2

MATHEMATICAL PRELIMINARIES

In the present chapter we introduce elementary concepts in the theory of maps, vector spaces and topology. A modest knowledge of undergraduate mathematics, such as set theory, calculus, complex analysis and linear algebra is assumed.

The main purpose of this book is to study the application of the theory of manifolds to the problems in physics. Vector spaces and topology are, in a sense, two extreme viewpoints of manifolds. A manifold is a space which locally looks like $\mathbb{R}^n$ (or $\mathbb{C}^n$) but not necessarily globally. As a first approximation, we may model a small part of a manifold by a Euclidean space $\mathbb{R}^n$ (or $\mathbb{C}^n$) (a small area around a point on a surface can be approximated by the tangent plane at that point); this is the viewpoint of a vector space. In topology, however, we study the manifold as a whole. We want to study the properties of manifolds and classify manifolds using some sort of 'measures'. Topology usually comes with an adjective: algebraic topology, differential topology, combinatorial topology, general topology and so on. These adjectives refer to the measure we use when classifying manifolds.

2.1 Maps

2.1.1 Definitions

Let X and Y be sets. A **map** (or **mapping**) f is a rule by which we assign $y \in Y$ for each $x \in X$. We write

$$f : X \to Y. \tag{2.1}$$

If f is defined by some explicit formula, we may write

$$f : x \mapsto f(x) \tag{2.2}$$

There may be more than two elements in X that correspond to the same $y \in Y$. A subset of X whose elements are mapped to $y \in Y$ under f is called the **inverse image** of y, denoted by $f^{-1}(y) = \{x \in X | f(x) = y\}$. The set X is called the **domain** of the map while Y is called the **range** of the map. The **image** of the map is $f(X) = \{y \in Y | y = f(x) \text{ for some } x \in X\} \subset Y$. The image $f(X)$ is also denoted by $\operatorname{im} f$. The reader should note that a map cannot be defined without specifying the domain and the range. Take $f(x) = \exp x$, for example. If both the domain and the range are $\mathbb{R}$, $f(x) = -1$ has no inverse image. If. however, the domain and the range are the complex plane $\mathbb{C}$, we find $f^{-1}(-1) = \{(2n+1)\pi\mathrm{i} | n \in Z\}$. The domain X and the range Y are as important as f itself in specifying a map.

Example 2.1. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = \sin x$. We also write $f : x \mapsto \sin x$. The domain and the range are $\mathbb{R}$ and the image $f(\mathbb{R})$ is $[-1, 1]$. The inverse image of 0 is $f^{-1}(0) = \{n\pi | n \in \mathbb{Z}\}$. Let us take the same function $f(x) = \sin x = (\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x})/2\mathrm{i}$ but $f : \mathbb{C} \to \mathbb{C}$ this time. The image $f(C)$ is the whole complex plane $\mathbb{C}$.

Definition 2.1. If a map satisfies a certain condition it bears a special name.

(a) A map $f : X \to Y$ is called **injective** (or **one to one**) if $x \neq x'$ implies $f(x) \neq f(x')$.
(b) A map $f : X \to Y$ is called **surjective** (or **onto**) if for each $y \in Y$ there exists at least one element $x \in X$ such that $f(x) = y$.
(c) A map $f : X \to Y$ is called **bijective** if it is both injective and surjective.

Example 2.2. A map $f : \mathbb{R} \to \mathbb{R}$ defined by $f : x \mapsto ax$ $(a \in \mathbb{R} - \{0\})$ is bijective. $f : \mathbb{R} \to \mathbb{R}$ defined by $f : x \mapsto x^2$ is neither injective nor surjective. $f : \mathbb{R} \to \mathbb{R}$ given by $f : x \mapsto \exp x$ is injective but not surjective.

Exercise 2.1. A map $f : \mathbb{R} \to \mathbb{R}$ defined by $f : x \mapsto \sin x$ is neither injective nor surjective. Restrict the domain and the range to make f bijective.

Example 2.3. Let M be an element of the general linear group $\mathrm{GL}(n, \mathbb{R})$ whose matrix representation is given by $n \times n$ matrices with non-vanishing determinant. Then $M : \mathbb{R}^n \to \mathbb{R}^n$, $x \mapsto Mx$ is bijective. If $\det M = 0$, it is neither injective nor surjective.

A **constant map** $c : X \to Y$ is defined by $c(x) = y_0$ where y_0 is a fixed element in Y and x is an arbitrary element in X. Given a map $f : X \to Y$, we may think of its **restriction** to $A \subset X$, which is denoted as $f|_A : A \to Y$. Given two maps $f : X \to Y$ and $g : Y \to Z$, the **composite map** of f and g is a map $g \circ f : X \to Z$ defined by $g \circ f(x) = g(f(x))$. A diagram of maps is called **commutative** if any composite maps between a pair of sets do not depend on how they are composed. For example, in figure 2.1, $f \circ g = h \circ j$ and $f \circ g = k$ etc.

Exercise 2.2. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f : x \to x^2$ and $g : \mathbb{R} \to \mathbb{R}$ by $g : x \to \exp x$. What are $g \circ f : \mathbb{R} \to \mathbb{R}$ and $f \circ g : \mathbb{R} \to \mathbb{R}$?

If $A \subset X$, an **inclusion map** $i : A \to X$ is defined by $i(a) = a$ for any $a \in A$. An inclusion map is often written as $i : A \hookrightarrow X$. The **identity map** $\mathrm{id}_X : X \to X$ is a special case of an inclusion map, for which $A = X$. If $f : X \to Y$ defined by $f : x \mapsto f(x)$ is bijective, there exists an **inverse map** $f^{-1} : Y \to X$, such that $f^{-1} : f(x) \to x$, which is also bijective. The maps f

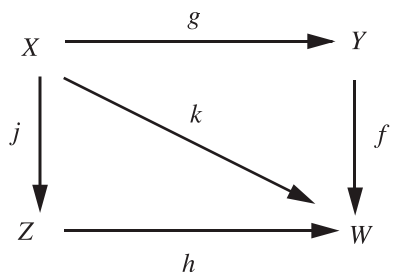

Figure 2.1. A commutative diagram of maps.

and f^{-1} satisfy $f \circ f^{-1} = \mathrm{id}_Y$ and $f^{-1} \circ f = \mathrm{id}_X$. Conversely, if $f : X \to Y$ and $g : Y \to X$ satisfy $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$, then f and g are bijections. This can be proved from the following exercise.

Exercise 2.3. Show that if $f : X \to Y$ and $g : Y \to X$ satisfy $g \circ f = \mathrm{id}_X$, f is injective and g is surjective. If this is applied to $f \circ g = \mathrm{id}_Y$ as well, we obtain the previous result.

Example 2.4. Let $f : \mathbb{R} \to (0, \infty)$ be a bijection defined by $f : x \mapsto \exp x$. Then the inverse map $f^{-1} : (0, \infty) \to \mathbb{R}$ is $f^{-1} : x \mapsto \ln x$. Let $g : (-\pi/2, \pi/2) \to (-1, 1)$ be a bijection defined by $g : x \to \sin x$. The inverse map is $g^{-1} : x \mapsto \sin^{-1} x$.

Exercise 2.4. The n-dimensional Euclidean group E^n is made of an n-dimensional translation $a : x \to x + a$ $(x, a \in \mathbb{R}^n)$ and an $\mathrm{O}(n)$ rotation $R : x \to Rx$, $R \in \mathrm{O}(n)$. A general element (R, a) of E^n acts on x by $(R, a) : x \mapsto Rx + a$. The product is defined by $(R_2, a_2) \times (R_1, a_1) : x \mapsto R_2(R_1 x + a_1) + a_2$, that is, $(R_2, a_2) \circ (R_1, a_1) = (R_2 R_1, R_2 a_1 + a_2)$. Show that the maps a, R and (R, a) are bijections. Find their inverse maps.

