Guide for instructors

The original MASS course on which this text is based ran for 14 weeks, with three 50-minute lectures in each week (given by the second author). This structure has been preserved in the present book and may be used as a rough guide if you wish to use this book as the text for a class. However, some circumstances of the original course are worth noting. As part of the MASS program, the students, primarily junior and senior undergraduate math majors, formed a tight-knit and energetic group, spending a significant amount of time outside of class hours working with each other on the course material. They had a one-hour tutorial session each week with the course TA (the first author), and smaller groups often met with the TA for informal tutorials outside of this time. All of this combined to allow a faster pace of presentation than would otherwise have been possible.

In addition to the accelerated pace of the original lectures, this book has various additions, expansions, and elaborations that increase the length; in particular, only the skeleton of Chapter 6 on geometric group theory appeared in the actual lectures, and throughout the book we have added many exercises and elaborated on many explanations. Thus, in a typical undergraduate setting it is probably not realistic to cover all six chapters completely in a single semester; as a masters level course, this may be feasible depending on the preparation and energy of the students.

With that in mind, you may wish to do one of two things: either teach this as a two-semester course, which would allow ample time to go on "side excursions" and fill in details on interesting topics related to the material here (for some of these, see the list of suggested projects at the back of the book); or teach a course that uses most, but not all, of the material herein. Although mileage may vary, we expect that covering two-thirds of the book would be reasonable, so you would need to omit material equivalent to roughly 12 lectures; this may be adjusted slightly depending on whether you plan to present the material completely or to thin it out somewhat in the interest of covering more ground. Below is a discussion of which lectures in each chapter are required for use later in the book and which could be omitted; we expect that some combination will be consistent with the background of the students and with your goals for the course.

Chapter 1. The 6 lectures in Chapter 1 cover the basics of group theory, rather than any geometric topics, and could be omitted or briefly reviewed if the students have already seen some group theory. One should be sure the students are familiar with group actions, which appear in Lecture 4(b) and play a central role in later chapters.

Chapter 2. Much of Chapter 2 does not play a crucial role in later chapters and could be omitted if the goal is to discuss "groups" more than it is to discuss "geometry". The exception to this is the material on semidirect products (Lecture 9), quotients by free and discrete actions (Lecture 10(b)–(c)), and scalar products (Lecture 14(a)), which is all essential for Chapters 3–6 and should be covered before moving on. Omitting the rest of Chapter 2 would save about 6 lectures and would result in the loss of the discussion of Euclidean isometries from the synthetic point of view but would leave more time for the discussion of matrix groups, different types of geometry, algebraic topology, and geometric group theory.

Chapter 3. Lectures 15–16 could easily be condensed if the students have a strong background in linear algebra. In particular, the material in Lecture 16 on Hermitian products and unitary, normal, and symmetric matrices is not essential for the rest of the book. Lectures 17, 19, and 20 are needed for the use of fractional linear transformations later in Chapter 5. The discussion of affine and projective transformations in Lecture 18 could be omitted, which would avoid the digression into field automorphisms; beware, however, that then the characterization of fractional linear transformations in Lecture 19(a) in terms of "lines and circles" will need to be stated without proof. Lecture 21 on solvable and nilpotent matrix groups plays an important role eventually, but not until Lecture 30(e), so it could be omitted if you do not plan to go past Lecture 30(d). Lecture 22 on Lie theory does not play an essential role later in the book.

Chapters 4 and 5. The material in these chapters is in some sense the culmination of the narrative begun in the first three chapters, and we do not recommend omitting any lectures here. That said, the heart of the geometric story runs until Lecture 30(d), and one could omit Lectures 30(e) and 31 without doing too much damage.

Chapter 6. The 6 lectures in this chapter could serve as the basis for student projects rather than as classroom lectures or could be omitted entirely if a discussion of geometric group theory is beyond the ambitions of the course.

The prerequisite knowledge that we assume of the reader is discussed in the preface and consists mainly of real analysis and linear algebra. We have placed exercises throughout the text rather than gathering them at the end of sections, so each exercise appears at the point where it most logically fits the story. Exercises are marked with a different font so as to be more readily visible. Hints to selected exercises are given at the end of the book.

An important part of the original MASS course was to have the students work on independent projects, learning a specific topic in greater depth and going beyond what the course itself could cover. At the end of the book we give a list of potential projects, along with some suggested references for further reading in these topics. Some of these are on the ambitious side for an undergraduate student, but all of them should lead to some interesting places even if they are not pursued to their conclusion. This list is of course not exhaustive, and we hope the broad view taken by the text will suggest other fruitful directions to explore.

Chapter 4

Fundamental group: A different kind of group associated to geometric objects

Lecture 23. Homotopies, paths, and π_1

a. Isometries vs. homeomorphisms. By and large, we have been considering groups that arise from geometric objects as collections of symmetries. Now we turn our attention to a different class of groups, which opens the door on the world of algebraic topology.[44]

We begin by highlighting the distinction between geometry and topology in the context of metric spaces. As with so many distinctions between various closely related fields of mathematics, the distinction hinges on the conditions under which we consider two metric spaces to be "the same" or "equivalent".

The natural equivalence relation in metric geometry is isometry. Recall that two metric spaces (X, d) and (X', d') are *isometric* if there exists an isometric bijection between them—that is, a bijection

[44]Although the groups we introduce in this chapter do not initially appear to be "symmetry groups" as most of our previous examples have been, we will, nevertheless, see in Lecture 27(c) that they do in fact appear as symmetry groups acting by *deck transformations of the universal cover*.

$f\colon X \to X'$ such that $d'(f(x_1), f(x_2)) = d(x_1, x_2)$ for all $x_1, x_2 \in X$. For example, any two circles in $\mathbb{R}^2$ with the same radius are isometric, regardless of their center, while the circles $x^2+y^2 = 1$ and $x^2+y^2 = 4$ are *not* isometric to each other, nor to the square with vertices at $(\pm 1, \pm 1)$, nor to the line $x = 1$.

Nevertheless, we feel that the two circles are in some sense more akin to each other than they are to either the square or the line, and that the circles and the square are somehow more akin than the circles and the line, or the square and the line. To make the first feeling precise, observe that there is a similarity transformation $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ taking the circle of radius 1 to the circle of radius 2; indeed, *any* two circles are equivalent up to a similarity transformation. Thus, passing from metric geometry to similarity geometry is a matter of weakening the conditions under which two objects may be considered equivalent.

Weakening these conditions still further, we may consider allow even more general maps f. Writing X for the square with vertices at $(\pm 1, \pm 1)$ and S^1 for the unit circle, we may define a bijection $f\colon X \to S^1$ by $f(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$, where $\|(x, y)\| = \sqrt{x^2 + y^2}$. One may easily verify that given a sequence of points $\mathbf{x}_n \in X$, we have $\mathbf{x}_n \to \mathbf{x}_0$ on the square if and only if $f(\mathbf{x}_n) \to f(\mathbf{x}_0)$ on the circle—that is, both f and f^{-1} are continuous.

Such a map f is called a *homeomorphism*, and is the natural equivalence relation between topological spaces. Since general topology is not our subject we will not discuss the latter notion in any detail. It is enough to mention here that a metric allows to define the notion of a ball, hence that of an open set, and the notion of convergence. A topological space is a set X with a collection of subsets, called *open sets*, that contains X itself and the empty set, and satisfies the same basic properties as the collection of open sets in a metric space, namely invariance with respect to arbitrary unions and finite intersections. Then one defines a continuous map between topological spaces by a property familiar from continuous functions on $\mathbb{R}$: a map $f\colon X \to Y$ is continuous if $f^{-1}(U)$ is open in X whenever U is open in Y. If X and Y are metric spaces, this is equivalent to the definition in terms of sequences. Using this language, a homeomorphism is a bijection such that $f(U)$ is open if and only if U is open.

Given a metric space (X, d), there are in general many different metrics one can place on X that induce the same topology as d does—that is, there are many metrics d' on X such that $d(x_n, x_0) \to 0$ if and only if $d'(x_n, x_0) \to 0$. Consequently, when we are interested in topological matters, we may refer to X as a *metrizable space*, to indicate that the primary importance of d is not the geometric structure it induces, which is specific to that particular metric, but rather the topological structure, which is held in common by many equivalent metrics and depends on on the notion of convergence induced by d.[45]

Exercise 4.1. Let $X = \mathbb{R}^n$, and for every $p \geq 1$ consider the function

$$d_p(\mathbf{x}, \mathbf{y}) = ((x_1 - y_1)^p + \cdots + (x_n - y_n)^p)^{\frac{1}{p}}. \tag{4.1}$$

Show that d_p is a metric for every $p \geq 1$ and that each of these metrics defines the same topology on $\mathbb{R}^n$ (the standard one).

Exercise 4.2. Consider the following distance function in $\mathbb{R}^2$:

$$d_L((x_1, x_2), (y_1, y_2)) = |x_1 - y_1| + |x_2 - y_2| + 1 - \delta_{x_2, y_2}.$$

Prove that d_L is a metric that is not equivalent to the standard one.

b. Tori and $\mathbb{Z}^2$. As a good example of homeomorphic spaces with different metrics, let us consider the torus $\mathbb{T}^2$, which we discussed earlier in Example 2.30 of Lecture 10(c). On the one hand, the torus is the "surface of a bagel", that is, the surface of revolution in $\mathbb{R}^3$ given by (2.15) and illustrated in Figure 4.1 below. On the other hand, the torus is the quotient space $\mathbb{R}^2/\mathbb{Z}^2$, whose points are orbits of the free and discrete action of $\mathbb{Z}^2$ on $\mathbb{R}^2$ by translations.

Both of these models of the torus come equipped with a natural metric. Writing $X \subset \mathbb{R}^3$ for the surface of revolution, the Euclidean metric d on $\mathbb{R}^3$ restricts to a metric on X. Similarly, the quotient space $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ inherits a metric ρ from the Euclidean metric on $\mathbb{R}^2$, as given in (2.14) in Definition 2.28. The natural correspondence between the two models is given by the parametrization $\varphi\colon \mathbb{T}^2 \to X$ from Example 2.30.

[45]Since we will not concern ourselves with non-metrizable spaces, we are safe in conflating the topology of a space with the notion of convergence in that topology, even though the two notions are not completely equivalent for general topological spaces.

Exercise 4.3. Prove that $\varphi\colon (\mathbb{T}^2, \rho) \to (X, d)$ is a homeomorphism but is not an isometry.

