Chapter 3

Characterizing spaces

In this chapter, some basic concepts of topology are introduced, focusing on aspects that are of immediate use in physics and optics. A few of the more formal aspects of topology are discussed briefly in appendix A; for more detailed discussions and for proofs, see topology textbooks such as [1–7].

3.1 Loops, holes, and winding numbers

Loosely speaking, two spaces are topologically equivalent to each other if they can be continuously deformed into each other. One of the most obvious ways to show that two spaces are *not* topologically equivalent is to show that they have different numbers of holes in them. For example, the single-holed torus and the double-holed torus are inequivalent: there is no way to continuously deform the former into the latter: to go from the single to the double torus it is necessary to tear the space to create the second hole, and tearing is a discontinuous transformation.

So how do you characterize the number of holes? A simple way is to look at the sets of closed loops that can be continuously deformed into each other. Consider a plane with a single hole punctured in it (left panel of figure 3.1). The loops marked A and B can be continuously deformed into each other. In fact, they can both be continuously collapsed down to a single point. So we consider these loops to be equivalent to each other. However, loop C circles the hole. It cannot be continuously deformed into either A or B (or to a single point), because it gets snagged on the hole. Similarly, a loop that circles the hole twice cannot be deformed into A, B, or C. We therefore have an infinite set of equivalence classes of loops: the nth class consists of all the loops that circle the hole n times. The loops on this space are therefore characterized by a single integer, called the *winding number*, which will be defined in more detail in chapter 5. Note that the loop has an orientation given by the direction it rotates (clockwise or counter-clockwise). The winding number has a sign determined by this orientation: we take $n > 0$ if the loop circulates around the hole counter-clockwise, and $n < 0$ for clockwise.

doi:10.1088/2053-2571/aaddd5ch3

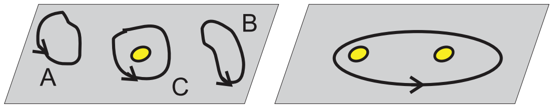

Figure 3.1. On the left, the plane with a single puncture in it contains loops that cannot be continuously deformed into each other without getting caught on the hole. So each loop can be characterized by an integer counting the number of times n the loop is circled. For the loops shown, A and B are equivalent to each other, with $n = 0$. C is not equivalent to the other two loops, since it has $n = 1$. For the plane on the right, with two punctures, loops are characterized by a *pair* of integers (n_1, n_2) counting the number of times each of the two holes is enclosed; the loop shown has $n_1 = n_2 = 1$.

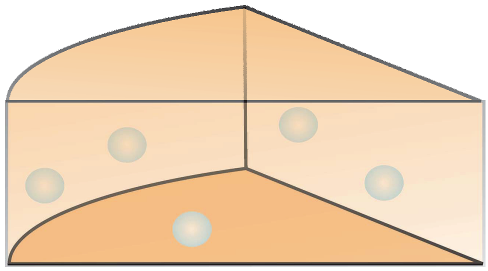

Figure 3.2. The air bubbles in a piece of Swiss cheese are a different type of hole than the puncture in the plane. Whereas a loop (which is deformable to a circle) can slide around any of the bubbles and be deformed into any other loop, a two-dimensional sphere that surrounds a bubble cannot be deformed into a sphere that does not. So the space is characterized by equivalence classes of two-dimensional spheres, rather than one-dimensional loops.

But now consider a second plane, with two holes punctured in it (right panel of figure 3.1). Here, characterizing equivalence classes of loops now requires specifying two integers, (n_1, n_2), where n_1 specifies how many times the loop circles the left-hand hole and n_2 counts the number of windings around the right-hand hole. A plane with three holes would require specifying three integer winding numbers, and so on.

This is still not the end of the story, however. A space may have holes of different types. Holes in the plane are not the same as holes in a piece of Swiss cheese (figure 3.2). A loop like those of figure 3.1 can slip around the holes in the cheese. So loops are incapable of detecting these ‘higher-dimensional’ holes. However, instead of loops (which are topologically equivalent to a circle or one-dimensional sphere, S^1) we can use two-dimensional spheres, S^2, and look at equivalence classes that can be continuously deformed into each other without crossing holes. So equivalence classes of spheres that wind around all of the holes the same number of times can be used to characterize the space. These integer spherical winding numbers will be called Chern numbers (chapter 5).

Continuing in this way, we can characterize the hole structure of a space by a series of integers representing equivalence classes of spheres of different dimensions. A method for testing whether or not two spaces are topologically equivalent is then apparent: compare the list of such integers for the two spaces. If the lists are not equal, then the spaces are distinct and cannot be smoothly deformed into each other.

The idea of treating spaces with the same hole structure as equivalent is formalized by the idea of *homotopy classes*, which will be defined in the next section. Winding numbers and other topological invariants will be discussed in more detail in chapter 5.