Suppose certain algebraic structures (product or addition, say) are endowed with the sets X and Y. If $f : X \to Y$ preserves these algebraic structures, then f is called a **homomorphism**. For example, let X be endowed with a product. If f is a homomorphism, it preserves the product, $f(ab) = f(a)f(b)$. Note that ab is defined by the product rule in X, and $f(a)f(b)$ by that in Y. If a homomorphism f is bijective, f is called an **isomorphism** and X is said to be **isomorphic** to Y, denoted $x \cong y$.

2.1.2 Equivalence relation and equivalence class

Some of the most important concepts in mathematics are **equivalence relations** and **equivalence classes**. Although these subjects are not directly related to maps, it is appropriate to define them at this point before we proceed further. A **relation** R defined in a set X is a subset of X^2. If a point $(a, b) \in X^2$ is in R, we may write aRb. For example, the relation $>$ is a subset of $\mathbb{R}^2$. If $(a, b) \in >$, then $a > b$.

Definition 2.2. An **equivalence relation** $\sim$ is a relation which satisfies the following requirements:

(i) $a \sim a$ (reflective).
(ii) If $a \sim b$, then $b \sim a$ (symmetric).
(iii) If $a \sim b$ and $b \sim c$, then $a \sim c$ (transitive).

Exercise 2.5. If an integer is divided by 2, the remainder is either 0 or 1. If two integers n and m yield the same remainder, we write $m \sim n$. Show that $\sim$ is an equivalence relation in $\mathbb{Z}$.

Given a set X and an equivalence relation $\sim$, we have a partition of X into *mutually disjoint* subsets called **equivalence classes**. A class $[a]$ is made of all the elements x in X such that $x \sim a$,

$$[a] = \{x \in X | x \sim a\} \tag{2.3}$$

$[a]$ cannot be empty since $a \sim a$. We now prove that if $[a] \cap [b] \neq \emptyset$ then $[a] = [b]$. First note that $a \sim b$. (Since $[a] \cap [b] \neq \emptyset$ there is at least one element in $[a] \cap [b]$ that satisfies $c \sim a$ and $c \sim b$. From the transitivity, we have $a \sim b$.) Next we show that $[a] \subset [b]$. Take an arbitrary element a' in $[a]$; $a' \sim a$. Then $a \sim b$ implies $b \sim a'$, that is $a' \in [b]$. Thus, we have $[a] \subset [b]$. Similarly, $[a] \supset [b]$ can be shown and it follows that $[a] = [b]$. Hence, two classes $[a]$ and $[b]$ satisfy either $[a] = [b]$ or $[a] \cap [b] = \emptyset$. In this way a set X is decomposed into mutually disjoint equivalence classes. The set of all classes is called the **quotient space**, denoted by $X/\sim$. The element a (or any element in $[a]$) is called the **representative** of a class $[a]$. In exercise 2.5, the equivalence relation $\sim$ divides integers into two classes, even integers and odd integers. We may choose the representative of the even class to be 0, and that of the odd class to be 1. We write this quotient space $\mathbb{Z}/\sim$. $\mathbb{Z}/\sim$ is isomorphic to $\mathbb{Z}_2$, the **cyclic group** of order 2, whose algebra is defined by $0 + 0 = 0$, $0 + 1 = 1 + 0 = 1$ and $1 + 1 = 0$. If all integers are divided into equivalence classes according to the remainder of division by n, the quotient space is isomorphic to $\mathbb{Z}_n$, the cyclic group of order n.

Let X be a space in our usual sense. (To be more precise, we need the notion of topological space, which will be defined in section 2.3. For the time being we depend on our intuitive notion of 'space'.) Then quotient spaces may be realized as geometrical figures. For example, let x and y be two points in $\mathbb{R}$.

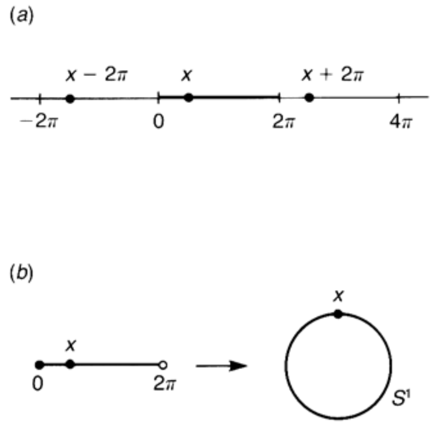

Figure 2.2. In (a) all the points $x + 2n\pi$, $n \in \mathbb{Z}$ are in the same equivalence class $[x]$. We may take $x \in [0, 2\pi)$ as a representative of $[x]$. (b) The quotient space $\mathbb{R}/\sim$ is the circle S^1.

Introduce a relation $\sim$ by: $x \sim y$ if there exists $n \in \mathbb{Z}$ such that $y = x + 2\pi n$. It is easily shown that $\sim$ is an equivalence relation. The class $[x]$ is the set $\{\ldots, x - 2\pi, x, x + 2\pi, \ldots\}$. A number $x \in [0, 2\pi)$ serves as a representative of an equivalence class $[x]$, see figure 2.2(a). Note that 0 and 2π are different points in $\mathbb{R}$ but, according to the equivalence relation, these points are looked upon as the same element in $\mathbb{R}/\sim$. We arrive at the conclusion that the quotient space $\mathbb{R}/\sim$ is the circle $S^1 = \{e^{i\theta} | 0 \leq \theta < 2\pi\}$; see figure 2.2($b$). Note that a point ε is close to a point $2\pi - \varepsilon$ for infinitesimal ε. Certainly this is the case for S^1, where an angle ε is close to an angle $2\pi - \varepsilon$, but not the case for $\mathbb{R}$. The concept of closeness of points is one of the main ingredients of topology.

Example 2.5. (a) Let X be a square disc $\{(x, y) \in \mathbb{R}^2 | \, |x| \geq 1, |y| \geq 1\}$. If we identify the points on a pair of facing edges, $(-1, y) \sim (1, y)$, for example, we obtain the cylinder, see figure 2.3(a). If we identify the points $(-1, -y) \sim (1, y)$, we find the Möbius strip, see figure 2.3(b). [*Remarks*: If readers are not familiar with the Möbius strip, they may take a strip of paper and glue up its ends after a π-twist. Because of the twist, one side of the strip has been joined to the other side, making the surface single sided. The Möbius strip is an example of a **non-orientable** surface, while the cylinder has definite sides and is said to be **orientable**. Orientability will be discussed in terms of differential forms in section 5.5.]

