discrete topology. But D is compact, being a closed subset of the compact space C, so D must be finite. By the definition of D this means that C can meet only finitely many vertices and edges.

A **subgraph** of a graph X is a subspace $Y \subset X$ that is a union of vertices and edges of X, such that $e_\alpha \subset Y$ implies $\overline{e}_\alpha \subset Y$. The latter condition just says that Y is a closed subspace of X. A **tree** is a contractible graph. By a tree in a graph X we mean a subgraph that is a tree. We call a tree in X **maximal** if it contains all the vertices of X. This is equivalent to the more obvious meaning of maximality, as we will see below.

Proposition 1A.1. *Every connected graph contains a maximal tree, and in fact any tree in the graph is contained in a maximal tree.*

Proof: Let X be a connected graph. We will describe a construction that embeds an arbitrary subgraph $X_0 \subset X$ as a deformation retract of a subgraph $Y \subset X$ that contains all the vertices of X. By choosing X_0 to be any subtree of X, for example a single vertex, this will prove the proposition.

As a preliminary step, we construct a sequence of subgraphs $X_0 \subset X_1 \subset X_2 \subset \cdots$, letting X_{i+1} be obtained from X_i by adjoining the closures $\overline{e}_\alpha$ of all edges $e_\alpha \subset X - X_i$ having at least one endpoint in X_i. The union $\bigcup_i X_i$ is open in X since a neighborhood of a point in X_i is contained in X_{i+1}. Furthermore, $\bigcup_i X_i$ is closed since it is a union of closed edges and X has the weak topology. So $X = \bigcup_i X_i$ since X is connected.

Now to construct Y we begin by setting $Y_0 = X_0$. Then inductively, assuming that $Y_i \subset X_i$ has been constructed so as to contain all the vertices of X_i, let Y_{i+1} be obtained from Y_i by adjoining one edge connecting each vertex of $X_{i+1} - X_i$ to Y_i, and let $Y = \bigcup_i Y_i$. It is evident that Y_{i+1} deformation retracts to Y_i, and we may obtain a deformation retraction of Y to $Y_0 = X_0$ by performing the deformation retraction of Y_{i+1} to Y_i during the time interval $[1/2^{i+1}, 1/2^i]$. Thus a point $x \in Y_{i+1} - Y_i$ is stationary until this interval, when it moves into Y_i and thereafter continues moving until it reaches Y_0. The resulting homotopy $h_t : Y \to Y$ is continuous since it is continuous on the closure of each edge and Y has the weak topology. $\square$

Given a maximal tree $T \subset X$ and a base vertex $x_0 \in T$, then each edge e_α of $X - T$ determines a loop f_α in X that goes first from x_0 to one endpoint of e_α by a path in T, then across e_α, then back to x_0 by a path in T. Strictly speaking, we should first orient the edge e_α in order to specify which direction to cross it. Note that the homotopy class of f_α is independent of the choice of the paths in T since T is simply-connected.

Proposition 1A.2. *For a connected graph X with maximal tree T, $\pi_1(X)$ is a free group with basis the classes $[f_\alpha]$ corresponding to the edges e_α of $X - T$.*

In particular this implies that a maximal tree is maximal in the sense of not being contained in any larger tree, since adjoining any edge to a maximal tree produces a graph with nontrivial fundamental group. Another consequence is that a graph is a tree iff it is simply-connected.

Proof: The quotient map $X \to X/T$ is a homotopy equivalence by Proposition 0.17. The quotient X/T is a graph with only one vertex, hence is a wedge sum of circles, whose fundamental group we showed in Example 1.21 to be free with basis the loops given by the edges of X/T, which are the images of the loops f_α in X. $\square$

Here is a very useful fact about graphs:

Lemma 1A.3. *Every covering space of a graph is also a graph, with vertices and edges the lifts of the vertices and edges in the base graph.*

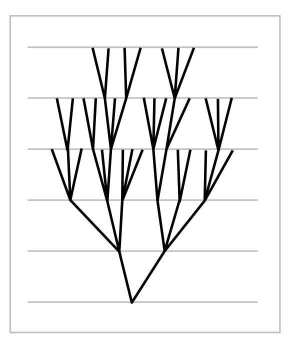
Proof: Let $p:\widetilde{X} \to X$ be the covering space. For the vertices of $\widetilde{X}$ we take the discrete set $\widetilde{X}^0 = p^{-1}(X^0)$. Writing X as a quotient space of $X^0 \coprod_\alpha I_\alpha$ as in the definition of a graph and applying the path lifting property to the resulting maps $I_\alpha \to X$, we get a unique lift $I_\alpha \to \widetilde{X}$ passing through each point in $p^{-1}(x)$, for $x \in e_\alpha$. These lifts define the edges of a graph structure on $\widetilde{X}$. The resulting topology on $\widetilde{X}$ is the same as its original topology since both topologies have the same basic open sets, the covering projection $\widetilde{X} \to X$ being a local homeomorphism. $\square$

We can now apply what we have proved about graphs and their fundamental groups to prove a basic fact of group theory:

Theorem 1A.4. *Every subgroup of a free group is free.*

Proof: Given a free group F, choose a graph X with $\pi_1(X) \approx F$, for example a wedge of circles corresponding to a basis for F. For each subgroup G of F there is by Proposition 1.36 a covering space $p:\widetilde{X} \to X$ with $p_*(\pi_1(\widetilde{X})) = G$, hence $\pi_1(\widetilde{X}) \approx G$ since p_* is injective by Proposition 1.31. Since $\widetilde{X}$ is a graph by the preceding lemma, the group $G \approx \pi_1(\widetilde{X})$ is free by Proposition 1A.2. $\square$

The structure of trees can be elucidated by looking more closely at the constructions in the proof of Proposition 1A.1. If X is a tree and v_0 is any vertex of X, then the construction of a maximal tree $Y \subset X$ starting with $Y_0 = \{v_0\}$ yields an increasing sequence of subtrees $Y_n \subset X$ whose union is all of X since a tree has only one maximal subtree, namely itself. We can think of the vertices in $Y_n - Y_{n-1}$ as being at 'height' n, with the edges of $Y_n - Y_{n-1}$ connecting these vertices to vertices of height $n-1$. In this way we get a 'height function' $h:X \to \mathbb{R}$ assigning to each vertex its height, and monotone on edges.