In fact, there are many ways to realize the torus as a quotient of $\mathbb{R}^2$ by a free and discrete $\mathbb{Z}^2$-action. Given any two linearly independent vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^2$, we may consider the *lattice* $L = \{a\mathbf{v}+b\mathbf{w} \mid a, b \in \mathbb{Z}^2\}$, which is a normal subgroup of $\mathbb{R}^2$, and also defines a free and discrete action of $\mathbb{Z}^2$ on $\mathbb{R}^2$ by translations; the quotient $\mathbb{R}^2/L$ inherits both a group structure and a metric from $\mathbb{R}^2$.[46]

Exercise 4.4. Given a lattice $L = \langle \mathbf{v}, \mathbf{w} \rangle \subset \mathbb{R}^2$, show that L and $\mathbb{Z}^2$ are isomorphic subgroups of $\mathbb{R}^2$—that is, there exists a group isomorphism (invertible additive map) $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $f(\mathbb{Z}^2) = L$. Conclude that the two tori $\mathbb{R}^2/\mathbb{Z}^2$ and $\mathbb{R}^2/L$ are homeomorphic by using the map f to exhibit a homeomorphism between them.

One can also build homeomorphic tori in other ways: observing that $\mathbb{R}^2/\mathbb{Z}^2 = (\mathbb{R}/\mathbb{Z}) \times (\mathbb{R}/\mathbb{Z}) = S^1 \times S^1$, one can take the direct product of two circles (or two closed curves) in the plane, which is naturally embedded into $\mathbb{R}^4$. One can also generalize the surface X (the "bagel example") by constructing a surface of revolution that starts not with a circle but rather with another closed curve that has no internal symmetries, thus putting a "warp" on the surface.

All the tori we have described are homeomorphic, but they are geometrically quite different from each other, as quickly becomes apparent when we consider their isometry groups. We describe these here without giving complete proofs; it is a useful exercise to prove the statements below.

(1) Writing X for the surface of revolution in (2.15), the isometries of X are rotations around the z-axis by an arbitrary angle, rotations around any axis through the origin in the xy-plane by an angle of exactly π, and reflections in either the xy-plane or any plane that contains the z-axis. In particular, $\mathrm{Isom}(X)$ has one continuous degree of freedom.

(2) A torus obtained by rotating a non-symmetric closed curve has fewer isometries in $\mathrm{Isom}(\mathbb{R}^3)$ than the bagel; symmetry in the

[46] We write $\mathbb{R}^2/L$ for the quotient space instead of $\mathbb{R}^2/\mathbb{Z}^2$ to emphasize the specific action that is considered.

xy-plane is lost. It turns out, however, that something like it is recovered when one considers the intrinsic metric defined as the shortest length of a path *on the surface on the torus*, which is distinct from restriction of the Euclidean metric to the surface.

(3) By putting a further warp on the bagel's surface, one may destroy all non-trivial isometries.

(4) The quotient space $\mathbb{R}^2/L$ acts on itself by isometries; every translation $T_{\mathbf{u}} \colon \mathbb{R}^2 \to \mathbb{R}^2$ descends to an isometry of $\mathbb{R}^2/L$, and $\mathbf{u}, \mathbf{u}'$ induce the same isometry of the torus if and only if $\mathbf{u}+L = \mathbf{u}'+L$. Thus, all of these quotient tori are in some sense more symmetric than X, since their isometry groups have two continuous degrees of freedom.

(5) The translation subgroup of $\text{Isom}(\mathbb{R}^2/L)$ has finite index, but the exact value of this index depends on L. Generically, it is index 2; the other coset comprises rotations by π around different centers. One exceptional case occurs when the vectors $\mathbf{v}, \mathbf{w}$ generating the lattice have the same length as each other and form an angle of $\pi/2$; in this case the lattice is the square lattice $\mathbb{Z}^2$ and $\text{Isom}(\mathbb{R}^2/L) = (\mathbb{R}^2/\mathbb{Z}^2) \rtimes D_4$. Another exceptional case occurs if $\mathbf{v}, \mathbf{w}$ have the same length and make an angle of $\pi/3$ or $2\pi/3$; in this case the lattice is triangular and the isometry group is $(\mathbb{R}^2/\mathbb{Z}^2) \rtimes D_6$.

(6) The direct product of two circles inherits a metric from $\mathbb{R}^4$, in which it is isometric to $\mathbb{R}^2/\mathbb{Z}^2$. On the other hand, the product of two other curves has only a discrete group of isometries in $\text{Isom}(\mathbb{R}^4)$. However, if we consider instead the *intrinsic* metric defined on both tori in terms of the length of the shortest curve on the surface between two points, rather than the extrinsic metric inherited from $\mathbb{R}^4$, then both tori are isometric to $\mathbb{R}^2/\mathbb{Z}^2$.

As we stressed above, all of these tori are homeomorphic to each other despite the differences in their geometry. For the tori $\mathbb{R}^2/L$ with different lattices L, this was Exercise 4.4, where a key tool was the fact that the lattices L are all isomorphic to $\mathbb{Z}^2$, so that all of the tori $\mathbb{R}^2/L$ can be obtained as the quotient space by a free and discrete action of $\mathbb{Z}^2$. This suggests that the group $\mathbb{Z}^2$ somehow plays an important role in the topology of the torus—but how? We will devote the remainder of this lecture to answering this question.

When we consider the torus $\mathbb{T}^2 = \mathbb{R}^2/L$ as a factor group, the role of $\mathbb{Z}^2$ is clear, since it sits inside $\mathbb{R}^2$ as the lattice L. However, this does not tell us how $\mathbb{Z}^2$ is related to the *intrinsic* structure of the torus $\mathbb{T}^2$—after all, every point in the integer lattice in $\mathbb{R}^2$ corresponds to the *same* point on the torus, and a single point does not have terribly much internal structure!

Another way of stating the problem is to observe that if the algebraic structure of $\mathbb{Z}^2$ characterizes some aspect of the topological structure of the torus, then we should be able to describe $\mathbb{Z}^2$ in terms of *any* torus homeomorphic to $\mathbb{R}^2/\mathbb{Z}^2$. In particular, we want to produce $\mathbb{Z}^2$ in terms of objects on the embedded torus X given in (2.15). But how do we do this?

c. Paths and loops. Thinking once more in terms of the factor space $\mathbb{R}^2/\mathbb{Z}^2$, what we want is a description of the lattice points $\mathbb{Z}^2 \subset \mathbb{R}^2$ that is able to distinguish between different points on the lattice even after we pass to the quotient space $\mathbb{R}^2/\mathbb{Z}^2$. To this end, we consider not just lattice points, but *paths* between lattice points, as in Figure 4.1.

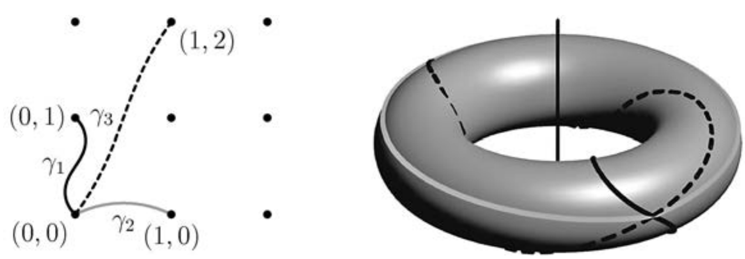

Figure 4.1. Paths in the plane and loops on the torus.

A path in $\mathbb{R}^2$ is given by a continuous function $\gamma\colon [0,1] \to \mathbb{R}^2$; such a path also defines a path on the factor torus by $\tilde{\gamma}\colon t \mapsto \gamma(t) + \mathbb{Z}^2$, and on the embedded torus by $\tilde{\gamma}(t) = f(\gamma(t))$, where f is the parametrization from (2.16). Let $\mathbf{p}$ be the point on the embedded torus that corresponds to the lattice points in $\mathbb{Z}^2$ under the map

f, and observe that if γ is a path between lattice points, then $\tilde{\gamma}$ is a *loop* on the torus based at $\mathbf{p}$—that is, it has the property that $\tilde{\gamma}(0) = \tilde{\gamma}(1) = \mathbf{p}$. Figure 4.1 shows three such paths, both as paths in $\mathbb{R}^2$ and loops on the torus.

Of course, there are many paths in $\mathbb{R}^2$ that connect a particular pair of lattice points. For example, γ_1 is only one possible path from $\mathbf{0}$ to $\mathbf{x} = (0, 1)$; a more natural choice would be $\gamma_0(t) = (0, t)$, which goes between the points along a straight line with uniform speed. These two paths are equivalent in the sense that one can be continuously deformed into the other while keeping the endpoints fixed—this visually obvious property is made precise as follows. Define a map $\Gamma\colon [0, 1] \times [0, 1] \to \mathbb{R}^2$ by

$$\Gamma(s, t) = (1 - s)\gamma_0(t) + s\gamma_1(t). \tag{4.2}$$

The map Γ has several important properties:

(1) Γ depends continuously on both s and t.

(2) $\Gamma(0, t) = \gamma_0(t)$ and $\Gamma(1, t) = \gamma_1(t)$ for all $t \in [0, 1]$.

(3) $\Gamma(s, 0) = \mathbf{0}$ and $\Gamma(s, 1) = \mathbf{x}$ for all $s \in [0, 1]$.

The cross-sections $\Gamma(s, \cdot)$ each define a path by $\gamma_s(t) = \Gamma(s, t)$. The first property above states that the paths γ_s are each continuous and that they vary continuously with s. The second property states that the family of paths γ_s connects γ_0 and γ_1—that is, it continuously deforms one into the other. Finally, the third property states every path γ_s runs from $\mathbf{0}$ to $\mathbf{x}$—that is, the endpoints are held fixed even as the rest of the curve moves. We say that γ_0 and γ_1 are *homotopic relative to* $\{\mathbf{0}, \mathbf{x}\}$.

The condition that the endpoints be held fixed is essential. Indeed, if we remove this condition, then any two paths in $\mathbb{R}^2$ can be related by a linear homotopy as in (4.2); but this homotopy does not project to the torus as a family of closed paths. One may of course consider an intermediate condition: a homotopy between loops on the torus that does not fix a point. While this condition (called *free homotopy*) makes perfect sense geometrically, classes of free homotopic paths are not easily amenable to algebraic manipulations, unlike the classes of paths homotopic relative to a point.