(b) Let (x_1, y_1) and (x_2, y_2) be two points in $\mathbb{R}^2$ and introduce an equivalence relation $\sim$ by: $(x_1, y_1) \sim (x_2, y_2)$ if $x_2 = x_1 + 2\pi n_x$ and $y_2 = y_1 + 2\pi n_y$, $n_x, n_y \in \mathbb{Z}$. Then $\sim$ is an equivalence relation. The quotient space $\mathbb{R}_2/\sim$ is the **torus** T^2 (the surface of a doughnut), see figure 2.4(a). Alternatively, T^2 is

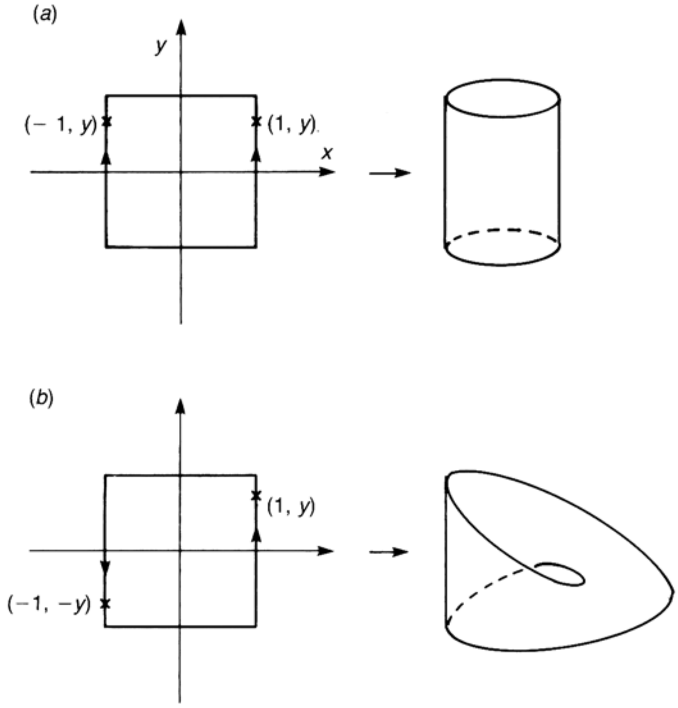

Figure 2.3. (*a*) The edges $|x| = 1$ are identified in the direction of the arrows to form a cylinder. (*b*) If the edges are identified in the opposite direction, we have a Möbius strip.

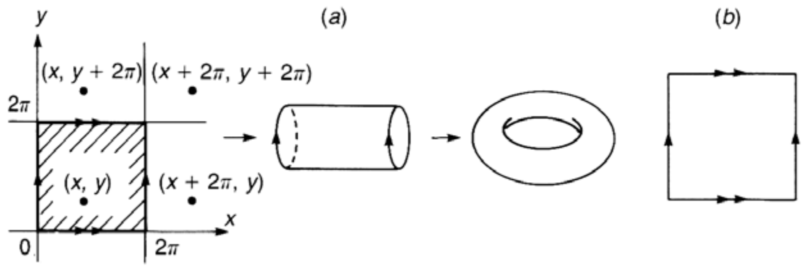

Figure 2.4. If all the points $(x + 2\pi n_x, y + 2\pi n_y), n_x, n_y \in \mathbb{Z}$ are identified as in (*a*), the quotient space is taken to be the shaded area whose edges are identified as in (*b*). This resulting quotient space is the torus T^2.

represented by a rectangle whose edges are identified as in figure 2.4(*b*).

(c) What if we identify the edges of a rectangle in other ways? Figure 2.5 gives possible identifications. The spaces obtained by these identifications are

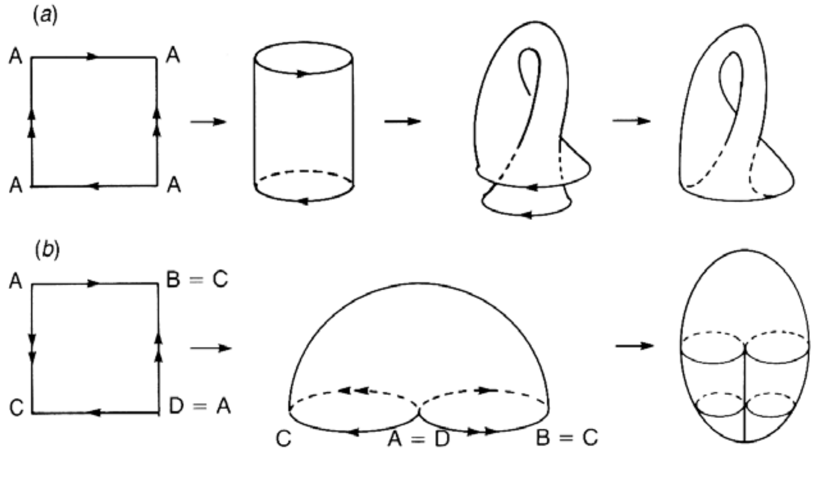

Figure 2.5. The Klein bottle (a) and the projective plane (b).

called the **Klein bottle**, figure 2.5(a), and the **projective plane**, figure 2.5(b), neither of which can be realized (or *embedded*) in the Euclidean space $\mathbb{R}^3$ without intersecting with itself. They are known to be non-orientable.

The projective plane, which we denote RP^2, is visualized as follows. Let us consider a unit vector $\boldsymbol{n}$ and identify $\boldsymbol{n}$ with $-\boldsymbol{n}$, see figure 2.6. This identification takes place when we describe a rod with no head or tail, for example. We are tempted to assign a point on S^2 to specify the 'vector' $\boldsymbol{n}$. This works except for one point. Two antipodal points $\boldsymbol{n} = (\theta, \phi)$ and $-\boldsymbol{n} = (\pi - \theta, \pi + \phi)$ represent the same state. Then we may take a northern hemisphere as the coset space $S^2/\sim$ since only a half of S^2 is required. However, the coset space is not just an ordinary hemisphere since the antipodal points on the equator are identified. By continuous deformation of this hemisphere into a square, we obtain the square in figure 2.5(b).

(d) Let us identify pairs of edges of the octagon shown in figure 2.7(a). The quotient space is the torus with two handles, denoted by Σ_2, see figure 2.7(b). Σ_g, the torus with g handles, can be obtained by a similar identification, see problem 2.1. The integer g is called the **genus** of the torus.

(e) Let $D^2 = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 1\}$ be a closed disc. Identify the points on the boundary $\{(x, y) \in \mathbb{R}^2 | x^2 + y^2 = 1\}$; $(x_1, y_1) \sim (x_2, y_2)$ if $x_1^2 + y_1^2 = x_2^2 + y_2^2 = 1$. Then we obtain the sphere S^2 as the quotient space $D^2/\sim$, also written as D^2/S^1, see figure 2.8. If we take an n-dimensional disc $D^n = \{(x_0, \ldots, x^n) \in \mathbb{R}^{n+1} | (x_0)^2 + \cdots + (x^n)^2 \leq 1\}$ and identify the points on the surface S^{n-1}, we obtain the n-sphere S^n, namely $D^n/S^{n-1} = S^n$.

Exercise 2.6. Let H be the upper-half complex plane $\{\tau \in \mathbb{C} | \operatorname{Im} \tau \geq 0\}$. Define a

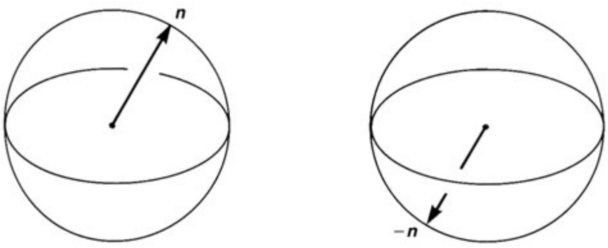

Figure 2.6. If $\boldsymbol{n}$ has no head or tail, one cannot distinguish $\boldsymbol{n}$ from $-\boldsymbol{n}$ and they must be identified. One obtains the projective plane RP^2 by this identification $\boldsymbol{n} \sim -\boldsymbol{n}$; $RP^2 \simeq S^2/\sim$. It suffices to take a hemisphere to describe the coset space. Note, however, that the antipodal points on the equator are identified.