For each vertex v of X there is exactly one edge leading downward from v, so by following these downward edges we obtain a path from v to the base vertex v_0. This is an example of an **edgepath**, which is a composition of finitely many paths each consisting of a single edge traversed monotonically. For any edgepath joining v to v_0 other than the downward edgepath, the height function would not be monotone and hence would have local maxima, occurring when the edgepath backtracked, retracing some edge it had just crossed. Thus in a tree there is a unique nonbacktracking edgepath joining any two points. All the vertices and edges along this edgepath are distinct.

A tree can contain no subgraph homeomorphic to a circle, since two vertices in such a subgraph could be joined by more than one nonbacktracking edgepath. Conversely, if a connected graph X contains no circle subgraph, then it must be a tree. For if T is a maximal tree in X that is not equal to X, then the union of an edge of $X-T$ with the nonbacktracking edgepath in T joining the endpoints of this edge is a circle subgraph of X. So if there are no circle subgraphs of X, we must have $X=T$, a tree.

For an arbitrary connected graph X and a pair of vertices v_0 and v_1 in X there is a unique nonbacktracking edgepath in each homotopy class of paths from v_0 to v_1. This can be seen by lifting to the universal cover $\widetilde{X}$, which is a tree since it is simply-connected. Choosing a lift $\widetilde{v}_0$ of v_0, a homotopy class of paths from v_0 to v_1 lifts to a homotopy class of paths starting at $\widetilde{v}_0$ and ending at a unique lift $\widetilde{v}_1$ of v_1. Then the unique nonbacktracking edgepath in $\widetilde{X}$ from $\widetilde{v}_0$ to $\widetilde{v}_1$ projects to the desired nonbacktracking edgepath in X.

Exercises

1. Let X be a graph in which each vertex is an endpoint of only finitely many edges. Show that the weak topology on X is a metric topology.

2. Show that a connected graph retracts onto any connected subgraph.

3. For a finite graph X define the Euler characteristic $\chi(X)$ to be the number of vertices minus the number of edges. Show that $\chi(X)=1$ if X is a tree, and that the rank (number of elements in a basis) of $\pi_1(X)$ is $1-\chi(X)$ if X is connected.

4. If X is a finite graph and Y is a subgraph homeomorphic to S^1 and containing the basepoint x_0, show that $\pi_1(X,x_0)$ has a basis in which one element is represented by the loop Y.

5. Construct a connected graph X and maps $f,g:X\to X$ such that $fg=1\!\!1$ but f and g do not induce isomorphisms on π_1. [Note that $f_*g_*=1\!\!1$ implies that f_* is surjective and g_* is injective.]

6. Let F be the free group on two generators and let F' be its commutator subgroup. Find a set of free generators for F' by considering the covering space of the graph $S^1\vee S^1$ corresponding to F'.

7. If F is a finitely generated free group and N is a nontrivial normal subgroup of infinite index, show, using covering spaces, that N is not finitely generated.

8. Show that a finitely generated group has only a finite number of subgroups of a given finite index. [First do the case of free groups, using covering spaces of graphs. The general case then follows since every group is a quotient group of a free group.]

9. Using covering spaces, show that an index n subgroup H of a group G has at most n conjugate subgroups gHg^{-1} in G. Apply this to show that there exists a normal subgroup $K \subset G$ of finite index with $K \subset H$. [For the latter statement, consider the intersection of all the conjugate subgroups gHg^{-1}. This is the maximal normal subgroup of G contained in H.]

10. Let X be the wedge sum of n circles, with its natural graph structure, and let $\widetilde{X} \to X$ be a covering space with $Y \subset \widetilde{X}$ a finite connected subgraph. Show there is a finite graph $Z \supset Y$ having the same vertices as Y, such that the projection $Y \to X$ extends to a covering space $Z \to X$.

11. Apply the two preceding problems to show that if F is a finitely generated free group and $x \in F$ is not the identity element, then there is a normal subgroup $H \subset F$ of finite index such that $x \notin H$. Hence x has nontrivial image in a finite quotient group of F. In this situation one says F is *residually finite*.

12. Let F be a finitely generated free group, $H \subset F$ a finitely generated subgroup, and $x \in F - H$. Show there is a subgroup K of finite index in F such that $K \supset H$ and $x \notin K$. [Apply Exercise 10.]

13. Let x be a nontrivial element of a finitely generated free group F. Show there is a finite-index subgroup $H \subset F$ in which x is one element of a basis. [Exercises 4 and 10 may be helpful.]

14. Show that the existence of maximal trees is equivalent to the Axiom of Choice.

1.B K(G,1) Spaces and Graphs of Groups

In this section we introduce a class of spaces whose homotopy type depends only on their fundamental group. These spaces arise many places in topology, especially in its interactions with group theory.