Given $\mathbf{x} \in \mathbb{Z}^2$ and a path γ in $\mathbb{R}^2$ with $\gamma(0) = \mathbf{0}$ and $\gamma(1) = \mathbf{x}$, let $[\gamma]$ denote the set of all paths in $\mathbb{R}^2$ that are homotopic to γ relative to $\{\mathbf{0}, \mathbf{x}\}$—that is, the set of all paths that can be continuously deformed into γ without moving their endpoints. Observe that $[\gamma]$ comprises all paths that start at $\mathbf{0}$ and end at $\mathbf{x}$, and that this gives a one-to-one correspondence between lattice points $\mathbb{Z}^2$ and equivalence classes of paths starting at $\mathbf{0}$.

Thus, we have associated the elements of the group $\mathbb{Z}^2$ to equivalence classes of paths in $\mathbb{R}^2$; we will now see that these equivalence classes are still distinguishable when we pass to the torus.

As remarked above, paths γ in $\mathbb{R}^2$ with endpoints in $\mathbb{Z}^2$ correspond to loops $\tilde{\gamma}$ on the torus—paths with $\tilde{\gamma}(0) = \tilde{\gamma}(1) = \mathbf{p}$. We can define equivalence classes just as before: two loops $\tilde{\gamma}_0$ and $\tilde{\gamma}_1$ based at $\mathbf{p}$ are homotopic relative to $\mathbf{p}$ if they can be deformed into each other via a continuous family of continuous paths, each of which is also a loop based at $\mathbf{p}$.

In $\mathbb{R}^2$, we were able to characterize $[\gamma]$ as the set of all paths from $\mathbf{0}$ with the same endpoint as γ; this no longer holds on the torus, since all lattice points are identified with the point $\mathbf{p}$. However, it is not the case that all loops on the torus are homotopic—for example, $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ in Figure 4.1 cannot be continuously deformed into each other. So what characterizes the different homotopy classes?

Heuristically, the answer is as follows (for the torus at least). Let Z denote the z-axis, and let C denote the circle in the xy-plane of radius 2 centered at the origin (the embedded torus in $\mathbb{R}^3$ is the set of all points whose distance from C is exactly 1). Observe that $\tilde{\gamma}_1$, which corresponds to the path γ_1 from $\mathbf{0}$ to $(0, 1)$, wraps around C exactly once, and Z not at all; similarly, $\tilde{\gamma}_2$, which corresponds to the path γ_2 from $\mathbf{0}$ to $(1, 0)$, wraps around Z exactly once, and C not at all. A slightly more careful look at Figure 4.1 shows that $\tilde{\gamma}_3$, which corresponds to the path γ_3 from $\mathbf{0}$ to $(1, 2)$, wraps around Z exactly once, and around C twice.

In general, if γ is a path from $\mathbf{0}$ to (a, b), then the corresponding curve $\tilde{\gamma}$ on the embedded torus wraps a times around Z and b times around C. Thus, we may think of $(1, 0)$ and $(0, 1)$, the generators of $\mathbb{Z}^2$, as representing the two "holes" in the torus: If we think of the embedded torus as a hollowed-out doughnut, then one hole (corresponding to $(1,0)$ and Z) is the "doughnut hole" through the center, and the other hole (corresponding to $(0,1)$ and C) is the hollowed-out part (where the jelly would go, perhaps).

One thing is not yet clear. We wanted to give an intrinsic description of the group $\mathbb{Z}^2$ in terms of the embedded torus; so far we have described the *elements* of the group as loops on the torus (or rather, as equivalence classes of loops), but have not specified a binary operation. There is a fairly natural candidate, though, using which we can complete the construction, and we do this in the next section.

d. The fundamental group. Consider now an arbitrary metric space X, and fix a point $p \in X$ (this will be our base point). Given any two paths $\gamma_1, \gamma_2 \colon [0,1] \to X$ with $\gamma_1(1) = \gamma_2(0)$, we can define a concatenated path $\gamma_1 \star \gamma_2$ by

$$(\gamma_1 \star \gamma_2)(t) = \begin{cases} \gamma_1(2t) & 0 \leq t \leq 1/2, \\ \gamma_2(2t-1) & 1/2 \leq t \leq 1. \end{cases} \tag{4.3}$$

That is, $\gamma_1 \star \gamma_2$ is the path that follows first γ_1 and then γ_2, moving with twice the speed of the original parametrizations so as to parametrize the entire path by the interval $[0,1]$. In particular, if γ_1 and γ_2 are *loops* from p, then $\gamma_1 \star \gamma_2$ is a loop from p as well.

We saw in the previous section that the key objects are not loops *per se*, but equivalence classes of loops. Thus, we formalize the discussion there as follows.

Definition 4.1. Let $\gamma_0, \gamma_1 \colon [0,1] \to X$ be continuous paths with $\gamma_0(0) = \gamma_1(0) = \gamma_0(1) = \gamma_1(1) = p$. We say that γ_0 and γ_1 are *homotopic relative to* p if there exists a continuous function $\Gamma \colon [0,1] \times [0,1] \to X$ such that

(1) $\Gamma(0,t) = \gamma_0(t)$ and $\Gamma(1,t) = \gamma_1(t)$ for all $0 \leq t \leq 1$.

(2) $\Gamma(s,0) = \Gamma(s,1) = p$ for all $0 \leq s \leq 1$.

In this case we write $\gamma_0 \sim \gamma_1$. The set of all loops from p that are homotopic to γ relative to p is called the *homotopy class* of γ, and is denoted $[\gamma]$.

The binary operation of concatenation works not just on loops, but on homotopy classes of loops: given loops γ and η, we define $[\gamma] \star [\eta]$ to be the homotopy class $[\gamma \star \eta]$. We must check that this is well defined, but once we do so, we will finally have in our hands the fundamental object of algebraic topology.

Definition 4.2. Given a metric space X and a point $p \in X$, the *fundamental group* of X with base point p is the collection of homotopy classes of loops based at p together with the binary operation of concatenation. We denote this group by $\pi_1(X, p)$.

Of course, this terminology puts the cart before the horse. Right now all we have is a set together with a binary operation (which may not even be well defined, for all we know). Why is this a group?

Proposition 4.3. *The binary operation $\star$ is well defined on $\pi_1(X, p)$ and makes it into a group.*

Proof. We first show that $\star$ is well defined—that is, that $\gamma_1 \star \eta_1 \sim \gamma_2 \star \eta_2$ whenever $\gamma_1 \sim \gamma_2$ and $\eta_1 \sim \eta_2$. An equivalent way of stating this condition is that the equivalence class $[\gamma \star \eta]$ is the same no matter which representatives of $[\gamma]$ and $[\eta]$ we work with.

The proof of this is straightforward: If Γ and H are homotopies demonstrating $\gamma_1 \sim \gamma_2$ and $\eta_1 \sim \eta_2$, respectively, we can concatenate them to obtain a homotopy between $\gamma_1 \star \eta_1$ and $\gamma_2 \star \eta_2$. To wit, define a continuous function $G\colon [0,1] \times [0,1] \to X$ as follows:

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

One may easily verify that G is the required homotopy.

The remaining elements of the proof are essentially contained in Figure 4.2(a)–(c). Let us explain this claim.

Once we know $\star$ is well-defined, we must show that it is associative. The lines inside the square $[0,1] \times [0,1]$ in part (a) of the figure are level sets of the function G—that is, values of s and t which G sends to the same point in X. The bottom edge of part (b) represents $(\gamma_1 \star \gamma_2) \star \gamma_3$, the curve which traverses γ_1 from $t = 0$ to $t = 1/4$, then γ_2 from $t = 1/4$ to $t = 1/2$, and finally γ_3 from $t = 1/2$ to $t = 1$. The

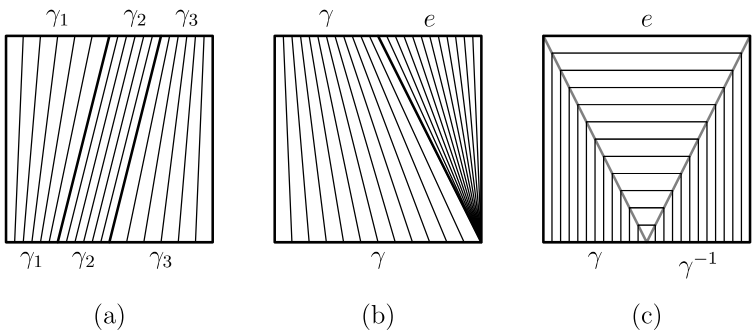

Figure 4.2. Homotopy equivalences that make $\pi_1(X)$ a group.

top edge represents $\gamma_1 \star (\gamma_2 \star \gamma_3)$, for which the points traversed are the same, but the parametrization is different. Using the piecewise linear homotopy

$$G(s,t) = \begin{cases} \gamma_1((s+1)t) & 0 \le t \le \frac{s+1}{4}, \\ \gamma_2(t-(s+1)/4) & \frac{s+1}{4} \le t \le \frac{s+2}{4}, \\ \gamma_3(1-(s+1)t) & \frac{s+2}{4} \le t \le 1, \end{cases}$$

we see that $[(\gamma_1 \star \gamma_2) \star \gamma_3] = [\gamma_1 \star (\gamma_2 \star \gamma_3)]$, and hence $\star$ is associative. The lines in Figure 4.2(a) correspond to values of s and t which G sends to the same place in X.

Observe that G does not change the geometry of the above paths at all—indeed, it is nothing more than a reparametrization! This is an important special case of homotopy equivalence, and is also what we need in order to satisfy the next group axiom, the existence of an identity element. The natural candidate for the identity element in the fundamental group $\pi_1(X,p)$ is the trivial loop $e\colon [0,1] \to X$, for which $e(t) = p$ for all $0 \le t \le 1$. Concatenating any loop γ with e does not change its geometry, and the simple piecewise linear reparametrization shown in Figure 4.2(b) suffices to show that $[\gamma] \star [e] = [\gamma \star e] = [\gamma]$ for all loops γ, and similarly $[e] \star [\gamma] = [\gamma]$.

Reparametrization is *not* enough to get us the final group axiom, the existence of inverse elements. Indeed, as soon as a loop γ is nontrivial and goes to points other than p, it cannot be a reparametrization of the trivial loop. Rather, a genuine homotopy is required; the key is that we consider loops not just as geometric objects (the image $\gamma([0,1])$), but we also record the "history" of movement along the path. Thus, the inverse γ^{-1} ought to be the loop which "undoes" γ, so we write $\gamma^{-1}(t) = \gamma(1-t)$ to obtain a loop that traverses the same curve as γ, but does so in the reverse direction.