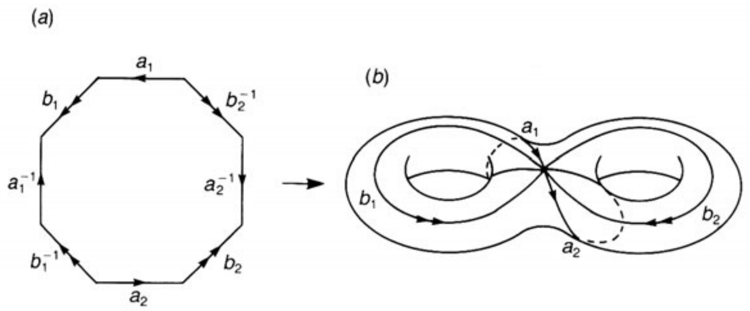

Figure 2.7. If the edges of (a) are identified a torus with two holes (genus two) is obtained.

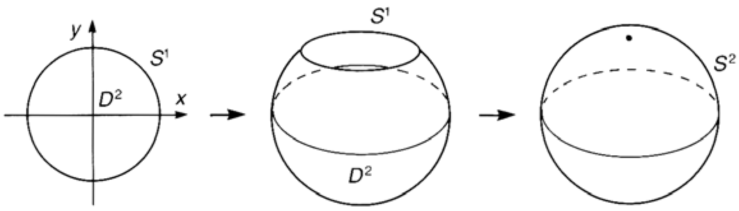

Figure 2.8. A disc D^2 whose boundary S^1 is identified is the sphere S^2.

group

$$\mathrm{SL}(2, \mathbb{Z}) \equiv \left\{ \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) \middle| a, b, c, d \in \mathbb{Z}, ad - bc = 1 \right\}. \tag{2.4}$$

Introduce a relation $\sim$, for $\tau, \tau' \in \mathrm{H}$, by $\tau \sim \tau'$ if there exists a matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$$

such that

$$\tau' = (a\tau + b)/(c\tau + d). \tag{2.5}$$

Show that this is an equivalence relation. (The quotient space $\mathrm{H}/\mathrm{SL}(2, \mathbb{Z})$ is shown in figure 8.3.)

Example 2.6. Let G be a group and H a subgroup of G. Let $g, g' \in G$ and introduce an equivalence relation $\sim$ by $g \sim g'$ if there exists $h \in H$ such that $g' = gh$. We denote the equivalence class $[g] = \{gh|h \in H\}$ by gH. The class gH is called a (**left**) **coset**. gH satisfies either $gH \cap g'H = \emptyset$ or $gH = g'H$. The quotient space is denoted by G/H. In general G/H is not a group unless H is a **normal subgroup** of G, that is, $ghg^{-1} \in H$ for any $g \in G$ and $h \in H$. If H is a normal subgroup of G, G/H is called the **quotient group**, whose group operation is given by $[g] * [g'] = [gg']$, where $*$ is the product in G/H. Take $gh\epsilon[g]$ and $g'h'\epsilon[g']$. Then there exists $h''\epsilon H$ such that $hg' = g'h''$ and hence $ghg'h' = gg'h''h'\epsilon[gg']$. The unit element of G/H is the equivalence class $[e]$ and the inverse element of $[g]$ is $[g^{-1}]$.

Exercise 2.7. Let G be a group. Two elements $a, b \in G$ are said to be conjugate to each other, denoted by $a \simeq b$, if there exists $g \in G$ such that $b = gag^{-1}$. Show that $\simeq$ is an equivalence relation. The equivalence class $[a] = \{gag^{-1}|g \in G\}$ is called the **conjugacy class**.

2.2 Vector spaces

2.2.1 Vectors and vector spaces

A **vector space** (or a **linear space**) V over a field K is a set in which two operations, addition and multiplication by an element of K (called a **scalar**), are defined. (In this book we are mainly interested in $K = \mathbb{R}$ and $\mathbb{C}$.) The elements (called **vectors**) of V satisfy the following axioms:

(i) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.
(ii) $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$.
(iii) There exists a zero vector $\mathbf{0}$ such that $\boldsymbol{v} + \mathbf{0} = \boldsymbol{v}$.
(iv) For any $\boldsymbol{u}$, there exists $-\boldsymbol{u}$, such that $\boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}$.
(v) $c(\boldsymbol{u} + \boldsymbol{v}) = c\boldsymbol{u} + c\boldsymbol{v}$.
(vi) $(c + d)\boldsymbol{u} = c\boldsymbol{u} + d\boldsymbol{u}$.
(vii) $(cd)\boldsymbol{u} = c(d\boldsymbol{u})$.
(viii) $1\boldsymbol{u} = \boldsymbol{u}$.

Here $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and $c, d \in K$ and 1 is the unit element of K.

Let $\{\boldsymbol{v}_i\}$ be a set of k (>0) vectors. If the equation

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_k\boldsymbol{v}_k = \boldsymbol{0} \tag{2.6}$$

has a non-trivial solution, $x_i \neq 0$ for some i, the set of vectors $\{\boldsymbol{v}_j\}$ is called **linearly dependent**, while if (2.6) has only a trivial solution, $x_i = 0$ for any i, $\{\boldsymbol{v}_i\}$ is said to be **linearly independent**. If at least one of the vectors is a zero vector $\boldsymbol{0}$, the set is always linearly dependent.

A set of linearly independent vectors $\{\boldsymbol{e}_i\}$ is called a basis of V, if any element $\boldsymbol{v} \in V$ is written *uniquely* as a linear combination of $\{\boldsymbol{e}_i\}$:

$$\boldsymbol{v} = v^1\boldsymbol{e}_1 + v^2\boldsymbol{e}_2 + \cdots + v^n\boldsymbol{e}_n. \tag{2.7}$$

The numbers $v^i \in K$ are called the **components** of $\boldsymbol{v}$ with respect to the basis $\{\boldsymbol{e}_j\}$. If there are n elements in the basis, the dimension of V is n, denoted by $\dim V = n$. We usually write the n-dimensional vector space over K as $V(n, K)$ (or simply V if n and K are understood from the context). We assume n is finite.

2.2.2 Linear maps, images and kernels

Given two vector spaces V and W, a map $f : V \to W$ is called a **linear map** if it satisfies $f(a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2) = a_1 f(\boldsymbol{v}_1) + a_2 f(\boldsymbol{v}_2)$ for any $a_1, a_2 \in K$ and $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$. A linear map is an example of a homomorphism that preserves the vector addition and the scalar multiplication. The **image** of f is $f(V) \subset W$ and the **kernel** of f is $\{\boldsymbol{v} \in V | f(\boldsymbol{v}) = \boldsymbol{0}\}$ and denoted by $\mathrm{im}\, f$ and $\ker f$ respectively. $\ker f$ cannot be empty since $f(\boldsymbol{0})$ is always $\boldsymbol{0}$. If W is the field K itself, f is called a **linear function**. If f is an isomorphism, V is said to be **isomorphic** to W and *vice versa*, denoted by $V \cong W$. It then follows that $\dim V = \dim W$. In fact, all the n-dimensional vector spaces are isomorphic to K^n, and they are regarded as identical vector spaces. The isomorphism between the vector spaces is an element of $\mathrm{GL}(n, K)$.

Theorem 2.1. If $f : V \to W$ is a linear map, then

$$\dim V = \dim(\ker f) + \dim(\mathrm{im} f). \tag{2.8}$$

Proof. Since f is a linear map, it follows that $\ker f$ and $\mathrm{im}\, f$ are vector spaces, see exercise 2.8. Let the basis of $\ker f$ be $\{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_r\}$ and that of $\mathrm{im}\, f$ be $\{\boldsymbol{h}'_1, \ldots, \boldsymbol{h}'_s\}$. For each i $(1 \leq i \leq s)$, take $\boldsymbol{h}_i \in V$ such that $f(\boldsymbol{h}_i) = \boldsymbol{h}'_i$ and consider the set of vectors $\{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_r, \boldsymbol{h}_1, \ldots, \boldsymbol{h}_s\}$.