A path-connected space whose fundamental group is isomorphic to a given group G and which has a contractible universal covering space is called a **K(G,1) space.** The '1' here refers to π_1. More general $K(G,n)$ spaces are studied in §4.2. All these spaces are called Eilenberg–MacLane spaces, though in the case $n = 1$ they were studied by

Hurewicz before Eilenberg and MacLane took up the general case. Here are some examples:

Example 1B.1. S^1 is a $K(\mathbb{Z},1)$. More generally, a connected graph is a $K(G,1)$ with G a free group, since by the results of §1.A its universal cover is a tree, hence contractible.

Example 1B.2. Closed surfaces with infinite π_1, in other words, closed surfaces other than S^2 and $\mathbb{R}P^2$, are $K(G,1)$'s. This will be shown in Example 1B.14 below. It also follows from the theorem in surface theory that the only simply-connected surfaces without boundary are S^2 and $\mathbb{R}^2$, so the universal cover of a closed surface with infinite fundamental group must be $\mathbb{R}^2$ since it is noncompact. Nonclosed surfaces deformation retract onto graphs, so such surfaces are $K(G,1)$'s with G free.

Example 1B.3. The infinite-dimensional projective space $\mathbb{R}P^\infty$ is a $K(\mathbb{Z}_2,1)$ since its universal cover is S^∞, which is contractible. To show the latter fact, a homotopy from the identity map of S^∞ to a constant map can be constructed in two stages as follows. First, define $f_t:\mathbb{R}^\infty \to \mathbb{R}^\infty$ by $f_t(x_1,x_2,\cdots) = (1-t)(x_1,x_2,\cdots) + t(0,x_1,x_2,\cdots)$. This takes nonzero vectors to nonzero vectors for all $t \in [0,1]$, so $f_t/|f_t|$ gives a homotopy from the identity map of S^∞ to the map $(x_1,x_2,\cdots) \mapsto (0,x_1,x_2,\cdots)$. Then a homotopy from this map to a constant map is given by $g_t/|g_t|$ where $g_t(x_1,x_2,\cdots) = (1-t)(0,x_1,x_2,\cdots) + t(1,0,0,\cdots)$.

Example 1B.4. Generalizing the preceding example, we can construct a $K(\mathbb{Z}_m,1)$ as an infinite-dimensional lens space $S^\infty/\mathbb{Z}_m$, where $\mathbb{Z}_m$ acts on S^∞, regarded as the unit sphere in $\mathbb{C}^\infty$, by scalar multiplication by m^{th} roots of unity, a generator of this action being the map $(z_1,z_2,\cdots) \mapsto e^{2\pi i/m}(z_1,z_2,\cdots)$. It is not hard to check that this is a covering space action.

Example 1B.5. A product $K(G,1)\times K(H,1)$ is a $K(G\times H,1)$ since its universal cover is the product of the universal covers of $K(G,1)$ and $K(H,1)$. By taking products of circles and infinite-dimensional lens spaces we therefore get $K(G,1)$'s for arbitrary finitely generated abelian groups G. For example the n-dimensional torus T^n, the product of n circles, is a $K(\mathbb{Z}^n,1)$.

Example 1B.6. For a closed connected subspace K of S^3 that is nonempty, the complement S^3-K is a $K(G,1)$. This is a theorem in 3-manifold theory, but in the special case that K is a torus knot the result follows from our study of torus knot complements in Examples 1.24 and 1.35. Namely, we showed that for K the torus knot $K_{m,n}$ there is a deformation retraction of S^3-K onto a certain 2-dimensional complex $X_{m,n}$ having contractible universal cover. The homotopy lifting property then implies that the universal cover of S^3-K is homotopy equivalent to the universal cover of $X_{m,n}$, hence is also contractible.

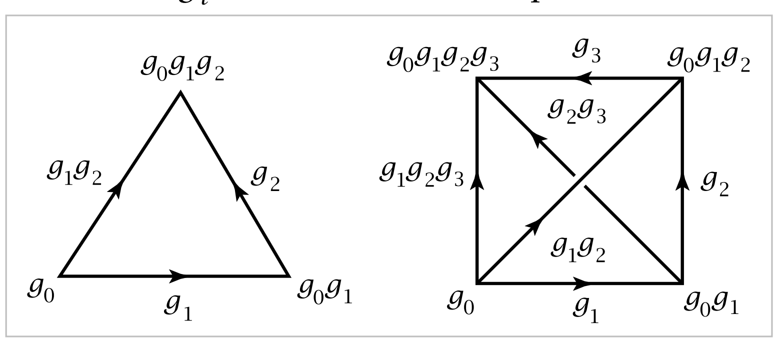
Example 1B.7. It is not hard to construct a $K(G,1)$ for an arbitrary group G, using the notion of a Δ-complex defined in §2.1. Let EG be the Δ-complex whose n-simplices are the ordered $(n+1)$-tuples $[g_0,\cdots,g_n]$ of elements of G. Such an n-simplex attaches to the $(n-1)$-simplices $[g_0,\cdots,\widehat{g}_i,\cdots,g_n]$ in the obvious way, just as a standard simplex attaches to its faces. (The notation $\widehat{g}_i$ means that this vertex is deleted.) The complex EG is contractible by the homotopy h_t that slides each point $x\in[g_0,\cdots,g_n]$ along the line segment in $[e,g_0,\cdots,g_n]$ from x to the vertex $[e]$, where e is the identity element of G. This is well-defined in EG since when we restrict to a face $[g_0,\cdots,\widehat{g}_i,\cdots,g_n]$ we have the linear deformation to $[e]$ in $[e,g_0,\cdots,\widehat{g}_i,\cdots,g_n]$. Note that h_t carries $[e]$ around the loop $[e,e]$, so h_t is not actually a deformation retraction of EG onto $[e]$.