To show that $\gamma \star \gamma^{-1} \sim e$, we use the homotopy shown in Figure 4.2(c), which may be given the following explicit form:

$$G(s,t) = \begin{cases} \gamma(t) & 0 \le t \le \frac{1-s}{2}, \\ \gamma(\frac{1-s}{2}) = \gamma^{-1}(\frac{1+s}{2}) & \frac{1-s}{2} \le t \le \frac{1+s}{2}, \\ \gamma^{-1}(t) & \frac{1+s}{2} \le t \le 1. \end{cases}$$

The path $G(s,\cdot)$ follows γ as far as $\gamma((1-s)/2)$, then stops and thinks about things for a while, and finally retraces its steps to end where it began, at p. As s goes from 0 to 1, the amount of γ that $G(s,\cdot)$ traverses gets smaller and smaller, until finally $G(1,\cdot)$ is just the trivial loop e. This homotopy establishes that $[\gamma] \star [\gamma^{-1}] = [e]$, and hence $\pi_1(X,p)$ is indeed a group. □

We have successfully produced a group from the intrinsic topological data of X. However, several questions remain. The definition involves an arbitrarily chosen point p; what happens if we choose a different point p as our base point? Do we get a different group? What does this group look like for familiar examples, such as the circle, the sphere, the plane, the torus, etc.? Part of our motivation was to recover the group $\mathbb{Z}^2$ from the intrinsic properties of the torus—did it work? Or is $\pi_1(\mathbb{T}^2, \mathbf{p})$ something else?

We will defer specific examples until the next lecture; for now we address the first question, and consider the groups $\pi_1(X,p)$ and $\pi_1(X,q)$ for points $p \neq q \in X$.

Definition 4.4. A metric space X is *path-connected* if for every $p, q \in X$ there exists a continuous path $\gamma\colon [0,1] \to X$ such that $\gamma(0) = p$ and $\gamma(1) = q$.

Proposition 4.5. *If X is a path-connected metric space, then $\pi_1(X,p)$ and $\pi_1(X,q)$ are isomorphic for any $p,q\in X$.*

Proof. Given $p,q\in X$, let $\alpha\colon [0,1]\to X$ be a continuous path such that $\alpha(0)=p$ and $\alpha(1)=q$. Define a map $\varphi\colon \pi_1(X,p)\to\pi_1(X,q)$ by $\varphi([\gamma])=[\alpha^{-1}\star\gamma\star\alpha]$. The proof that φ is well-defined exactly mirrors the proof for $\star$ in Proposition 4.3. Furthermore, φ is a homomorphism, since

$$\begin{aligned}\varphi([\gamma]\star[\eta]) &= [\alpha^{-1}\star\gamma\star\eta\star\alpha]\\ &= [\alpha^{-1}\star\gamma\star\alpha\star\alpha^{-1}\star\eta\star\alpha]=\varphi([\gamma])\star\varphi([\eta]),\end{aligned}$$

where the second equality uses the fact that $\alpha\star\alpha^{-1}\sim e_p$, and that $\alpha^{-1}\star\gamma\star e_p\star\eta\star\alpha$ is a reparametrization of $\alpha^{-1}\star\gamma\star\eta\star\alpha$ (see Figure 4.3).

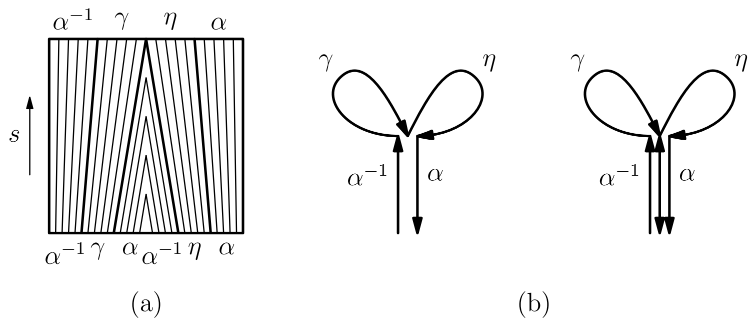

Figure 4.3. Changing base points is an isomorphism.

Now we observe that φ is onto, since φ^{-1} can be defined by $[\zeta]\mapsto[\alpha\star\zeta\star\alpha^{-1}]$ for every $[\zeta]\in\pi_1(X,q)$. Furthermore, $\varphi([\gamma])=[e_q]$ implies $\gamma\sim\alpha\star e_q\star\alpha^{-1}\sim e_p$, and so φ is one-to-one. It follows that φ is an isomorphism. □

As a consequence of Proposition 4.5, we can (and will) speak of the fundamental group of X, and write $\pi_1(X)$, without explicitly mentioning the base point, since changing the base point yields an isomorphic group.

e. Algebraic topology. In algebraic topology one associates to various kinds of topological spaces algebraic objects, usually groups, moduli or rings. The fundamental group we just described is a premiere and arguably most geometrically transparent example of such an association. Two leading principles of algebraic topology are *invariance* and *functoriality.*

Invariance requires that equivalent spaces are associated with isomorphic objects and that the association is independent of auxiliary elements involved in the construction of an algebraic object. We already have an example in the case of fundamental group of a path connected space: The construction does not depend on the base point used and homeomorphic spaces have isomorphic fundamental groups.

Functoriality in its simplest form requires that continuous maps between spaces "naturally" induce homomorphisms between the associated algebraic object. The direction of this homomorphism may be the same as for the map (*covariant* constructions) or the opposite (*contravariant* constructions)—the fundamental group is an example of the former. Furthermore, those homomorphisms should behave properly under the composition of maps.

Proposition 4.6. *Let* $f\colon X \to Y$ *be a continuous map and* $p \in X$*. Then for* $[\gamma] \in \pi_1(X,p)$ *the path* $f \circ \gamma : [0,1] \to Y$ *defines an element* $f_*([\gamma]) \in \pi_1(Y, f(p))$ *and* $f_*\colon \pi_1(X,p) \to \pi_1(Y, f(p))$ *is a group homomorphism. Moreover, given* $g\colon Y \to Z$*, the composition satisfies* $(g \circ f)_* = g_* \circ f_*$*.*

Proof. Since composition of a path homotopy in X with a continuous map is a path homotopy in Y the map f_* is correctly defined. Concatenation of paths goes into concatenation of their images, hence the map f_* is a homomorphism. The last statement is obvious since it is true already at the level of paths. □

Lecture 24. Computation of π_1 for some examples

a. Homotopy equivalence and contractible spaces. The notion of homotopy from the previous lecture can be applied not just to paths, but to *any* continuous maps. Given two metric spaces X

and Y, we say that two continuous maps $f, g\colon X \to Y$ are *homotopic* if there exists a continuous function $\Gamma\colon [0,1] \times X \to Y$ such that $\Gamma(0,x) = f(x)$ and $\Gamma(1,x) = g(x)$ for all $x \in X$. Heuristically, this means that the functions $\Gamma(s,\cdot)\colon X \to Y$ are a continuous one-parameter family of continuous maps that deform f into g.

This is the notion of *absolute* homotopy; observe that we place no restrictions on the functions $\Gamma(s,\cdot)$, in contrast to the previous lecture, where we required each of the paths $\Gamma(s,\cdot)$ to have endpoints at a fixed base point. Even though we later showed that the isomorphism class of the fundamental group is independent of the choice of base point, this base point still plays a prominent role in the definitions. This is emblematic of many topological constructions: In order to define a very general object, one must use definitions which in and of themselves depend on an arbitrary choice, but in the end the objects so defined are independent of which particular choice is made.

For the fundamental group, the "particular choice" is a choice of base point, which appears in the definitions via the notion of *relative* homotopy. Given two continuous maps $f, g\colon X \to Y$ and a subset $A \subset X$, we say that f and g are *homotopic relative to* A if there exists a continuous homotopy $\Gamma\colon [0,1] \times X \to Y$ with the properties above, along with the additional property that $\Gamma(s,x) = f(x) = g(x)$ for all $s \in [0,1]$ and $x \in A$. Thus, relative homotopy is a matter of continuously deforming the map f into the map g, while keeping the action of the map $\Gamma(s,\cdot)$ on the set A fixed; in the previous lecture, we used homotopy relative to the set of endpoints $A = \{0,1\}$.

Once we have a definition of homotopy for maps, it is natural to ask what the possible homotopy classes of maps from X to Y are. For example, if $X = Y = S^1$, then it is intuitively clear that the homotopy class of $f\colon S^1 \to S^1$ is the set of all maps that "wind around the circle the same number of times as f does". We will make this precise shortly.

In the meantime, we note that given *any* metric space X, there are two natural types of maps from X to itself. One is the identity map, $\mathrm{Id}\colon x \to x$, and the other is the trivial (or constant) map $e_p\colon x \to p$, where p is some arbitrarily chosen point in X. Thus, Id fixes every point in X, while e_p collapses all of X to a single point. We say that

X is *contractible to the point* p if these two maps are homotopic—that is, if there exists a continuous map $\Gamma\colon [0,1] \times X \to X$ such that $\Gamma(0,x) = x$ and $\Gamma(1,x) = p$ for all $x \in X$.

Proposition 4.7. *Given any two points $p, q \in X$, the space X is contractible to p if and only if X is contractible to q.*

Proof. First notice that if X is contractible to a point p it is path connected, since for any $q \in X$ the homotopy $\Gamma(t,q)$ is a path connecting q with p. Combining the contraction to p with this path in the opposite direction, i.e., $\Gamma(1-t,q)$, gives a contraction of X to q. □

Thanks to Proposition 4.7, we may simply refer to X as being *contractible* without mentioning which point it is contractible to, since if it is contractible to one point, then it is contractible to any point. This is another example of a general property that must be defined with reference to an arbitrarily chosen object, whose precise choice turns out not to matter.

Example 4.8. $\mathbb{R}^n$ is contractible: consider the homotopy $\Gamma(s,\mathbf{x}) = (1-s)\mathbf{x}$. We have $\Gamma(0,\cdot) = \mathrm{Id}$ and $\Gamma(1,\cdot) = e_{\mathbf{0}}$. Similarly, any open or closed ball in $\mathbb{R}^n$ is contractible: Given $\mathbf{p} \in \mathbb{R}^n$ and $r > 0$, the identity map on the closed ball $X = \{\mathbf{x} \in \mathbb{R}^n \mid d(\mathbf{p},\mathbf{x}) \le r\}$ can be homotoped to the trivial map by

$$\Gamma(s, \mathbf{p} + \mathbf{x}) = (1-s)\mathbf{x} + \mathbf{p}. \tag{4.4}$$

In fact, this gives a broad class of contractible spaces: We say that $X \subset \mathbb{R}^n$ is *star-shaped* if there exists $\mathbf{p} \in X$ such that the line segment from $\mathbf{p}$ to $\mathbf{x}$ is contained in X for every $\mathbf{x} \in X$. If X is star-shaped, then (4.4) gives a homotopy between Id_X and $e_{\mathbf{p}}$.