3.2 Homotopy classes

Homotopy classes are a means of classifying the structure of a topological space by equivalence classes of circles (one-dimensional loops), or more generally, of spheres of different dimensions. The idea of classifying surfaces by means of loops apparently goes back to Jordan in the 1860s. The idea of imposing a group structure on the set of homotopy classes originated with the work of Poincaré in the 1890s.

Let X be a topological space (see the appendix), and let I denote the unit interval, $I = [0, 1]$. A *path* α in X with endpoints x_0 and x_1 is a continuous map α: $I \rightarrow X$ (in other words the map has image $\alpha(t)$ in X, for $0 \leqslant t \leqslant 1$), such that $\alpha(0) = x_0$ and $\alpha(1) = x_1$ (figure 3.3(a)). A *loop* in X is a path whose ends are identified, $x_0 = x_1$ (figure 3.3(b)). The loop is said to be *based* at x_0.

Given two loops α and β based at the same point, one can define a product path $\alpha * \beta$ as the path that follows one loop until it returns to the base point, then follows the other loop (figure 3.3(c)):

$$\alpha * \beta(t) = \begin{cases} \alpha(2t), & \text{for } 0 \leqslant t \leqslant \frac{1}{2} \\ \beta(2t - 1), & \text{for } \frac{1}{2} < t \leqslant 1. \end{cases} \tag{3.1}$$

The constant loop is simply the loop that stays fixed at x_0 for all t: $\alpha(t) = x_0$ for $0 \leqslant t \leqslant 1$, and the inverse of the loop is obtained by running the parameter t in the opposite direction: $\alpha^{-1}(t) = \alpha(1 - t)$. With these definitions, it can easily be shown that the set of loops based at a given point form a mathematical group, with the constant loop playing the role of the group identity element.

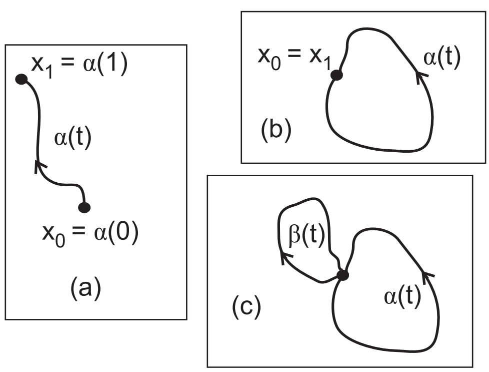

Figure 3.3. (a) A path going from x_0 to x_1. (b) A closed loop obtained by identifying the two endpoints of a path. (c) The product of two loops.

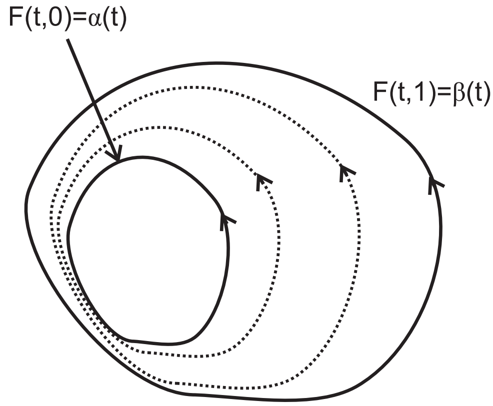

Figure 3.4. Homotopy of two loops $\alpha(t)$ and $\beta(t)$. The innermost loop is $F(t, 0) = \alpha(t)$, the outermost loop is $F(t, 1) = \beta(t)$. The dotted loops are representative examples of $F(t, s)$ for two other values of s ($0 < s < 1$). As s increases, $\alpha(t)$ gradually evolves into $\beta(t)$.

But a more interesting and useful group may be obtained by adding a second parameter. Given two loops $\alpha(t)$ and $\beta(t)$ in X, based at the same point, we define a two-parameter continuous map $F: I \times I \to X$, which provides a continuous deformation of α into β:

$$F(t, 0) = \alpha(t), \qquad F(t, 1) = \beta(t), \qquad F(0, s) = F(1, s) = x_0 \tag{3.2}$$

(see figure 3.4). The first parameter t carries us along the loop, while varying the second parameter continuously deforms one loop into the other. Such a deformation is called a *homotopy*. The idea of homotopy can be generalized in an obvious manner from loops to arbitrary continuous maps.

If two loops α and β are homotopic to each other, we write $\alpha \sim \beta$. Homotopy is an equivalence relation, so we define the *homotopy classes* $[\alpha]$ of loop α to be the equivalence class of loops homotopic to α. In other words $[\alpha]$ is the set of loops at x_0 continuously deformable into α.