Now we show that these vectors form a linearly independent basis of V. Take an arbitrary vector $\boldsymbol{v} \in V$. Since $f(\boldsymbol{v}) \in \mathrm{im}\, f$, it can be expanded as $f(\boldsymbol{v}) = c^i\boldsymbol{h}'_i = c^i f(\boldsymbol{h}_i)$. From the linearity of f, it then follows that $f(\boldsymbol{v} - c^i\boldsymbol{h}_i) = \boldsymbol{0}$, that is $\boldsymbol{v} - c^i\boldsymbol{h}_i \in \ker f$. This shows that an arbitrary vector $\boldsymbol{v}$ is a linear combination of $\{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_r, \boldsymbol{h}_1, \ldots, \boldsymbol{h}_s\}$. Thus, V is spanned by $r + s$ vectors. Next let us assume $a^i \boldsymbol{g}_i + b^i \boldsymbol{h}_i = \mathbf{0}$. Then $\mathbf{0} = f(\mathbf{0}) = f(a^i \boldsymbol{g}_i + b^i \boldsymbol{h}_i) = b^i f(\boldsymbol{h}_i) = b^i \boldsymbol{h}'_i$, which implies that $b^i = 0$. Then it follows from $a^i \boldsymbol{g}_i = \mathbf{0}$ that $a^i = 0$, and the set $\{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_r, \boldsymbol{h}_1, \ldots, \boldsymbol{h}_s\}$ is linearly independent in V. Finally we find $\dim V = r + s = \dim(\ker f) + \dim(\operatorname{im} f)$. □

[*Remark*: The vector space spanned by $\{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_s\}$ is called the **orthogonal complement** of $\ker f$ and is denoted by $(\ker f)^{\perp}$.]

Exercise 2.8. (1) Let $f : V \to W$ be a linear map. Show that both $\ker f$ and $\operatorname{im} f$ are vector spaces.

(2) Show that a linear map $f : V \to V$ is an isomorphism if and only if $\ker f = \{\mathbf{0}\}$.

2.2.3 Dual vector space

The dual vector space has already been introduced in section 1.2 in the context of quantum mechanics. The exposition here is more mathematical and complements the materials presented there.

Let $f : V \to K$ be a linear function on a vector space $V(n, K)$ over a field K. Let $\{\boldsymbol{e}_i\}$ be a basis and take an arbitrary vector $\boldsymbol{v} = v^1 \boldsymbol{e}_1 + \cdots + v^n \boldsymbol{e}_n$. From the linearity of f, we have $f(\boldsymbol{v}) = v^1 f(\boldsymbol{e}_1) + \cdots + v^n f(\boldsymbol{e}_n)$. Thus, if we know $f(\boldsymbol{e}_i)$ for all i, we know the result of the operation of f on any vector. It is remarkable that the set of linear functions is made into a vector space, namely a linear combination of two linear functions is also a linear function.

$$(a_1 f_1 + a_2 f_2)(\boldsymbol{v}) = a_1 f_1(\boldsymbol{v}) + a_2 f_2(\boldsymbol{v}) \tag{2.9}$$

This linear space is called the **dual vector space** to $V(n, K)$ and is denoted by $V^*(n, K)$ or simply by V^*. If $\dim V$ is finite, $\dim V^*$ is equal to $\dim V$. Let us introduce a basis $\{e^{*i}\}$ of V^*. Since e^{*i} is a linear function it is completely specified by giving $e^{*i}(\boldsymbol{e}_j)$ for all j. Let us choose the **dual basis**,

$$e^{*i}(\boldsymbol{e}_j) = \delta^i_j. \tag{2.10}$$

Any linear function f, called a **dual vector** in this context, is expanded in terms of $\{e^{*i}\}$,

$$f = f_i e^{*i}. \tag{2.11}$$

The action of f on $\boldsymbol{v}$ is interpreted as an **inner product** between a column vector and a row vector,

$$f(\boldsymbol{v}) = f_i e^{*i}(v^j \boldsymbol{e}_j) = f_i v^j e^{*i}(\boldsymbol{e}_j) = f_i v^i. \tag{2.12}$$

We sometimes use the notation $\langle\ ,\ \rangle : V^* \times V \to K$ to denote the inner product.

Let V and W be vector spaces with a linear map $f : V \to W$ and let $g : W \to K$ be a linear function on W ($g \in W^*$). It is easy to see that the

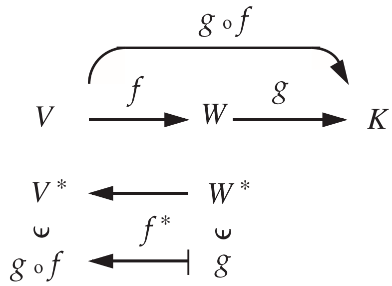

Figure 2.9. The pullback of a function g is a function $f^*(g) = g \circ f$.

composite map $g \circ f$ is a linear function on V. Thus, f and g give rise to an element $h \in V^*$ defined by

$$h(\boldsymbol{v}) \equiv g(f(\boldsymbol{v})) \qquad \boldsymbol{v} \in V. \tag{2.13}$$

Given $g \in W^*$, a map $f : V \to W$ has induced a map $h \in V^*$. Accordingly, we have an induced map $f^* : W^* \to V^*$ defined by $f^* : g \mapsto h = f^*(g)$, see figure 2.9. The map h is called the **pullback** of g by f^*.

Since $\dim V^* = \dim V$, there exists an isomorphism between V and V^*. However, this isomorphism is not canonical; we have to specify an inner product in V to define an isomorphism between V and V^* and *vice versa*, see the next section. The equivalence of a vector space and its dual vector space will appear recurrently in due course.

Exercise 2.9. Suppose $\{\boldsymbol{f}_j\}$ is another basis of V and $\{f^{*i}\}$ the dual basis. In terms of the old basis, $\boldsymbol{f}_i$ is written as $\boldsymbol{f}_i = A_i{}^j \boldsymbol{e}_j$ where $A \in \mathrm{GL}(n, K)$. Show that the dual bases are related by $e^{*i} = f^{*j} A_j{}^i$.

2.2.4 Inner product and adjoint

Let $V = V(m, K)$ be a vector space with a basis $\{\boldsymbol{e}_i\}$ and let g be a vector space isomorphism $g : V \to V^*$, where g is an arbitrary element of $\mathrm{GL}(m, K)$. The component representation of g is

$$g : v^j \to g_{ij} v^j. \tag{2.14}$$

Once this isomorphism is given, we may define the **inner product** of two vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ by

$$g(\boldsymbol{v}_1, \boldsymbol{v}_2) \equiv \langle g\boldsymbol{v}_1, \boldsymbol{v}_2 \rangle. \tag{2.15}$$

Let us assume that the field K is a real number $\mathbb{R}$. for definiteness. Then equation (2.15) has a component expression,

$$g(\boldsymbol{v}_1, \boldsymbol{v}_2) = v_1{}^i g_{ji} v_2{}^j. \tag{2.16}$$

We require that the matrix (g_{ij}) be positive definite so that the inner product $g(\boldsymbol{v}, \boldsymbol{v})$ has the meaning of the squared norm of $\boldsymbol{v}$. We also require that the metric be symmetric: $g_{ij} = g_{ji}$ so that $g(\boldsymbol{v}_1, \boldsymbol{v}_2) = g(\boldsymbol{v}_2, \boldsymbol{v}_1)$.