Remark. The fact that open balls are contractible emphasizes the fact that for “nice” spaces that look like Euclidean spaces at a local level (such spaces are called *manifolds*), any two paths that are in the same local neighborhood are homotopic; in particular, for this class of spaces, homotopy is really a *global* theory, which captures large-scale properties of spaces and maps.

Another class of contractible spaces appears as follows.

Definition 4.9. A *graph* is a finite or countable collection of vertices (which we may think of as lying in $\mathbb{R}^n$) together with a collection of edges joining certain pairs of vertices. A *cycle* is a collection of edges that forms a closed loop, and a graph without cycles is a *tree*.

Proposition 4.10. *Every finite tree is contractible.*

Proof. Use induction in the number of edges. The tree with zero edges is a point and hence contractible. Now let $\mathcal{T}$ be a tree with n edges. Removing an edge e makes the rest of the tree disconnected; otherwise the endpoints of e could be connected in $\mathcal{T} \setminus e$ and adding e would give a cycle. Thus, $\mathcal{T}$ with e removed is the union of two disjoint trees, each having fewer than n edges. By inductive hypothesis each of the two parts can be contracted to the corresponding endpoint of the removed edge. Combining these contractions with a contraction of the edge to a single point completes the argument. $\square$

As it turns out, countable trees are also contractible. This will be proven later on, when we study certain graphs as geometric objects related to certain groups.

Example 4.11. Let $X = \mathbb{R}^2 \setminus \{\mathbf{0}\}$. Then as we will shortly see, X is *not* contractible.

We have observed that homeomorphic spaces have the same fundamental group; however, there is a weaker condition under which two spaces must have the same fundamental group. The condition that X and Y be homeomorphic may be stated as the existence of maps $f\colon X \to Y$ and $g\colon Y \to X$ such that $f \circ g = \mathrm{Id}_Y$ and $g \circ f = \mathrm{Id}_X$. Since the fundamental group is stated not in terms of paths but rather of homotopy classes of paths, it makes sense to weaken these equalities to homotopic equivalences.

Definition 4.12. Two metric spaces X and Y are *homotopic* (or *homotopy equivalent*) if there exist maps $f\colon X \to Y$ and $g\colon Y \to X$ such that $f \circ g \sim \mathrm{Id}_Y$ and $g \circ f \sim \mathrm{Id}_X$.

Example 4.13. Any contractible space is homotopic to a point; to see this, let X be contractible, fix a point $p \in X$, and let $Y = \{p\}$.

Then defining $f\colon X \to \{p\}$ by $e_p\colon x \to p$ and $g\colon \{p\} \to X$ as the inclusion map $g(p) = p$, we see that $f \circ g = \mathrm{Id}_Y$ and $g \circ f = e_p \sim \mathrm{Id}_X$, where the last statement follows from the definition of contractibility.

Example 4.14. Writing S^1 for the circle $\{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x}\| = 1\}$, we see that the punctured plane $\mathbb{R}^2 \setminus \{\mathbf{0}\}$ is homotopic to S^1. Indeed, we may let $f\colon \mathbb{R}^2 \setminus \{\mathbf{0}\} \to S^1$ be the radial projection $f(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ and $g\colon S^1 \to \mathbb{R}^2 \setminus \{\mathbf{0}\}$ be the inclusion map $g(\mathbf{x}) = \mathbf{x}$. Then $f \circ g = \mathrm{Id}_{S^1}$, and $g \circ f$ is homotopic to $\mathrm{Id}_{\mathbb{R}^2\setminus\{\mathbf{0}\}}$ via the linear homotopy

$$\Gamma(s, \mathbf{x}) = s\mathbf{x} + (1 - s)\frac{\mathbf{x}}{\|\mathbf{x}\|}.$$

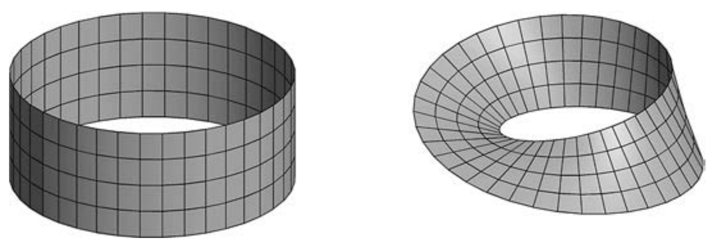

Figure 4.4. The cylinder and the Möbius strip

Similarly, one may show that both the cylinder and the Möbius strip, which are shown in Figure 4.4, are homotopic to the circle, and hence to each other (since homotopy is an equivalence relation), despite the fact that they are not homeomorphic (and indeed, their topologies are quite different in many other ways).

We already saw that homeomorphic spaces have isomorphic fundamental groups. In fact, the fundamental group is invariant under the weaker condition of homotopy equivalent.

Proposition 4.15. *If X and Y are homotopically equivalent, then $\pi_1(X)$ and $\pi_1(Y)$ are isomorphic.*

Proof. Let $f\colon X \to Y$ and $g\colon Y \to X$ establish the homotopy equivalence. Proposition 4.6 implies that the maps f_* and g_* are homomorphisms, and so it suffices to check that f_* and g_* are bijections.

We do this by showing that $g_*: f_*: \pi_1(X,p) \to \pi_1(X, g(f(p)))$ is a bijection.

Since f and g establish the homotopy equivalence, there is a homotopy $\Gamma: [0,1] \times X \to X$ such that $\Gamma(0,x) = x$ and $\Gamma(1,x) = g(f(x))$. Thus, if $\gamma: [0,1] \to X$ is any path, we see that $\tilde{\Gamma}: [0,1] \times [0,1] \to X$ given by $\tilde{\Gamma}(s,t) = \Gamma(s,\gamma(t))$ is a homotopy from γ to $g(f(\gamma))$ relative to the endpoints. It follows that a loop γ is null-homotopic if and only if the loop $g(f(\gamma))$ is null-homotopic, and hence $g_* \circ f_*$ is a bijection. □

b. The fundamental group of the circle. While we proved that certain spaces are contractible by explicit constructions of contractions, now we are ready to show for the first time that a space is not contractible. We will describe the homotopy classes of maps from S^1 to itself, which also lets us compute its fundamental group $\pi_1(S^1)$. We need to formalize the notion of a map $f: S^1 \to S^1$ as "wrapping the circle around itself".

To do this, we recall from Examples 1.66 and 2.29 that the circle S^1 can also be obtained as the factor space $\mathbb{R}/\mathbb{Z}$. Thus, any continuous map $F: \mathbb{R} \to \mathbb{R}$ projects to a map $f: \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ provided this projection is well-defined—that is, provided $F(x) - F(y) \in \mathbb{Z}$ whenever $x - y \in \mathbb{Z}$. Furthermore, for any such map, the quantity $F(x+1) - F(x)$ varies continuously in x and takes integer values, and hence is independent of x; it is called the *degree* of the map f and is denoted by $\deg f$. We may think of the degree as the number of times f wraps the circle around itself.

Does it go in the other direction? Do we get *every* map of the circle this way? That is, given a continuous map $f: \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$, can we produce a continuous map $F: \mathbb{R} \to \mathbb{R}$ such that the following diagram commutes?

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{F} & \mathbb{R} \\ \big\downarrow{\scriptstyle\pi} & & \big\downarrow{\scriptstyle\pi} \\ \mathbb{R}/\mathbb{Z} & \xrightarrow{f} & \mathbb{R}/\mathbb{Z} \end{array} \tag{4.5}$$

Here $\pi: \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ is the natural projection $\pi(x) = x + \mathbb{Z}$.

It turns out that such a map F does indeed exist; we call this the *lift* of f. To produce F, we begin by specifying $F(0)$ as any element of $f(0+\mathbb{Z})$. Once this is done, the requirement that F be continuous specifies it uniquely; by fixing $\epsilon > 0$ small enough, we can guarantee that $\pi^{-1}(f((-\epsilon,\epsilon)))$ is the disjoint union of small intervals around the elements of $f(0+\mathbb{Z})$, each of which is sent homeomorphically to the interval $f((-\epsilon,\epsilon))$ by π. Our choice of $F(0)$ specifies $F(y)$ uniquely for $y \in (-\epsilon,\epsilon)$ as the element of $f(y+\mathbb{Z})$ that lies nearest to $F(0)$. Continuing in this manner, we can define F on $(-2\epsilon, 2\epsilon)$, $(-3\epsilon, 3\epsilon)$, and so on.

As before, once we have a lift $F\colon \mathbb{R} \to \mathbb{R}$, the value $F(x+1) - F(x)$ is independent of the choice of x. Furthermore, every other lift of f is of the form $F + n$ for some integer n, and thus has the same degree. Thus, we may legitimately define the degree of f by taking any lift F and computing $F(1) - F(0)$. Once again we see the phenomenon where the precise choice of the lift F and the reference point 0 to determine the degree is irrelevant, although *some* choice was necessary.

Proposition 4.16. *If $f\colon S^1 \to S^1$ and $g\colon S^1 \to S^1$ are homotopic, then* $\deg f = \deg g$.

Proof. We need the following fact.

Exercise 4.5. Show that if $f, g\colon S^1 \to S^1$ are homotopic, then any choice of lifts $F, G\colon \mathbb{R} \to \mathbb{R}$ are homotopic as well.

Let Γ be a homotopy between F and G as in the exercise, and observe that $\Gamma(s,\cdot)$ varies continuously in s, so $\Gamma(s,1) - \Gamma(s,0)$ varies continuously in s as well. Since it takes integer values, it must be constant. □

We can easily define a circle map with any given degree: For any $n \in \mathbb{Z}$, let $E_n\colon S^1 \to S^1$ be the linear map $E_n(x+\mathbb{Z}) = nx + \mathbb{Z}$—that is, E_n is the projection of the map $x \mapsto nx$ from the real line onto the circle. In fact, from the point of view of homotopy, these maps are all there is.