Given two topological spaces, X and Y, they are said to be of the same *homotopy type* if there exist continuous maps $f: X \to Y$ and $g: Y \to X$ such that

$$f \circ g \sim \mathcal{I}_Y \quad \text{and} \quad g \circ f \sim \mathcal{I}_X, \tag{3.3}$$

where $\sim$ denotes equivalence under homotopy and $\mathcal{I}_{X,\,Y}$ denote the identity maps on spaces X and Y.

A topological space X is called *arc-wise connected* if, given any two points $x_0, x_1 \in X$, there is a path such that x_0 and x_1 are its endpoints. On an arc-wise connected space, the set of homotopy classes at each base point is isomorphic to the homotopy classes at any other base point. In this case, all base points are equivalent, and so there is no need to specify which base point is used. Henceforth, we only consider arc-wise connected spaces and will often omit mention of the base point.

Given the product of loops defined above, the set of homotopy classes inherits a natural product, which will also be denoted by $*$:

$$[\alpha] * [\beta] = [\alpha * \beta]. \tag{3.4}$$

Inverses, associativity and the existence of an identity element then follow:

$$[\alpha]^{-1} = [\alpha^{-1}], \tag{3.5}$$

$$([\alpha] * [\beta]) * [\gamma] = [\alpha] * ([\beta] * [\gamma]), \tag{3.6}$$

$$[\alpha] * [c] = [c] * [\alpha] = \alpha, \tag{3.7}$$

where the equivalence class $[c]$ of the constant loop c serves as the identity element. The homotopy classes therefore form a group, called the *fundamental group* or the *first homotopy group* of X, denoted by $\pi_1(X, x_0)$ or simply as $\pi_1(X)$.

Homotopy type is an equivalence relation among topological spaces and it can be shown that two spaces with the same fundamental group are of the same homotopy type. Thus, the fundamental group can be used to classify homotopy-equivalent sets of spaces.

Consider some simple examples:

(i) *Euclidean spaces*, $\mathbb{R}^n$. All loops are deformable to each other, so there is a single homotopy class. The homotopy group has a single entry, and since every group must contain the identity element $\mathcal{I}$, the group in this case consists of *just* the identity: $\pi_1(\mathbb{R}^n) = \{\mathcal{I}\}$.

(ii) The *n-dimensional sphere*: S^n. For $n > 1$, all loops on the sphere can be smoothly deformed into each other simply by sliding them around on the sphere's surface, so that once again $\pi_1(S^n) = \{\mathcal{I}\}$ for $n > 1$. However, for $n = 1$, the one-dimensional sphere is a circle; closed loops on the circle are distinguished from each other by a single integer, the number of times they wind around the circle. So the homotopy group is simply the group of integers: $\pi_1(S^1) = \mathbb{Z}$.

(iii) *Tori*: An n-dimensional torus T^n is formed from the product of n circles, so there are n integers counting the windings about each hole. Therefore, $\pi_1(T^n) = \mathbb{Z}^n$.

Evaluating the fundamental group allows us to detect holes such as those in a punctured plane or the hole in a donut. As a more physical example, the interior of the solenoid in the Aharonov–Bohm (AB) effect serves as a hole in the charged particle's configuration space, so that the AB effect can be viewed as a consequence of the nontrivial first homotopy group. But as mentioned earlier, there are other types of holes that cannot be detected by looking at deformations of circles. To study these holes, we must move from circles (one-dimensional spheres) to spheres of higher dimension; this leads us to define the higher homotopy groups.

Let the symbol ∂ denote the boundary operator; ∂M is the set of points on the boundary of M. Consider a unit cube, the set of points

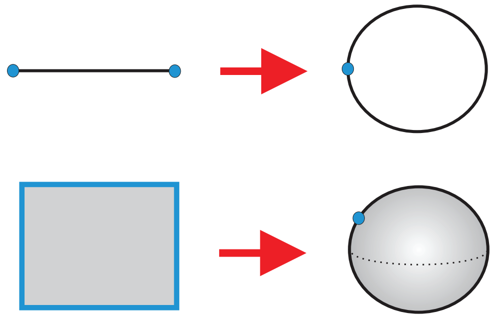

Figure 3.5. Forming n-spheres by identifying the boundaries of n-cubes. For $n = 1$, the endpoints of the unit interval or one-cube (the blue dots) are identified with each other to form a circle. For $n = 2$, the boundary of the square is collapsed to a point to form a two-sphere.

$$I_n = \{s_1, s_2, \ldots, s_n\}, \text{ with } 0 \leqslant s_i \leqslant 1. \tag{3.8}$$