The group G acts on EG by left multiplication, an element $g\in G$ taking the simplex $[g_0,\cdots,g_n]$ linearly onto the simplex $[gg_0,\cdots,gg_n]$. Only the identity e takes any simplex to itself, so by an exercise at the end of this section, the action of G on EG is a covering space action. Hence the quotient map $EG\to EG/G$ is the universal cover of the orbit space $BG=EG/G$, and BG is a $K(G,1)$.

Since G acts on EG by freely permuting simplices, BG inherits a Δ-complex structure from EG. The action of G on EG identifies all the vertices of EG, so BG has just one vertex. To describe the Δ-complex structure on BG explicitly, note first that every n-simplex of EG can be written uniquely in the form

$$[g_0,g_0g_1,g_0g_1g_2,\cdots,g_0g_1\cdots g_n]=g_0[e,g_1,g_1g_2,\cdots,g_1\cdots g_n]$$

The image of this simplex in BG may be denoted unambiguously by the symbol $[g_1|g_2|\cdots|g_n]$. In this 'bar' notation the g_i's and their ordered products can be used to label edges, viewing an edge label as the ratio between the two labels on the vertices at the endpoints of the edge, as indicated in the figure. With this notation, the boundary of a simplex $[g_1|\cdots|g_n]$ of BG consists of the simplices $[g_2|\cdots|g_n]$, $[g_1|\cdots|g_{n-1}]$, and $[g_1|\cdots|g_ig_{i+1}|\cdots|g_n]$ for $i=1,\cdots,n-1$.



This construction of a $K(G,1)$ produces a rather large space, since BG is always infinite-dimensional, and if G is infinite, BG has an infinite number of cells in each positive dimension. For example, $B\mathbb{Z}$ is much bigger than S^1, the most efficient $K(\mathbb{Z},1)$. On the other hand, BG has the virtue of being functorial: A homomorphism $f:G\to H$ induces a map $Bf:BG\to BH$ sending a simplex $[g_1|\cdots|g_n]$ to the simplex $[f(g_1)|\cdots|f(g_n)]$. A different construction of a $K(G,1)$ is given in §4.2. Here one starts with any 2-dimensional complex having fundamental group G, for example the complex X_G associated to a presentation of G, and then one attaches cells of dimension 3 and higher to make the universal cover contractible without affecting π_1. In general, it is hard to get any control on the number of higher-dimensional cells needed in this construction, so it too can be rather inefficient. Indeed, finding an efficient $K(G,1)$ for a given group G is often a difficult problem.

It is a curious and almost paradoxical fact that if G contains any elements of finite order, then every $K(G,1)$ CW complex must be infinite-dimensional. This is shown in Proposition 2.45. In particular the infinite-dimensional lens space $K(\mathbb{Z}_m,1)$'s in Example 1B.4 cannot be replaced by any finite-dimensional complex.

In spite of the great latitude possible in the construction of $K(G,1)$'s, there is a very nice homotopical uniqueness property that accounts for much of the interest in $K(G,1)$'s:

Theorem 1B.8. *The homotopy type of a CW complex $K(G,1)$ is uniquely determined by G.*

Having a unique homotopy type of $K(G,1)$'s associated to each group G means that algebraic invariants of spaces that depend only on homotopy type, such as homology and cohomology groups, become invariants of groups. This has proved to be a quite fruitful idea, and has been much studied both from the algebraic and topological viewpoints. The discussion following Proposition 2.45 gives a few references.

The preceding theorem will follow easily from:

Proposition 1B.9. *Let X be a connected CW complex and let Y be a $K(G,1)$. Then every homomorphism $\pi_1(X,x_0)\to\pi_1(Y,y_0)$ is induced by a map $(X,x_0)\to(Y,y_0)$ that is unique up to homotopy fixing x_0.*

To deduce the theorem from this, let X and Y be CW complex $K(G,1)$'s with isomorphic fundamental groups. The proposition gives maps $f:(X,x_0)\to(Y,y_0)$ and $g:(Y,y_0)\to(X,x_0)$ inducing inverse isomorphisms $\pi_1(X,x_0)\approx\pi_1(Y,y_0)$. Then fg and gf induce the identity on π_1 and hence are homotopic to the identity maps.

Proof of 1B.9: Let us first consider the case that X has a single 0-cell, the basepoint x_0. Given a homomorphism $\varphi:\pi_1(X,x_0)\to\pi_1(Y,y_0)$, we begin the construction of a map $f:(X,x_0)\to(Y,y_0)$ with $f_* = \varphi$ by setting $f(x_0) = y_0$. Each 1-cell e^1_α of X has closure a circle determining an element $[e^1_\alpha]\in\pi_1(X,x_0)$, and we let f on the closure of e^1_α be a map representing $\varphi([e^1_\alpha])$. If $i:X^1\hookrightarrow X$ denotes the inclusion, then $\varphi i_* = f_*$ since $\pi_1(X^1,x_0)$ is generated by the elements $[e^1_\alpha]$.

$$\begin{tikzcd} \pi_1(X^1,x_0) \arrow[rr, "f_*"] \arrow[dr, "i_*"'] & & \pi_1(Y,y_0) \\ & \pi_1(X,x_0) \arrow[ur, "\varphi"'] & \end{tikzcd}$$