Proposition 4.17. *Every circle map of degree n is homotopic to E_n.*

Proof. Let $f\colon S^1 \to S^1$ have degree n, and let $F\colon \mathbb{R} \to \mathbb{R}$ be its lift to the real line. Consider the linear homotopy

$$\Gamma(s,x) = (1-s)F(x) + snx, \tag{4.6}$$

and observe that $\Gamma(0,x) = F(x)$ and $\Gamma(1,x) = nx$. Furthermore, we have

$$\begin{aligned}\Gamma(s,x+1) &= (1-s)F(x+1) + sn(x+1)\\ &= (1-s)(F(x)+n) + snx + sn = \Gamma(s,x) + n,\end{aligned}$$

and so $\Gamma(s,\cdot)\colon \mathbb{R} \to \mathbb{R}$ projects to a well-defined continuous map $\gamma(s,\cdot)\colon \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$. Since $\gamma(0,x+\mathbb{Z}) = f(x+\mathbb{Z})$ and $\gamma(1,x+\mathbb{Z}) = nx+\mathbb{Z}$, we see that γ is the desired homotopy. □

Corollary 4.18. *The fundamental group of the circle is* $\pi_1(S^1) = \{[E_n]\} \cong \mathbb{Z}$*, where the group operation is* $[E_n] \star [E_m] = [E_{n+m}]$.

Proof. A loop in X with base point p can be written as a continuous map $S^1 = \mathbb{R}/\mathbb{Z} \to X$ which maps $0+\mathbb{Z}$ to p. Taking $p = 0+\mathbb{Z} \in S^1$, we see that E_n has this property as well, and so any loop in S^1 of degree n with base point $0+\mathbb{Z}$ is homotopic to E_n via the homotopy coming from (4.6). □

Remark. In the end, this result is purely topological, and applies to any space homotopic to the circle—a punctured plane, a Möbius strip, a cylinder, etc. However, in order to prove it, we found it beneficial to consider a very particular representative from this homotopy class—namely, the factor circle $\mathbb{R}/\mathbb{Z}$, which carries an extra algebraic structure that was essential in the proof. Such a course of action is not infrequent in topology.

c. Tori and spheres. Upon observing that the torus $\mathbb{R}^2/\mathbb{Z}^2$ is the direct product of two copies of S^1, we can finally complete our description of the fundamental group of the torus, using the following result.

Theorem 4.19. *Let X and Y be path-connected metric spaces. Then* $\pi_1(X \times Y) \cong \pi_1(X) \times \pi_1(Y)$.

Proof. Fix base points $x_0 \in X$ and $y_0 \in Y$, and let $P_X : (x, y) \mapsto x$ and $P_Y : (x, y) \mapsto y$ be the natural projections from $X \times Y$ to X and Y, respectively.

Now if γ_X and γ_Y are loops in X and Y with base points x_0 and y_0, then they determine a unique loop in $X \times Y$ with base point (x_0, y_0) by

$$\gamma(t) = (\gamma_X(t), \gamma_Y(t)). \tag{4.7}$$

Conversely, every loop γ in $X \times Y$ based at (x_0, y_0) determines loops in X and Y based at x_0 and y_0 by the projections $P_X(\gamma)$ and $P_Y(\gamma)$. This map also works for homotopies, so it defines a map $\pi_1(X \times Y) \to \pi_1(X) \times \pi_1(Y)$; similarly, the map (4.7) defines a map $\pi_1(X) \times \pi_1(Y) \to \pi_1(X \times Y)$. Writing down definitions in a straightforward way one sees that these maps are homomorphisms and are inverses of each other, which proves the result. □

Corollary 4.20. *The n-dimensional torus $\mathbb{R}^n/\mathbb{Z}^n$ has fundamental group $\mathbb{Z}^n$.*

Corollary 4.20 is a concrete example of an important general scheme: Many interesting spaces are obtained as X/G, where X is a topological space and G is a group acting on X. We will see in the next lecture that for "simple" enough X, and a "nice" action of G one can obtain $\pi_1(X/G) = G$.

Definition 4.21. If $\pi_1(X)$ is trivial, we say that X is *simply connected.*

Obviously, every contractible space is simply connected. The converse fails, however: being contractible is a stronger property than being simply connected. A simplest but fundamental example is provided by the spheres. To see this, consider the sphere S^2. The sphere is not contractible (this looks clear intuitively but requires a proof that will be given later), but has trivial fundamental group.

Proposition 4.22. *The sphere S^2 is simply connected.*

Proof. We observe that if x is any point on the sphere, then $S^2 \setminus \{x\}$ is homeomorphic to $\mathbb{R}^2$ via stereographic projection (see Figure 3.2). Since $\mathbb{R}^2$ is contractible, any curve on $\mathbb{R}^2$ is homotopic to a point and, in particular, any loop $\gamma\colon S^1 \to S^2$ which misses a point (that is, $\gamma(S^1) \neq S^2$) can be homotoped to a point by using stereographic projection from a point $x \in S^2 \setminus \gamma(S^1)$.

However, one must deal with the fact that there are continuous and surjective functions $\gamma\colon S^1 \to S^2$—these so-called *Peano curves* cannot be immediately dealt with in the above fashion. They turn out not to cause too much trouble, as any curve γ is homotopic to a piecewise smooth approximation. In the plane, this can be seen by letting $\gamma\colon [0,1] \to \mathbb{R}^2$ be any curve and considering the piecewise linear approximations γ_n that are defined by the property $\gamma_n(k/n) = \gamma(k/n)$ for integers $0 \leq k \leq n$, and are linear in between these points. We have $\gamma_n \sim \gamma$, and a similar construction works on the sphere, replacing line segments with arcs of great circles. Since the curves γ_n *cannot* cover the entire sphere (being piecewise smooth), this suffices to show that $\pi_1(S^2) = \{e\}$. □

Remark. This argument extends straightforwardly to higher dimensions, and one obtains that for every $n \geq 2$, the fundamental group of the n-sphere is trivial: $\pi_1(S^n) = \{e\}$.

d. Abelian fundamental groups. In general, fundamental groups can have a very complicated algebraic structure. However, there is one instance worth noting in which this structure simplifies significantly, and the fundamental group $\pi_1(X)$ turns out to be abelian. This occurs when the space X is not just a topological space, but carries a group structure as well.

Definition 4.23. A *metrizable topological group* is a metrizable space G equipped with a binary operation making it a group, and with the additional property that multiplication $(g,h) \mapsto gh$ and inversion $g \mapsto g^{-1}$ are continuous.

Theorem 4.24. *Let G be a metrizable path-connected topological group. Then $\pi_1(G)$ is abelian.*

Proof. We take the identity element e as our base point, and consider two loops $\alpha, \beta\colon [0,1] \to G$ with $\alpha(0) = \beta(0) = \alpha(1) = \beta(1) = e$. We must show that $\alpha \star \beta \sim \beta \star \alpha$ by using group multiplication in G to produce a homotopy.

Using the fact that α and β take the value e at the endpoints of the interval, we observe that

$$\alpha \star \beta(t) = \begin{cases} \alpha(2t)\beta(0) & 0 \leq t \leq 1/2, \\ \alpha(0)\beta(2t-1) & 1/2 \leq t \leq 1, \end{cases}$$

and a similar formula holds for $\beta \star \alpha$. Observe that if $\Gamma(s,t)$ is the desired homotopy—that is, $\Gamma(0,t) = \alpha \star \beta(t)$ and $\Gamma(1,t) = \beta \star \alpha(t)$—then for $0 \leq t \leq 1/2$, we must have

$$\Gamma(s,t) = \begin{cases} \alpha(2t)\beta(0) & s = 0, \\ \alpha(0)\beta(2t) & s = 1, \end{cases} \tag{4.8}$$

and for $1/2 \leq t \leq 1$,

$$\Gamma(s,t) = \begin{cases} \alpha(1)\beta(2t-1) & s = 0, \\ \alpha(2t-1)\beta(1) & s = 1. \end{cases} \tag{4.9}$$

It is now easy to see that the following homotopy works:

$$\Gamma(s,t) = \begin{cases} \alpha((1-s)(2t))\beta(s(2t)) & 0 \leq t \leq 1/2, \\ \alpha(s(2t-1)+1-s)\beta((1-s)(2t-1)+s) & 1/2 \leq t \leq 1. \end{cases}$$

One need only observe that Γ satisfies (4.8) and (4.9), is continuous, and has $\Gamma(s,0) = \Gamma(s,1) = e$ for all $0 \leq s \leq 1$. □

Lecture 25. Fundamental group of a bouquet of circles

a. Covering of bouquets of circles. Now we extend the example of $\pi_1(S^1)$ in a different direction. Let $B_n(S^1)$ denote the "bouquet" of n circles shown in Figure 4.5. What is the fundamental group of these spaces? We will focus our attention on the case $n = 2$, where we get the figure-eight shape shown in Figure 4.5(a), as the situation for larger values of n is analogous.

The case $n = 1$ is just the circle S^1, where the key to deciphering the fundamental group was to classify curves in terms of how often they looped around the circle. Thus, we expect a similar classification to be important here, and indeed, given a loop γ, we may profitably ask what the degree of γ is on each "leaf" of the bouquet (to mix our metaphors a little). However, we soon find that this is not quite

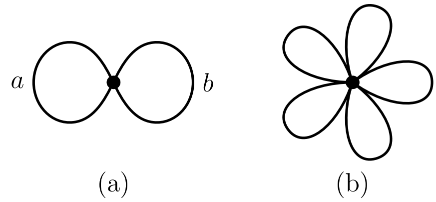

Figure 4.5. Bouquets of circles.

sufficient for a complete understanding. Labeling the leaves of the figure-eight as shown in Figure 4.5(a), we write a for the loop that goes around the left-hand leaf in the clockwise direction, and a^{-1} for the loop that goes around it counterclockwise, and similarly for b and b^{-1}. Then the loop $\gamma = a \star b \star a \star b^{-1}$ has degree 0 around both leaves, but is *not* homotopic to the identity.

This indicates that the fundamental group in this case is more complicated; in particular, it is non-abelian. The following exercise suggests a sense in which it is "larger" than $\mathbb{Z}^2$.

Exercise 4.6.

(1) Consider the embedding f of figure eight $B_2(S^1)$ into the torus with one circle going along the "parallel" and the other along the "meridian" of the torus (corresponding more or less to the curves γ_1 and γ_2, respectively, in Figure 4.1). Show that the induced homomorphism of the fundamental groups $f_*\colon \pi_1(B_2(S^1)) \to \pi_1(\mathbb{T}^2)$ is surjective.

(2) Generalize this construction and the statement to $B_n(S^1)$.

To say more, we recall that we have used three different techniques to compute the fundamental group of a space X:

(1) Show that $\pi_1(X)$ is trivial by showing that any loop can be contracted to a point.