Next, let $W = W(n, \mathbb{R})$ be a vector space with a basis $\{\boldsymbol{f}_\alpha\}$ and a vector space isomorphism $G : W \to W^*$. Given a map $f : V \to W$, we may define the **adjoint** of f, denoted by $\tilde{f}$, by

$$G(\boldsymbol{w}, f\boldsymbol{v}) = g(\boldsymbol{v}, \tilde{f}\boldsymbol{w}) \tag{2.17}$$

where $\boldsymbol{v} \in V$ and $\boldsymbol{w} \in W$. It is easy to see that $\widetilde{(\tilde{f})} = f$. The component expression of equation (2.17) is

$$w^\alpha G_{\alpha\beta} f^\beta{}_i v^i = v^i g_{ij} \tilde{f}^j{}_\alpha w^\alpha \tag{2.18}$$

where $f^\beta{}_i$ and $\tilde{f}^j{}_\alpha$ are the matrix representations of f and $\tilde{f}$ respectively. If $g_{ij} = \delta_{ij}$ and $G_{\alpha\beta} = \delta_{\alpha\beta}$, the adjoint $\tilde{f}$ reduces to the transpose f^{t} of the matrix f.

Let us show that $\dim \operatorname{im} f = \dim \operatorname{im} \tilde{f}$. Since (2.18) holds for any $\boldsymbol{v} \in V$ and $\boldsymbol{w} \in W$, we have $G_{\alpha\beta} f^\beta{}_i = g_{ij} \tilde{f}^j{}_\alpha$, that is

$$\tilde{f} = g^{-1} f^{\mathrm{t}} G^{\mathrm{t}}. \tag{2.19}$$

Making use of the result of the following exercise, we obtain $\operatorname{rank} f = \operatorname{rank} \tilde{f}$, where the rank of a map is defined by that of the corresponding matrix (note that $g \in \mathrm{GL}(m, \mathbb{R})$ and $G \in \mathrm{GL}(n, \mathbb{R})$). It is obvious that $\dim \operatorname{im} f$ is the rank of a matrix representing the map f and we conclude $\dim \operatorname{im} f = \dim \operatorname{im} \tilde{f}$.

Exercise 2.10. Let $V = V(m, \mathbb{R})$ and $W = W(n, \mathbb{R})$ and let f be a matrix corresponding to a linear map from V to W. Verify that $\operatorname{rank} f = \operatorname{rank} f^{\mathrm{t}} = \operatorname{rank}(M f^{\mathrm{t}} N)$, where $M \in \mathrm{GL}(m, \mathbb{R})$ and $N \in \mathrm{GL}(n, \mathbb{R})$.

Exercise 2.11. Let V be a vector space over $\mathbb{C}$. The inner product of two vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ is defined by

$$g(\boldsymbol{v}_1, \boldsymbol{v}_2) = \overline{v_1}^i g_{ij} v_2{}^j \tag{2.20}$$

where $\bar{\ }$ denotes the complex conjugate. From the positivity and symmetry of the inner product, $g(\boldsymbol{v}_1, \boldsymbol{v}_2) = \overline{g(\boldsymbol{v}_2, \boldsymbol{v}_1)}$, the vector space isomorphism $g : V \to V^*$ is required to be a positive-definite Hermitian matrix. Let $f : V \to W$ be a (complex) linear map and $G : W \to W^*$ be a vector space isomorphism. The adjoint of f is defined by $g(\boldsymbol{v}, \tilde{f}\boldsymbol{w}) = \overline{G(\boldsymbol{w}, f\boldsymbol{v})}$. Repeat the analysis to show that

(a) $\tilde{f} = g^{-1} f^\dagger G^\dagger$, where † denotes the Hermitian conjugate, and
(b) $\dim \operatorname{im} f = \dim \operatorname{im} \tilde{f}$.

Theorem 2.2. (**Toy index theorem**) Let V and W be finite-dimensional vector spaces over a field K and let $f : V \to W$ be a linear map. Then

$$\dim \ker f - \dim \ker \tilde{f} = \dim V - \dim W. \tag{2.21}$$

Proof. Theorem 2.1 tells us that

$$\dim V = \dim\ker f + \dim\operatorname{im} f$$

and, if applied to $\tilde{f} : W \to V$,

$$\dim W = \dim\ker \widetilde{f} + \dim\operatorname{im} \widetilde{f}.$$

We saw earlier that $\dim\operatorname{im} f = \dim\operatorname{im} \widetilde{f}$, from which we obtain

$$\dim V - \dim\ker f = \dim W - \dim\ker \widetilde{f}. \qquad \square$$

Note that in (2.21), each term on the LHS depends on the details of the map f. The RHS states, however, that the *difference* in the two terms is independent of f! This may be regarded as a finite-dimensional analogue of the index theorems, see chapter 12.

2.2.5 Tensors

A dual vector is a linear object that maps a vector to a scalar. This may be generalized to multilinear objects called **tensors**, which map several vectors and dual vectors to a scalar. A tensor T of type (p, q) is a multilinear map that maps p dual vectors and q vectors to $\mathbb{R}$,

$$T : \bigotimes^{p} V^* \bigotimes^{q} V \to \mathbb{R}. \tag{2.22}$$

For example, a tensor of type $(0, 1)$ maps a vector to a real number and is identified with a dual vector. Similarly, a tensor of type $(1, 0)$ is a vector. If ω maps a dual vector and two vectors to a scalar, $\omega : V^* \times V \times V \to \mathbb{R}$, ω is of type $(1, 2)$.

The set of all tensors of type (p, q) is called the **tensor space** of type (p, q) and denoted by $\mathcal{T}^p_q$. The **tensor product** $\tau = \mu \otimes \nu \in \mathcal{T}^p_q \otimes \mathcal{T}^{p'}_{q'}$ is an element of $\mathcal{T}^{p+p'}_{q+q'}$ defined by

$$\begin{aligned}\tau(\omega_1, \ldots, \omega_p, \xi_1, \ldots, \xi_{p'}; u_1, \ldots, u_q, v_1, \ldots, v_{q'})\\ = \mu(\omega_1, \ldots, \omega_p; u_1, \ldots, u_q)\nu(\xi_1, \ldots, \xi_{p'}; v_1, \ldots, v_{q'}).\end{aligned} \tag{2.23}$$

Another operation in a tensor space is the **contraction**, which is a map from a tensor space of type (p, q) to type $(p - 1, q - 1)$ defined by

$$\tau(\ldots, e^{*i}, \ldots; \ldots, e_i, \ldots) \tag{2.24}$$

where $\{\boldsymbol{e}_i\}$ and $\{e^{*i}\}$ are the dual bases.

Exercise 2.12. Let V and W be vector spaces and let $f : V \to W$ be a linear map. Show that f is a tensor of type $(1, 1)$.

2.3 Topological spaces

The most general structure with which we work is a topological space. Physicists often tend to think that all the spaces they deal with are equipped with metrics. However, this is not always the case. In fact, metric spaces form a subset of manifolds and manifolds form a subset of topological spaces.

2.3.1 Definitions

Definition 2.3. Let X be any set and $\mathcal{T} = \{U_i | i \in I\}$ denote a certain collection of subsets of X. The pair $(X, \mathcal{T})$ is a **topological space** if $\mathcal{T}$ satisfies the following requirements.

(i) $\emptyset, X \in \mathcal{T}$.
(ii) If $\mathcal{T}$ is any (maybe infinite) subcollection of I, the family $\{U_j | j \in J\}$ satisfies $\cup_{j \in J} U_j \in \mathcal{T}$.
(iii) If K is any *finite* subcollection of I, the family $\{U_k | k \in K\}$ satisfies $\cap_{k \in K} U_k \in \mathcal{T}$.

X alone is sometimes called a topological space. The U_i are called the **open sets** and $\mathcal{T}$ is said to give a **topology** to X.