To extend f over a cell e^2_β with attaching map $\psi_\beta:S^1\to X^1$, all we need is for the composition $f\psi_\beta$ to be nullhomotopic. Choosing a basepoint $s_0\in S^1$ and a path in X^1 from $\psi_\beta(s_0)$ to x_0, ψ_β determines an element $[\psi_\beta]\in\pi_1(X^1,x_0)$, and the existence of a nullhomotopy of $f\psi_\beta$ is equivalent to $f_*([\psi_\beta])$ being zero in $\pi_1(Y,y_0)$. We have $i_*([\psi_\beta]) = 0$ since the cell e_β^2 provides a nullhomotopy of ψ_β in X. Hence $f_*([\psi_\beta]) = \varphi i_*([\psi_\beta]) = 0$, and so f can be extended over e_β^2.

Extending f inductively over cells e_γ^n with $n > 2$ is possible since the attaching maps $\psi_\gamma : S^{n-1} \to X^{n-1}$ have nullhomotopic compositions $f\psi_\gamma : S^{n-1} \to Y$. This is because $f\psi_\gamma$ lifts to the universal cover of Y if $n > 2$, and this cover is contractible by hypothesis, so the lift of $f\psi_\gamma$ is nullhomotopic, hence also $f\psi_\gamma$ itself.

Turning to the uniqueness statement, if two maps $f_0, f_1 : (X,x_0) \to (Y,y_0)$ induce the same homomorphism on π_1, then we see immediately that their restrictions to X^1 are homotopic, fixing x_0. To extend the resulting map $X^1 \times I \cup X \times \partial I \to Y$ over the remaining cells $e^n \times (0,1)$ of $X \times I$ we can proceed just as in the preceding paragraph since these cells have dimension $n+1 > 2$. Thus we obtain a homotopy $f_t : (X,x_0) \to (Y,y_0)$, finishing the proof in the case that X has a single 0-cell.

The case that X has more than one 0-cell can be treated by a small elaboration on this argument. Choose a maximal tree $T \subset X$. To construct a map f realizing a given φ, begin by setting $f(T) = y_0$. Then each edge e_α^1 in $X - T$ determines an element $[e_\alpha^1] \in \pi_1(X,x_0)$, and we let f on the closure of e_α^1 be a map representing $\varphi([e_\alpha^1])$. Extending f over higher-dimensional cells then proceeds just as before. Constructing a homotopy f_t joining two given maps f_0 and f_1 with $f_{0*} = f_{1*}$ also has an extra step. Let $h_t : X^1 \to X^1$ be a homotopy starting with $h_0 = \mathbb{1}$ and restricting to a deformation retraction of T onto x_0. (It is easy to extend such a deformation retraction to a homotopy defined on all of X^1.) We can construct a homotopy from $f_0|X^1$ to $f_1|X^1$ by first deforming $f_0|X^1$ and $f_1|X^1$ to take T to y_0 by composing with h_t, then applying the earlier argument to obtain a homotopy between the modified $f_0|X^1$ and $f_1|X^1$. Having a homotopy $f_0|X^1 \simeq f_1|X^1$ we extend this over all of X in the same way as before. ☐

The first part of the preceding proof also works for the 2-dimensional complexes X_G associated to presentations of groups. Thus every homomorphism $G \to H$ is realized as the induced homomorphism of some map $X_G \to X_H$. However, there is no uniqueness statement for this map, and it can easily happen that different presentations of a group G give X_G's that are not homotopy equivalent.

Graphs of Groups

As an illustration of how $K(G,1)$ spaces can be useful in group theory, we shall describe a procedure for assembling a collection of $K(G,1)$'s together into a $K(G,1)$ for a larger group G. Group-theoretically, this gives a method for assembling smaller groups together to form a larger group, generalizing the notion of free products.

Let Γ be a graph that is connected and oriented, that is, its edges are viewed as arrows, each edge having a specified direction. Suppose that at each vertex v of Γ we place a group G_v and along each edge e of Γ we put a homomorphism φ_e from the group at the tail of the edge to the group at the head of the edge. We call this data a **graph of groups.** Now build a space $B\Gamma$ by putting the space BG_v from Example 1B.7 at each vertex v of Γ and then filling in a mapping cylinder of the map $B\varphi_e$ along each edge e of Γ, identifying the two ends of the mapping cylinder with the two BG_v's at the ends of e. The resulting space $B\Gamma$ is then a CW complex since the maps $B\varphi_e$ take n-cells homeomorphically onto n-cells. In fact, the cell structure on $B\Gamma$ can be canonically subdivided into a Δ-complex structure using the prism construction from the proof of Theorem 2.10, but we will not need to do this here.

More generally, instead of BG_v one could take any CW complex $K(G_v,1)$ at the vertex v, and then along edges put mapping cylinders of maps realizing the homomorphisms φ_e. We leave it for the reader to check that the resulting space $K\Gamma$ is homotopy equivalent to the $B\Gamma$ constructed above.

Example 1B.10. Suppose Γ consists of one central vertex with a number of edges radiating out from it, and the group G_v at this central vertex is trivial, hence also all the edge homomorphisms. Then van Kampen's theorem implies that $\pi_1(K\Gamma)$ is the free product of the groups at all the outer vertices.

In view of this example, we shall call $\pi_1(K\Gamma)$ for a general graph of groups Γ the **graph product** of the vertex groups G_v with respect to the edge homomorphisms φ_e. The name for $\pi_1(K\Gamma)$ that is generally used in the literature is the rather awkward phrase, 'the fundamental group of the graph of groups'.