(2) In the case $X = S^1$, use the fact that we can lift loops to paths in $\mathbb{R}$ to define the *degree* of a loop, and show that this defines an isomorphism between $\pi_1(S^1)$ and $\mathbb{Z}$.

(3) Show that X is homotopic to a space whose fundamental group is known, or obtain X as the direct product of such spaces.

For the figure eight $X = B_2(S^1)$, the first and the third methods are useless (Exercise 4.6 is about as far as we can go in the latter direction),[47] and so we must look more closely at the second. As we did for the circle, we want to exhibit a standard family of mutually non-homotopic loops in $B_2(S^1)$ that carries a clear group structure, and which is universal in the sense that every loop in $B_2(S^1)$ is homotopic to something from this family.

The first step in obtaining this family for the circle was to use the fact that the circle is a factor space $\mathbb{R}/\mathbb{Z}$, and that loops on the circle (and homotopies between them) can be lifted to paths (and homotopies) in $\mathbb{R}$. The standard projection $\pi\colon \mathbb{R} \to S^1 = \mathbb{R}/\mathbb{Z}$ can be written in a number of different forms.

(1) If we think of the circle as the interval $[0, 1)$, where the missing endpoint 1 is identified with 0, then $\pi(x) = x \pmod 1$.
(2) If we think of the circle as the factor space $\mathbb{R}/\mathbb{Z}$, so that points on the circle are equivalence classes in $\mathbb{R}$, then $\pi(x) = x + \mathbb{Z}$.
(3) If we think of the circle as the unit circle in $\mathbb{C}$, then $\pi(x) = e^{2\pi i x}$.

Whichever model of the circle we use, the key property of the projection π that we used in Lecture 24(b) was that around each $x \in S^1$ there is a small neighborhood U such that $\pi^{-1}(U)$ is a disjoint union of neighborhoods around the elements of $\pi^{-1}(x)$, and that π is a homeomorphism from each of these neighborhoods to U. This is formalized by the following definition, illustrated in Figure 4.6.

Definition 4.25. Let C and X be metric spaces, and suppose that $\rho\colon C \to X$ is a continuous map such that for every $x \in X$, there exists a neighborhood $U \ni x$ such that $\rho^{-1}(U)$ is a (finite or countable) union of disjoint neighborhoods in C on each of which ρ is homeomorphic—that is, $\rho^{-1}(U) = \bigcup_i V_i$, where the V_i are disjoint open sets in X, and $\rho\colon V_i \to U$ is a homeomorphism for all i. Then ρ is called a *covering map*, and C is a *covering space* of X.

Remark. The number of connected components of the preimage $\rho^{-1}(U)$ takes discrete values and varies continuously in x. Thus,

[47] In fact, there are results such as the *Siefert–van Kampen Theorem* that let us deduce π_1 for constructions other than the direct product, such as gluing two spaces together at a common point, but these lie beyond the scope of this book.

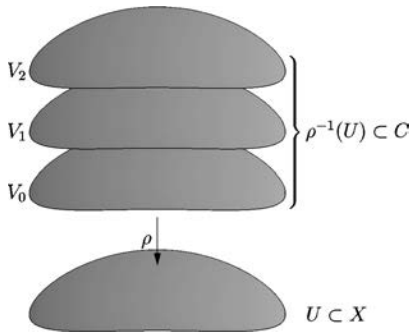

Figure 4.6. A local view of a covering map.

it is locally constant, and hence constant everywhere if X is path connected, the only case to be considered henceforth.

Example 4.26. The natural projection $S^2 \to \mathbb{R}P(2)$ that takes $\mathbf{x}$ to $\{\mathbf{x}, -\mathbf{x}\}$ is also a covering map. In this case, however, each point has only two preimages, rather than countably many, as is the case for the circle.

The covering maps $\mathbb{R} \to S^1$ and $S^2 \to \mathbb{R}P(2)$ are special cases of the following important result.

Exercise 4.7. Let X be a metric space on which a group G acts freely and discretely by isometries. Prove that the quotient map $X \mapsto X/G$ is a covering map.

We will postpone further general results and discussion of covering spaces until Lecture 27(b), and for the moment will focus instead on the techniques involved in the specific case $X = B_2(S^1)$.

One visualization of the covering map π is shown in Figure 4.7(a). Topologically, the helix $\{(\cos t, \sin t, t) \mid t \in \mathbb{R}\}$ is equivalent to the real line, and the projection $(x, y, z) \mapsto (x, y)$ is a local homeomorphism from the helix to the circle.

If we unwind just one of the (topological) circles in $B_2(S^1)$, say a, then we obtain the space X shown in Figure 4.7(b); the circle labeled by a unwinds into a copy of $\mathbb{R}$, just as S^1 did, but now the resulting line has a circle corresponding to b attached to every integer

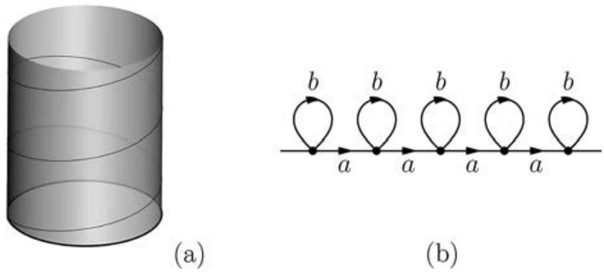

Figure 4.7. Unwinding the circle and the figure-eight.

value. There is a natural projection from X back down to $B_2(S^1)$; however, in the end X is not quite the space we were after. Recall that one of the key tools in our analysis of $\pi_1(S^1)$ was the fact that the homotopy type of a loop in S^1 only depended on the endpoints of its lift to a path in $\mathbb{R}$. In particular, this required every loop in $\mathbb{R}$ to be homotopic to the trivial loop; in other words, it was essential that $\mathbb{R}$ be simply connected. The space X is *not* simply connected, and so we will run into difficulties if we try to study $\pi_1(B_2(S^1))$ using X.

Thus, to obtain the proper covering space for $B_2(S^1)$, we must unwind X still further until we have something simply connected, which will be called the *universal covering space*. The space we get is to be locally homeomorphic to $B_2(S^1)$—that is, every point must either have a neighborhood that is a segment of a path or be a vertex from which four paths emanate. This means that the space we are looking for is a graph in which every vertex has degree 4. Furthermore, in order to be simply connected, it cannot have any loops, and hence must be a tree (recall Definition 4.9).

This is enough to describe the space completely—see Figures 4.8 and 4.9. Let p be the point at which the two circles in $B_2(S^1)$ intersect. We construct the universal covering space, which we call Γ_4, and the covering map $\pi\colon \Gamma_4 \to B_2(S^1)$ by beginning with a single preimage x of the point p, which is the center of the cross in Figure 4.8(a). There are four edges emanating from x, which correspond to the paths a, b, a^{-1}, b^{-1}; at the other end of each of these edges is another preimage of p, distinct from the first one.

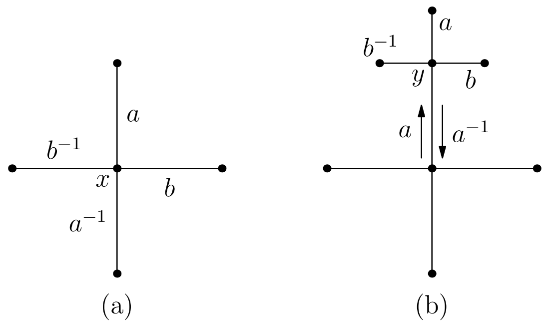

Figure 4.8. Building a tree of homogeneous degree 4.

Consider the point $y \in \Gamma_4$ shown in Figure 4.8(b); this point is the preimage of p lying at the other end of the edge labeled a. The loop a in $B_2(S^1)$ corresponds to following this edge from x to y; following the edge in the reverse direction, from y to x, corresponds to the loop a^{-1}. There must be three other edges emanating from y, and they are labeled a, b, b^{-1}, as shown.

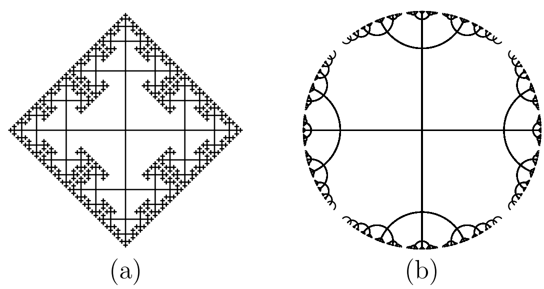

Figure 4.9. Two infinite trees of homogeneous degree 4.

Similarly, each of the other three vertices we constructed at the first step must be the source of three more edges; once these are all drawn, we have five vertices of degree 4 and twelve vertices of degree 1. Each of these twelve vertices must likewise be the source of three further edges, so that it has one edge corresponding to each of the four labels a, b, a^{-1}, b^{-1}; this process continues *ad infinitum*. Thus, Γ_4 is an infinite tree of the sort shown in Figure 4.9(a); observe that at every step, the vertices we add are disjoint from those that came before and from each other, since otherwise we would produce a loop.

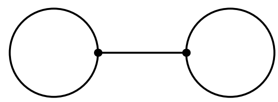

Figure 4.10. Two circles joined by an interval.

Exercise 4.8. Describe a simply connected cover of the space in Figure 4.10 and give it an appropriate name.

Remark. The lengths of the edges of Γ_4 are irrelevant for topological questions, which is what we are interested in at the moment. Nevertheless, the geometric nature of Figure 4.9 is worth noting. Edges further away from x are drawn to be shorter; in particular, if we let $n(z)$ denote the minimum number of edges we must move along to reach x from a vertex z, then the edges emanating from z are drawn with length $2^{-(n(z)+1)}$. If we draw Γ_4 with curved edges rather than straight, we can put Γ_4 in the unit disc, as in Figure 4.9(b), which results in a fractal-like pattern near the edges of the disc that is reminiscent of some of M.C. Escher's artwork. Recalling that these drawings are based on the unit disc model of the hyperbolic plane, we may suspect that there is some connection between Γ_4 and hyperbolic geometry. This is indeed the case, although we shall defer a more detailed discussion until Lecture 29(b).

The projection map $\pi\colon \Gamma_4 \to B_2(S^1)$ is defined in the obvious way: every vertex of Γ_4 is mapped to p, and every edge is mapped to the loop corresponding to its label. In particular, π is a covering map; this is the key to the following result, which says that we can lift paths from $B_2(S^1)$ to Γ_4.