Example 2.7. (a) If X is a set and $\mathcal{T}$ is the collection of *all* the subsets of X, then (i)–(iii) are automatically satisfied. This topology is called the **discrete topology**.

(b) Let X be a set and $\mathcal{T} = \{\emptyset, X\}$. Clearly $\mathcal{T}$ satisfies (i)–(iii). This topology is called the **trivial topology**. In general the discrete topology is too stringent while the trivial topology is too trivial to give any interesting structures on X.

(c) Let X be the real line $\mathbb{R}$. All open intervals (a, b) and their unions define a topology called the **usual topology**; a and b may be $-\infty$ and ∞ respectively. Similarly, the usual topology in $\mathbb{R}^n$ can be defined. [Take a product $(a_1, b_1) \times \cdots \times (a_n, b_n)$ and their unions….]

Exercise 2.13. In definition 2.3, axioms (ii) and (iii) look somewhat unbalanced. Show that, if we allow infinite intersection in (iii), the usual topology in $\mathbb{R}$ reduces to the discrete topology (and is thus not very interesting).

A **metric** $d : X \times X \to \mathbb{R}$ is a function that satisfies the conditions:

(i) $d(x, y) = d(y, x)$
(ii) $d(x, y) \geq 0$ where the equality holds if and only if $x = y$
(iii) $d(x, y) + d(y, z) \geq d(x, z)$

for any $x, y, z \in X$. If X is endowed with a metric d, X is made into a topological space whose open sets are given by 'open discs',

$$U_\varepsilon(X) = \{y \in X | d(x, y) < \varepsilon\} \tag{2.25}$$

and all their possible unions. The topology $\mathcal{T}$ thus defined is called the **metric topology** determined by d. The topological space $(X, \mathcal{T})$ is called a **metric space**. [*Exercise*: Verify that a metric space $(X, \mathcal{T})$ is indeed a topological space.]

Let $(X, \mathcal{T})$ be a topological space and A be any subset of X. Then $\mathcal{T} = \{U_i\}$ induces the **relative topology** in A by $\mathcal{T}' = \{U_i \cap A | U_i \in \mathcal{T}\}$.

Example 2.8. Let $X = \mathbb{R}^{n+1}$ and take the n-sphere S^n,

$$(x^0)^2 + (x^1)^2 + \cdots + (x^n)^2 = 1. \tag{2.26}$$

A topology in S^n may be given by the relative topology induced by the usual topology on $\mathbb{R}^{n+1}$.

2.3.2 Continuous maps

Definition 2.4. Let X and Y be topological spaces. A map $f : X \to Y$ is **continuous** if the *inverse* image of an open set in Y is an open set in X.

This definition is in agreement with our intuitive notion of continuity. For instance, let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} -x + 1 & x \leq 0 \\ -x + \frac{1}{2} & x > 0. \end{cases} \tag{2.27}$$

We take the usual topology in $\mathbb{R}$, hence any open interval (a, b) is an open set. In the usual calculus, f is said to have a discontinuity at $x = 0$. For an open set $(3/2, 2) \subset Y$, we find $f^{-1}((3/2, 2)) = (-1, -1/2)$ which is an open set in X. If we take an open set $(1 - 1/4, 1 + 1/4) \subset Y$, however, we find $f^{-1}((1 - 1/4, 1 + 1/4)) = (-1/4, 0]$ which is not an open set in the usual topology.

Exercise 2.14. By taking a continuous function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ as an example, show that the reverse definition, '*a map f is continuous if it maps an open set in X to an open set in Y*', does not work. [*Hint*: Find where $(-\varepsilon, +\varepsilon)$ is mapped to under f.]

2.3.3 Neighbourhoods and Hausdorff spaces

Definition 2.5. Suppose $\mathcal{T}$ gives a topology to X. N is a **neighbourhood** of a point $x \in X$ if N is a subset of X and N contains some (at least one) open set U_i to which x belongs. (The subset N need not be an open set. If N happens to be an open set in $\mathcal{T}$, it is called an **open neighbourhood**.)

Example 2.9. Take $X = \mathbb{R}$ with the usual topology. The interval $[-1, 1]$ is a neighbourhood of an arbitrary point $x \in (-1, 1)$.

Definition 2.6. A topological space $(X, \mathcal{T})$ is a **Hausdorff space** if, for an arbitrary pair of distinct points $x, x' \in X$, there always exist neighbourhoods U_x of x and $U_{x'}$ of x' such that $U_x \cap U_{x'} = \emptyset$.

Exercise 2.15. Let $X = \{\text{John, Paul, Ringo, George}\}$ and $U_0 = \emptyset, U_1 = \{\text{John}\}, U_2 = \{\text{John, Paul}\}, U_3 = \{\text{John, Paul, Ringo, George}\}$. Show that $\mathcal{T} = \{U_0, U_1, U_2, U_3\}$ gives a topology to X. Show also that $(X, \mathcal{T})$ is not a Hausdorff space.

Unlike this exercise, most spaces that appear in physics satisfy the Hausdorff property. In the rest of the present book we always assume this is the case.

Exercise 2.16. Show that $\mathbb{R}$ with the usual topology is a Hausdorff space. Show also that any metric space is a Hausdorff space.

2.3.4 Closed set

Let $(X, \mathcal{T})$ be a topological space. A subset A of X is **closed** if its complement in X is an open set, that is $X - A \in \mathcal{T}$. According to the definition, X and $\emptyset$ are both open *and* closed. Consider a set A (either open or closed). The **closure** of A is the smallest closed set that contains A and is denoted by $\bar{A}$. The **interior** of A is the largest open subset of A and is denoted by A°. The **boundary** $b(A)$ of A is the complement of A° in A; $b(A) = A - A^\circ$. An open set is always disjoint from its boundary while a closed set always contains its boundary.

Example 2.10. Take $X = \mathbb{R}$ with the usual topology and take a pair of open intervals $(-\infty, a)$ and (b, ∞) where $a < b$. Since $(-\infty, a) \cup (b, \infty)$ is open under the usual topology, the complement $[a, b]$ is closed. Any closed interval is a closed set under the usual topology. Let $A = (a, b)$, then $\bar{A} = [a, b]$. The boundary $b(A)$ consists of two points $\{a, b\}$. The sets (a, b), $[a, b]$, $(a, b]$, and $[a, b)$ all have the same boundary, closure and interior. In $\mathbb{R}^n$, the product $[a_1, b_1] \times \cdots \times [a_n, b_n]$ is a closed set under the usual topology.

Exercise 2.17. Whether a set $A \subset X$ is open or closed depends on X. Let us take an interval $I = (0, 1)$ in the x-axis. Show that I is open in the x-axis $\mathbb{R}$ while it is neither closed nor open in the xy-plane $\mathbb{R}^2$.

2.3.5 Compactness

Let $(X, \mathcal{T})$ be a topological space. A family $\{A_i\}$ of subsets of X is called a **covering** of X, if

$$\bigcup_{i \in I} A_i = X.$$

If all the A_i happen to be the open sets of the topology $\mathcal{T}$, the covering is called an **open covering**.

Definition 2.7. Consider a set X and all possible coverings of X. The set X is **compact** if, for every open covering $\{U_i | i \in I\}$, there exists a *finite* subset J of I such that $\{U_j | j \in J\}$ is also a covering of X.

In general, if a set is compact in $\mathbb{R}^n$, it must be bounded. What else is needed? We state the result without the proof.

Theorem 2.3. Let X be a subset of $\mathbb{R}^n$. X is compact if and only if it is *closed* and *bounded*.

Example 2.11. (a) A point is compact.