Here is the main result we shall prove about graphs of groups:

Theorem 1B.11. *If all the edge homomorphisms φ_e are injective, then $K\Gamma$ is a $K(G,1)$ and the inclusions $K(G_v,1)\hookrightarrow K\Gamma$ induce injective maps on π_1.*

Before giving the proof, let us look at some interesting special cases:

Example 1B.12: Free Products with Amalgamation. Suppose the graph of groups is $A \leftarrow C \rightarrow B$, with the two maps monomorphisms. One can regard this data as specifying embeddings of C as subgroups of A and B. Applying van Kampen's theorem to the decomposition of $K\Gamma$ into its two mapping cylinders, we see that $\pi_1(K\Gamma)$ is the quotient of $A * B$ obtained by identifying the subgroup $C \subset A$ with the subgroup $C \subset B$. The standard notation for this group is $A *_C B$, the free product of A and B **amalgamated** along the subgroup C. According to the theorem, $A *_C B$ contains both A and B as subgroups.

For example, a free product with amalgamation $\mathbb{Z} *_{\mathbb{Z}} \mathbb{Z}$ can be realized by mapping cylinders of the maps $S^1 \leftarrow S^1 \rightarrow S^1$ that are m-sheeted and n-sheeted covering spaces, respectively. We studied this case in Examples 1.24 and 1.35 where we showed that the complex $K\Gamma$ is a deformation retract of the complement of a torus knot in S^3 if m and n are relatively prime. It is a basic result in 3-manifold theory that the complement of every smooth knot in S^3 can be built up by iterated graph of groups constructions with injective edge homomorphisms, starting with free groups, so the theorem implies that these knot complements are $K(G,1)$'s. Their universal covers are all $\mathbb{R}^3$, in fact.

Example 1B.13: HNN Extensions. Consider a graph of groups $C \overset{\varphi}{\underset{\psi}{\rightrightarrows}} A$ with φ and ψ both monomorphisms. This is analogous to the previous case $A \leftarrow C \rightarrow B$, but with the two groups A and B coalesced to a single group. The group $\pi_1(K\Gamma)$, which was denoted $A *_C B$ in the previous case, is now denoted $A*_C$. To see what this group looks like, let us regard $K\Gamma$ as being obtained from $K(A,1)$ by attaching $K(C,1)\times I$ along the two ends $K(C,1)\times\partial I$ via maps realizing the monomorphisms φ and ψ. Using a $K(C,1)$ with a single 0-cell, we see that $K\Gamma$ can be obtained from $K(A,1)\vee S^1$ by attaching cells of dimension two and greater, so $\pi_1(K\Gamma)$ is a quotient of $A*\mathbb{Z}$, and it is not hard to figure out that the relations defining this quotient are of the form $t\varphi(c)t^{-1} = \psi(c)$ where t is a generator of the $\mathbb{Z}$ factor and c ranges over C, or a set of generators for C. We leave the verification of this for the Exercises.

As a very special case, taking $\varphi = \psi = \mathbb{1}$ gives $A*_A = A\times\mathbb{Z}$ since we can take $K\Gamma = K(A,1)\times S^1$ in this case. More generally, taking $\varphi = \mathbb{1}$ with ψ an arbitrary automorphism of A, we realize any semidirect product of A and $\mathbb{Z}$ as $A*_A$. For example, the Klein bottle occurs this way, with φ realized by the identity map of S^1 and ψ by a reflection. In these cases when $\varphi = \mathbb{1}$ we could realize the same group $\pi_1(K\Gamma)$ using a slightly simpler graph of groups, with a single vertex, labeled A, and a single edge, labeled ψ.

Here is another special case. Suppose we take a torus, delete a small open disk, then identify the resulting boundary circle with a longitudinal circle of the torus. This produces a space X that happens to be homeomorphic to a subspace of the standard picture of a Klein bottle in $\mathbb{R}^3$; see Exercise 12 of §1.2. The fundamental group $\pi_1(X)$ has the form $(\mathbb{Z}*\mathbb{Z})*_{\mathbb{Z}}\mathbb{Z}$ with the defining relation $tb^{\pm1}t^{-1} = aba^{-1}b^{-1}$ where a is a meridional loop and b is a longitudinal loop on the torus. The sign of the exponent in the term $b^{\pm1}$ is immaterial since the two ways of glueing the boundary circle to the longitude produce homeomorphic spaces. The group $\pi_1(X) = \langle a,b,t \mid tbt^{-1}aba^{-1}b^{-1}\rangle$ abelianizes to $\mathbb{Z}\times\mathbb{Z}$, but to show that $\pi_1(X)$ is not isomorphic to $\mathbb{Z}*\mathbb{Z}$ takes some work. There is a surjection $\pi_1(X)\to\mathbb{Z}*\mathbb{Z}$ obtained by setting $b=1$. This has nontrivial kernel since b is nontrivial in $\pi_1(X)$ by the preceding theorem. If $\pi_1(X)$ were isomorphic to $\mathbb{Z}*\mathbb{Z}$ we would then have a surjective homomorphism $\mathbb{Z}*\mathbb{Z}\to\mathbb{Z}*\mathbb{Z}$ that was not an isomorphism. However, it is a theorem in group theory that a free group F is *hopfian* — every surjective homomorphism $F\to F$ must be injective. Hence $\pi_1(X)$ is not free.