Proposition 4.27. *Let $\gamma\colon [0,1] \to B_2(S^1)$ be a continuous path such that $\gamma(0) = \gamma(1) = p$, and let x be the point at the center of Γ_4. Then there exists a unique continuous path $\tilde{\gamma}\colon [0,1] \to \Gamma_4$ such that $\tilde{\gamma}(0) = x$ and $\pi \circ \tilde{\gamma} = \gamma$.*

Proof. The idea is this: Because π is a covering map, there exists a neighborhood $V \ni p = \gamma(0)$ such that if we write U for the connected component of $\pi^{-1}(V)$ containing x, then $\pi\colon U \to V$ is a homeomorphism. Write $\varphi\colon V \to U$ for the inverse of this homeomorphism; then taking $\epsilon > 0$ such that $\gamma(t) \in V$ for all $t \in [0, \epsilon)$, the unique path $\tilde{\gamma}$ satisfying $\pi \circ \tilde{\gamma} = \gamma$ on $[0, \epsilon)$ is $\tilde{\gamma} = \varphi \circ \gamma$. Repeating this construction and using compactness of the unit interval, one obtains the result.

To make this a little more precise, let $r > 0$ be such that every ball of radius r in $B_2(S^1)$ is simply connected. For example, if the two (topological) circles in Figure 4.5 each have diameter 1, then any $r < \frac{1}{2}$ will suffice, as a ball of radius r in $B_2(S^1)$ is not big enough to contain a complete loop. Now since $\gamma\colon [0,1] \to B_2(S^1)$ is continuous and $[0,1]$ is compact, γ is uniformly continuous, so there exists $\epsilon > 0$ such that $d(\gamma(s), \gamma(t)) < r$ whenever $|s - t| < \epsilon$.

Now for every $t \in [0,1]$, we may write $B(\gamma(t), r)$ for the ball of radius r centered at $\gamma(t)$ in $B_2(S^1)$, and we observe that if U is a connected component of $\pi^{-1}(B(\gamma(t), r))$, then $\pi|_U$ is a homeomorphism from U to $B(\gamma(t), r)$. Given ϵ as above, we see that once $\tilde{\gamma}(t)$ is chosen, the connected component is fixed, and so there exists a unique lift of γ to $\tilde{\gamma}$ on $(t - \epsilon, t + \epsilon)$.

Thus, we start with $\tilde{\gamma}(0) = x$, and observe that this determines a unique $\tilde{\gamma}$ on $[0, \epsilon)$. Applying the above argument to $t = \epsilon/2$, we get a unique $\tilde{\gamma}$ on $[0, 3\epsilon/2)$; applying it to $t = \epsilon$, we get $[0, 2\epsilon)$, and so on. Within a finite number of steps, we have determined $\tilde{\gamma}$ uniquely on $[0, 1]$. □

In fact, the above argument lets us lift more than just paths. We can also lift homotopies, as in Exercise 4.5, which gives a direct link between $\pi_1(B_2(S^1))$ and $\pi_1(\Gamma_4)$.

Proposition 4.28 (Principle of covering homotopy)**.** *Given two continuous loops $\gamma_0, \gamma_1\colon [0,1] \to B_2(S^1)$ based at p and a homotopy $H\colon [0,1] \times [0,1] \to B_2(S^1)$ from γ_0 to γ_1, there exists a unique lift of H to a homotopy from $\tilde{\gamma}_0$ to $\tilde{\gamma}_1$, the lifts guaranteed by Proposition* 4.27*. Furthermore, the lifted homotopy is a homotopy relative to endpoints.*

Proof. Apply Proposition 4.27 to $H(s, \cdot)$ for each $0 \leq s \leq 1$. We can (indeed, must) hold the endpoints fixed because the set of preimages of the base point is discrete. □

b. Standard paths and elements of the free group. In order to describe the homotopy classes of loops in $B_2(S^1)$, we need to give a list of standard representatives, along with a complete homotopy invariant that identifies which element from the list corresponds to a given loop. For the circle S^1, the homotopy invariant was the degree of a loop, which tracked how many times the loop went around the circle; upon being lifted to $\mathbb{R}$, this became the total displacement of the lifted path.

For the torus $\mathbb{T}^2$, the homotopy invariant was a pair of integers specifying the degrees of the projections of the lifted path; this integer pair corresponded to the second endpoint of the lifted path on the integer lattice, which was the lift of the base point.

For the figure eight $B_2(S^1)$, we may likewise expect that the homotopy invariant will be the second endpoint of the lifted path, which lies on the preimage of the base point under the projection map π. This preimage is the set of vertices of Γ_4, and every such vertex may be specified by the sequence of edges we follow to reach it from the "center" of Γ_4. To see this, we first consider a finite sequence of symbols from the set $\{a, a^{-1}, b, b^{-1}\}$—such a sequence is called a *word.* If a word w has the property that the symbols a and a^{-1} never appear next to each other, and similarly for b and b^{-1}, then w is called a *reduced word.* Any word can be transformed into a reduced word by repeatedly canceling all adjacent pairs of inverses. For brevity of notation, we abbreviate aa as a^2, aaa as a^3, and so on; for example, $aaab^{-1}aba^{-1}a^{-1}bbb$ may be written $a^3b^{-1}aba^{-2}b^3$.

Now labeling the edges of Γ_4 with the symbols a, a^{-1}, b, b^{-1} (see Figure 4.8), we associate to each reduced word w the following path in Γ_4. Beginning at the center x, follow the edge corresponding to the first symbol in w; once the second vertex of this edge is reached, follow the edge corresponding to the second symbol in w, and so on. Observe that because w never contains a symbol followed by its inverse, we will never backtrack. Parametrizing this path with uniform speed, one associates to each reduced word a standard path in Γ_4.

This exhibits a one-to-one correspondence between reduced words and standard paths; there is also a one-to-one correspondence between standard paths and vertices in Γ_4. By Proposition 4.28, any two homotopic loops in $B_2(S^1)$ lift to paths in Γ_4 that are homotopic relative to endpoints. In particular, they correspond to the same reduced word.

Let F_2 denote the set of all reduced words using the symbols a, b, a^{-1}, b^{-1}. We have now shown that the process of lifting loops in $B_2(S^1)$ to paths in Γ_4 gives a map $\psi\colon \pi_1(B_2(S^1)) \to F_2$. The previous paragraph shows that ψ is well-defined, and it is obvious that ψ is surjective.

Furthermore, ψ is one-to-one. To see this, we must show that any two loops in $B_2(S^1)$ that lift to paths with the same endpoint in Γ_4 are actually homotopic.

Lemma 4.29. *Every loop based at p in $B_2(S^1)$ is homotopic to one of the standard loops described above.*

Proof. As in the proof of Proposition 4.27, let $r > 0$ be such that every ball of radius r in $B_2(S^1)$ is contractible. Given a loop γ based at p, let $\epsilon > 0$ be such that $\gamma((t-\epsilon, t+\epsilon))$ is contained in such a ball for every $\epsilon > 0$. (This uses uniform continuity of γ.)

Now consider the set $E = \{t \in [0,1] \mid \gamma(t) = p\}$, which contains all parameter values t at which γ returns to the base point. Because γ is continuous, $E = \gamma^{-1}(p)$ is closed, and so $[0,1] \setminus E$ is open—in particular, this complement is a countable union of open intervals. Denote these intervals by (s_n, t_n), and observe that if $|t_n - s_n| < \epsilon$, then $\gamma|_{[s_n,t_n]}$ is homotopic to the constant map $t \mapsto p$.

This shows that γ is homotopic to a loop γ_1 with the property that $[0,1] \setminus \gamma_1^{-1}(p)$ is a *finite* union of open intervals (since there are at most $1/\epsilon$ values of n such that $t_n - s_n \geq \epsilon$). Again, denote these by (s_n, t_n), and observe that each $\gamma|_{[s_n,t_n]}$ is a loop on a circle (which corresponds to either a or b), and hence is homotopic to one of the standard representatives from $\pi_1(S^1)$.

We have shown that γ is homotopic to a concatenation of standard loops on circles; a straightforward reparametrization shows that such a concatenation is homotopic to one of the standard loops described above. □

Lemma 4.29 shows that ψ is a bijection between $\pi_1(B_2(S^1))$ and F_2. In order to complete our description of the fundamental group, it remains to put a group structure on F_2 and show that ψ is in fact an isomorphism.

As with paths, words can be multiplied by concatenation; in order to obtain a reduced word, we must then cancel adjacent inverse symbols. Thus, for example,

$$(aba^2b) \star (b^{-1}a^{-1}b) = aba^2bb^{-1}a^{-1}b = aba^2a^{-1}b = abab.$$

This gives a group structure on F_2, which we call the *free group* with two generators. The same operation with reduced words in the symbols $\{a_1, a_1^{-1}, \dots, a_n, a_n^{-1}\}$ gives F_n, the free group on n generators.

We will study free groups in more detail in the next chapter. For now, we observe that ψ is a homomorphism, since the operation in both groups is concatenation; upon observing that everything we have done generalizes immediately to $B_n(S^1)$ for $n > 2$, we have the following result.

Theorem 4.30. *The fundamental group of the bouquet of n circles is isomorphic to the free group with n generators: $\pi_1(B_n(S^1)) = F_n$.*

The free groups are in some sense the most strongly non-abelian groups possible, in that they have the fewest relations between their generators—none. Our study of these groups in the next chapter will demonstrate the utility of geometric methods in group theory.

Exercise 4.9. Let $H \subset F_n$ be the set of all words such that each generator appears with powers that sum to zero. Show that:

(1) H is a normal subgroup of F_n;

(2) the factor group F_n/H is isomorphic to $\mathbb{Z}^n$;

(3) $[F_n, F_n] \subset H$;

(4) H contains a subgroup isomorphic to F_2.

Exercise 4.10. Describe a free and discrete action of F_n on Γ_{2n} with the property that the quotient space Γ_{2n}/F_n is homeomorphic to $B_n(S^1)$.

Recalling Corollaries 4.18 and 4.20 together with Theorem 4.30, we see that

$$\pi_1(S^1) = \pi_1(\mathbb{R}/\mathbb{Z}) = \mathbb{Z},$$
$$\pi_1(\mathbb{T}^n) = \pi_1(\mathbb{R}^n/\mathbb{Z}^n) = \mathbb{Z}^n,$$
$$\pi_1(B_n(S^1)) = \pi_1(\Gamma_{2n}/F_n) = F_n.$$

One begins to suspect that there may be something going on here. Indeed, in Theorem 5.22 we will see that these two examples are manifestations of a more general phenomenon, wherein the fundamental group of a quotient space is the group doing the acting, as long as the action is "nice enough".