The boundary of I_n is the surface of the cube. In one dimension, the unit interval $I = I_1$ has boundary given by the pair of endpoints, $\partial I_1 = \{0, 1\}$; this interval is converted into a circle by identifying the endpoints with each other, or in other words, gluing the ends of the interval together (figure 3.5). Similarly, for $n > 1$ we can collapse the boundary of I_n to a single point (identify all points on the surface with each other), to convert the n-cube into something isomorphic to an n-sphere, S^n. This identification is formally written as a quotient (more precisely as a *coset space*):

$$S^n = I_n/\partial I_n. \tag{3.9}$$

Now that ∂I_n is collapsed to a point, x_0 we can use it as a base point for n-loops. The *n-loop* α is a continuous map of the cube to topological space X, leaving the boundary fixed:

$$\alpha \colon I_n \to X, \qquad \text{such that } \alpha \colon \partial I_n \to x_0. \tag{3.10}$$

Two such loops are then *homotopic* $\alpha \sim \beta$ if there exists a *homotopy* $F\colon I_n \times I \to X$ such that:

$$F(s_1, \ldots, s_n, 0) = \alpha(s_1, \ldots, s_n) \tag{3.11}$$

$$F(s_1, \ldots, s_n, 1) = \beta(s_1, \ldots, s_n) \tag{3.12}$$

$$F(s_1, \ldots, s_n, t) = x_0 \text{ for } (s_1, \ldots, s_n) \in \partial I_n. \tag{3.13}$$

the homotopy relation $\alpha \sim \beta$ is again an equivalence relation, so that we may define the corresponding homotopy equivalence classes $[\alpha]$. The product of n-loops, $\alpha * \beta$ is

$$\alpha * \beta(s_1, \ldots, s_n) = \begin{cases} \alpha(2s_1, s_2, \ldots, s_n), & \text{for } 0 \leqslant s_1 \leqslant \frac{1}{2} \\ \beta(2s_1 - 1, s_2, \ldots, s_n), & \text{for } \frac{1}{2} < s_1 \leqslant 1. \end{cases} \tag{3.14}$$

The *nth homotopy group* at x_0 for $n \geqslant 2$ is then defined in direct analogy to the fundamental group: $\pi_n(X, x_0)$ is the group of equivalence classes continuous maps $S^n \to X$, where we are identifying $I_n/\partial I_n$ with the n-sphere. π_n quantifies the set of topologically inequivalent n-spheres in the topological space that cannot be deformed into each other without being obstructed by holes.

We have now defined homotopy groups $\pi_n(X)$ for $n \geqslant 1$. We can carry out the analogous construction in zero dimensions as well to form a zeroth homotopy 'group': the zero-dimensional interval is simply a point, $I_0 = x_0$, with the boundary being the empty set: $\partial I_0 = \{\varnothing\}$. The zero sphere $S^0 \sim I_0/\partial I_0$ is then the point x_0, and all loops are just constant maps. Two such loops at points x and y in X will be homotopic, $\alpha \sim \beta$, if and only if x and y can be smoothly deformed into each other, i.e. if they are the endpoints of some continuous curve. A space may be composed of multiple components that are disconnected from each other, such as several concentric spheres nested inside each other. The set of equivalence classes, $\pi_0(X)$ is then just the set of connected components. A space is simply connected if $\pi_0(X)$ is the trivial group containing a single element, while a multiply connected space has $\pi_1(X)$ isomorphic to a finite set of integers that label the connected components. Note however, that $\pi_0(X)$ is *not* a group, unlike the $\pi_n(X)$ with $n > 0$.

A stronger notion of topological equivalence than homotopy type is homeomorphism, which means essentially that two spaces can be continuously deformed into each other. (See the appendix for the precise definition.) As mentioned above, spaces with the same fundamental group are of the same homotopy type. However, since there are types of holes that cannot be distinguished by the fundamental group, being of the same homotopy type is a weaker condition than being homeomorphic. Dimension is largely invisible to the fundamental group. For example, points and circles are homotopy equivalent to solid balls and Möbius bands, respectively, due to the ability to continuously contract one to the other; however these spaces are not homeomorphic (note that the contraction is not uniquely invertible). Including the higher homotopy classes is one way to help distinguish between spaces that are of the same homotopy type but which are not homeomorphic.

References

[1] Fulton W 1995 *Algebraic Topology: A First Course* (Berlin: Springer)

[2] Hatcher A 2002 *Algebraic Topology* (Cambridge: Cambridge University Press)

[3] Greenberg M and Harper J 2018 *Algebraic Topology: A First Course* (Boca Raton, FL: CRC Press)

[4] Munkres J R 2000 *Topology* (Upper Saddle River, NJ: Prentice-Hall)

[5] Hirsch M W 1997 *Differential Topology* (Berlin: Springer)
[6] Guillemin V and Pollock A 1974 *Differential Topology* (Englewood Cliffs, NJ: Prentice-Hall)
[7] Basener W F 2006 *Topology and Its Applications* (Hoboken, NJ: Wiley)