Example 1B.14: Closed Surfaces. A closed orientable surface M of genus two or greater can be cut along a circle into two compact surfaces M_1 and M_2 such that the closed surfaces obtained from M_1 and M_2 by filling in their boundary circle with a disk have smaller genus than M. Each of M_1 and M_2 is the mapping cylinder of a map from S^1 to a finite graph. Namely, view M_i as obtained from a closed surface by deleting an open disk in the interior of the 2-cell in the standard CW structure described in Chapter 0, so that M_i becomes the mapping cylinder of the attaching map of the 2-cell. This attaching map is not nullhomotopic, so it induces an injection on π_1 since free groups are torsionfree. Thus we have realized the original surface M as $K\Gamma$ for Γ a graph of groups of the form $F_1 \longleftarrow \mathbb{Z} \longrightarrow F_2$ with F_1 and F_2 free and the two maps injective. The theorem then says that M is a $K(G,1)$.

A similar argument works for closed nonorientable surfaces other than $\mathbb{RP}^2$. For example, the Klein bottle is obtained from two Möbius bands by identifying their boundary circles, and a Möbius band is the mapping cylinder of the 2-sheeted covering space $S^1 \to S^1$.

Proof of 1B.11: We shall construct a covering space $\widetilde{K} \to K\Gamma$ by gluing together copies of the universal covering spaces of the various mapping cylinders in $K\Gamma$ in such a way that $\widetilde{K}$ will be contractible. Hence $\widetilde{K}$ will be the universal cover of $K\Gamma$, which will therefore be a $K(G,1)$.

A preliminary observation: Given a universal covering space $p:\widetilde{X} \to X$ and a connected, locally path-connected subspace $A \subset X$ such that the inclusion $A \hookrightarrow X$ induces an injection on π_1, then each component $\widetilde{A}$ of $p^{-1}(A)$ is a universal cover of A. To see this, note that $p:\widetilde{A} \to A$ is a covering space, so we have injective maps $\pi_1(\widetilde{A}) \to \pi_1(A) \to \pi_1(X)$ whose composition factors through $\pi_1(\widetilde{X}) = 0$, hence $\pi_1(\widetilde{A}) = 0$. For example, if X is the torus $S^1 \times S^1$ and A is the circle $S^1 \times \{x_0\}$, then $p^{-1}(A)$ consists of infinitely many parallel lines in $\mathbb{R}^2$, each a universal cover of A.

For a map $f:A \to B$ between connected CW complexes, let $p:\widetilde{M}_f \to M_f$ be the universal cover of the mapping cylinder M_f. Then $\widetilde{M}_f$ is itself the mapping cylinder of a map $\widetilde{f}:p^{-1}(A) \to p^{-1}(B)$ since the line segments in the mapping cylinder structure on M_f lift to line segments in $\widetilde{M}_f$ defining a mapping cylinder structure. Since $\widetilde{M}_f$ is a mapping cylinder, it deformation retracts onto $p^{-1}(B)$, so $p^{-1}(B)$ is also simply-connected, hence is the universal cover of B. If f induces an injection on π_1, then the remarks in the preceding paragraph apply, and the components of $p^{-1}(A)$ are universal covers of A. If we assume further that A and B are $K(G,1)$'s, then $\widetilde{M}_f$ and the components of $p^{-1}(A)$ are contractible, and we claim that $\widetilde{M}_f$ deformation retracts onto each component $\widetilde{A}$ of $p^{-1}(A)$. Namely, the inclusion $\widetilde{A} \hookrightarrow \widetilde{M}_f$ is a homotopy equivalence since both spaces are contractible, and then Corollary 0.20 implies that $\widetilde{M}_f$ deformation retracts onto $\widetilde{A}$ since the pair $(\widetilde{M}_f, \widetilde{A})$ satisfies the homotopy extension property, as shown in Example 0.15.

Now we can describe the construction of the covering space $\widetilde{K}$ of $K\Gamma$. It will be the union of an increasing sequence of spaces $\widetilde{K}_1 \subset \widetilde{K}_2 \subset \cdots$. For the first stage, let $\widetilde{K}_1$ be the universal cover of one of the mapping cylinders M_f of $K\Gamma$. By the preceding remarks, this contains various disjoint copies of universal covers of the two $K(G_v,1)$'s at the ends of M_f. We build $\widetilde{K}_2$ from $\widetilde{K}_1$ by attaching to each of these universal covers of $K(G_v,1)$'s a copy of the universal cover of each mapping cylinder M_g of $K\Gamma$ meeting M_f at the end of M_f in question. Now repeat the process to construct $\widetilde{K}_3$ by attaching universal covers of mapping cylinders at all the universal covers of $K(G_v,1)$'s created in the previous step. In the same way, we construct $\widetilde{K}_{n+1}$ from $\widetilde{K}_n$ for all n, and then we set $\widetilde{K}=\bigcup_n \widetilde{K}_n$.

Note that $\widetilde{K}_{n+1}$ deformation retracts onto $\widetilde{K}_n$ since it is formed by attaching pieces to $\widetilde{K}_n$ that deformation retract onto the subspaces along which they attach, by our earlier remarks. It follows that $\widetilde{K}$ is contractible since we can deformation retract $\widetilde{K}_{n+1}$ onto $\widetilde{K}_n$ during the time interval $[1/2^{n+1},1/2^n]$, and then finish with a contraction of $\widetilde{K}_1$ to a point during the time interval $[{}^1\!/_2,1]$.

The natural projection $\widetilde{K}\to K\Gamma$ is clearly a covering space, so this finishes the proof that $K\Gamma$ is a $K(G,1)$.

The remaining statement that each inclusion $K(G_v,1)\hookrightarrow K\Gamma$ induces an injection on π_1 can easily be deduced from the preceding constructions. For suppose a loop $\gamma:S^1\to K(G_v,1)$ is nullhomotopic in $K\Gamma$. By the lifting criterion for covering spaces, there is a lift $\widetilde{\gamma}:S^1\to\widetilde{K}$. This has image contained in one of the copies of the universal cover of $K(G_v,1)$, so $\widetilde{\gamma}$ is nullhomotopic in this universal cover, and hence γ is nullhomotopic in $K(G_v,1)$. $\square$

The various mapping cylinders that make up the universal cover of $K\Gamma$ are arranged in a treelike pattern. The tree in question, call it $T\Gamma$, has one vertex for each copy of a universal cover of a $K(G_v,1)$ in $\widetilde{K}$, and two vertices are joined by an edge whenever the two universal covers of $K(G_v,1)$'s corresponding to these vertices are connected by a line segment lifting a line segment in the mapping cylinder structure of a mapping cylinder of $K\Gamma$. The inductive construction of $\widetilde{K}$ is reflected in an inductive construction of $T\Gamma$ as a union of an increasing sequence of subtrees $T_1\subset T_2\subset\cdots$. Corresponding to $\widetilde{K}_1$ is a subtree $T_1\subset T\Gamma$ consisting of a central vertex with a number of edges radiating out from it, an 'asterisk' with possibly an infinite number of edges. When we enlarge $\widetilde{K}_1$ to $\widetilde{K}_2$, T_1 is correspondingly enlarged to a tree T_2 by attaching a similar asterisk at the end of each outer vertex of T_1, and each subsequent enlargement is handled in the same way. The action of $\pi_1(K\Gamma)$ on $\widetilde{K}$ as deck transformations induces an action on $T\Gamma$, permuting its vertices and edges, and the orbit space of $T\Gamma$ under this action is just the original graph Γ. The action on $T\Gamma$ will not generally be a free action since the elements of a subgroup $G_v\subset\pi_1(K\Gamma)$ fix the vertex of $T\Gamma$ corresponding to one of the universal covers of $K(G_v,1)$.

There is in fact an exact correspondence between graphs of groups and groups acting on trees. See [Scott & Wall 1979] for an exposition of this rather nice theory. From the viewpoint of groups acting on trees, the definition of a graph of groups is usually taken to be slightly more restrictive than the one we have given here, namely, one considers only oriented graphs obtained from an unoriented graph by subdividing each edge by adding a vertex at its midpoint, then orienting the two resulting edges outward, away from the new vertex.

Exercises

1. Suppose a group G acts simplicially on a Δ-complex X, where 'simplicially' means that each element of G takes each simplex of X onto another simplex by a linear homeomorphism. If the action is free, show it is a covering space action.

2. Let X be a connected CW complex and G a group such that every homomorphism $\pi_1(X)\to G$ is trivial. Show that every map $X\to K(G,1)$ is nullhomotopic.

3. Show that every graph product of trivial groups is free.

4. Use van Kampen's theorem to compute $A*_C$ as a quotient of $A*\mathbb{Z}$, as stated in the text.

5. Consider the graph of groups Γ having one vertex, $\mathbb{Z}$, and one edge, the map $\mathbb{Z}\to\mathbb{Z}$ that is multiplication by 2, realized by the 2-sheeted covering space $S^1\to S^1$. Show that $\pi_1(K\Gamma)$ has presentation $\langle\, a,b \mid bab^{-1}a^{-2}\,\rangle$ and describe the universal cover of $K\Gamma$ explicitly as a product $T\times\mathbb{R}$ with T a tree. [The group $\pi_1(K\Gamma)$ is the first in a family of groups called Baumslag-Solitar groups, having presentations of the form $\langle\, a,b \mid ba^m b^{-1}a^{-n}\,\rangle$. These are HNN extensions $\mathbb{Z}*_{\mathbb{Z}}$.]

6. Show that for a graph of groups all of whose edge homomorphisms are injective maps $\mathbb{Z}\to\mathbb{Z}$, we can choose $K\Gamma$ to have universal cover a product $T\times\mathbb{R}$ with T a tree. Work out in detail the case that the graph of groups is the infinite sequence $\mathbb{Z}\xrightarrow{2}\mathbb{Z}\xrightarrow{3}\mathbb{Z}\xrightarrow{4}\mathbb{Z}\longrightarrow\cdots$ where the map $\mathbb{Z}\xrightarrow{n}\mathbb{Z}$ is multiplication by n. Show that $\pi_1(K\Gamma)$ is isomorphic to $\mathbb{Q}$ in this case. How would one modify this example to get $\pi_1(K\Gamma)$ isomorphic to the subgroup of $\mathbb{Q}$ consisting of rational numbers with denominator a power of 2?

7. Show that every graph product of groups can be realized by a graph whose vertices are partitioned into two subsets, with every oriented edge going from a vertex in the first subset to a vertex in the second subset.

8. Show that a finite graph product of finitely generated groups is finitely generated, and similarly for finitely presented groups.

9. If Γ is a finite graph of finite groups with injective edge homomorphisms, show that the graph product of the groups has a free subgroup of finite index by constructing a suitable finite-sheeted covering space of $K\Gamma$ from universal covers of the mapping cylinders in $K\Gamma$. [The converse is also true: A finitely generated group having a free subgroup of finite index is isomorphic to such a graph product. For a proof of this see [Scott & Wall 1979], Theorem 7.3.]