(b) Take an open interval (a, b) in $\mathbb{R}$ and choose an open covering $U_n = (a, b - 1/n)$, $n \in \mathbb{N}$. Evidently

$$\bigcup_{n \in \mathbb{Z}} U_n = (a, b).$$

However, no finite subfamily of $\{U_n\}$ covers (a, b). Thus, an open interval (a, b) is non-compact in conformity with theorem 2.3.

(c) S^n in example 2.8 with the relative topology is compact, since it is closed and bounded in $\mathbb{R}^{n+1}$.

The reader might not appreciate the significance of compactness from the definition and the few examples given here. It should be noted, however, that some mathematical analyses as well as physics become rather simple on a compact space. For example, let us consider a system of electrons in a solid. If the solid is non-compact with infinite volume, we have to deal with quantum statistical mechanics in an infinite volume. It is known that this is mathematically quite complicated and requires knowledge of the advanced theory of Hilbert spaces. What we usually do is to confine the system in a finite volume V surrounded by hard walls so that the electron wavefunction vanishes at the walls, or to impose periodic boundary conditions on the walls, which amounts to putting the system in a torus, see example 2.5(b). In any case, the system is now put in a compact space. Then we may construct the Fock space whose excitations are labelled by discrete indices. Another significance of compactness in physics will be found when we study extended objects such as instantons and Belavin–Polyakov monopoles, see section 4.8. In field theories, we usually assume that the field approaches some asymptotic form corresponding to the vacuum (or one of the vacua) at spatial infinities. Similarly, a class of order parameter distributions in which the spatial infinities have a common order parameter is an interesting class to study from various points of view as we shall see later. Since all points at infinity are mapped to a point, we have effectively compactified the non-compact space $\mathbb{R}^n$ to a compact space $S^n = \mathbb{R}^n \cup \{\infty\}$. This procedure is called the **one-point compactification**.

2.3.6 Connectedness

Definition 2.8. (a) A topological space X is **connected** if it cannot be written as $X = X_1 \cup X_2$, where X_1 and X_2 are both open and $X_1 \cap X_2 = \emptyset$. Otherwise X is called **disconnected**.

(b) A topological space X is called **arcwise connected** if, for any points $x, y \in X$, there exists a continuous map $f : [0, 1] \to X$ such that $f(0) = x$ and $f(1) = y$. With a few pathological exceptions, arcwise connectedness is practically equivalent to connectedness.

(c) A **loop** in a topological space X is a continuous map $f : [0, 1] \to X$ such that $f(0) = f(1)$. If any loop in X can be continuously shrunk to a point, X is called **simply connected**.

Example 2.12. (a) The real line $\mathbb{R}$ is arcwise connected while $\mathbb{R} - \{0\}$ is not. $\mathbb{R}^n$ $(n \geq 2)$ is arcwise connected and so is $\mathbb{R}^n - \{\mathbf{0}\}$.

(b) S^n is arcwise connected. The circle S^1 is not simply connected. If $n \geq 2$, S^n is simply connected. The n-dimensional torus

$$T^n = \underbrace{S^1 \times S^1 \times \cdots \times S^1}_{n} \qquad (n \geq 2)$$

is arcwise connected but not simply connected.

(c) $\mathbb{R}^2 - \mathbb{R}$ is not arcwise connected. $\mathbb{R}^2 - \{\mathbf{0}\}$ is arcwise connected but not simply connected. $\mathbb{R}^3 - \{\mathbf{0}\}$ is arcwise connected and simply connected.

2.4 Homeomorphisms and topological invariants

2.4.1 Homeomorphisms

As we mentioned at the beginning of this chapter, the main purpose of topology is to classify spaces. Suppose we have several figures and ask ourselves which are equal and which are different. Since we have not defined what is meant by *equal* or *different*, we may say ‘they are all different from each other’ or ‘they are all the same figures’. Some of the definitions of equivalence are too stringent and some are too loose to produce any sensible classification of the figures or spaces. For example, in elementary geometry, the equivalence of figures is given by congruence, which turns out to be too stringent for our purpose. In topology, we define two figures to be equivalent if it is possible to deform one figure into the other by *continuous deformation*. Namely we introduce the equivalence relation under which geometrical objects are classified according to whether it is possible to deform one object into the other by continuous deformation. To be more mathematical, we need to introduce the following notion of homeomorphism.

Definition 2.9. Let X_1 and X_2 be topological spaces. A map $f : X_1 \to X_2$ is a **homeomorphism** if it is continuous and has an inverse $f^{-1} : X_2 \to X_1$ which is

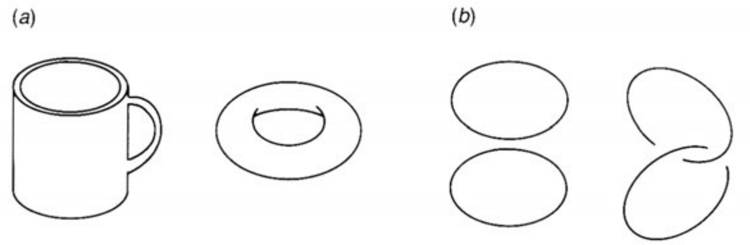

Figure 2.10. (*a*) A coffee cup is homeomorphic to a doughnut. (*b*) The linked rings are homeomorphic to the separated rings.

also continuous. If there exists a homeomorphism between X_1 and X_2, X_1 is said to be **homeomorphic** to X_2 and *vice versa*.

In other words, X_1 is homeomorphic to X_2 if there exist maps $f : X_1 \to X_2$ and $g : X_2 \to X_1$ such that $f \circ g = \mathrm{id}_{X_2}$, and $g \circ f = \mathrm{id}_{X_1}$. It is easy to show that a homeomorphism is an equivalence relation. Reflectivity follows from the choice $f = \mathrm{id}_X$, while symmetry follows since if $f : X_1 \to X_2$ is a homeomorphism so is $f^{-1} : X_2 \to X_1$ by definition. Transitivity follows since, if $f : X_1 \to X_2$ and $g : X_2 \to X_3$ are homeomorphisms so is $g \circ f : X_1 \to X_3$. Now we divide all topological spaces into equivalence classes according to whether it is possible to deform one space into the other by a homeomorphism. Intuitively speaking, we suppose the topological spaces are made out of ideal rubber which we can deform at our will. Two topological spaces are homeomorphic to each other if we can deform one into the other *continuously*, that is, without tearing them apart or pasting.

Figure 2.10 shows some examples of homeomorphisms. It seems impossible to deform the left figure in figure 2.10(*b*) into the right one by continuous deformation. However, this is an artefact of the embedding of these objects in $\mathbb{R}^3$. In fact, they are continuously deformable in $\mathbb{R}^4$, see problem 2.3. To distinguish one from the other, we have to embed them in S^3, say, and compare the complements of these objects in S^3. This approach is, however, out of the scope of the present book and we will content ourselves with homeomorphisms.

2.4.2 Topological invariants

Now our main question is: '*How can we characterize the equivalence classes of homeomorphism?*' In fact, we do not know the complete answer to this question yet. Instead, we have a rather modest statement, that is, if two spaces have different '**topological invariants**', they are not homeomorphic to each other. Here topological invariants are those quantities which are conserved under homeomorphisms. A topological invariant may be a number such as the number of connected components of the space, an algebraic structure such as a group or a ring which is constructed out of the space, or something like connectedness, compactness or the Hausdorff property. (Although it seems to be intuitively clear that these are topological invariants, we have to prove that they indeed are. We omit the proofs. An interested reader may consult any text book on topology.) If we knew the complete set of topological invariants we could specify the equivalence class by giving these invariants. However, so far we know a partial set of topological invariants, which means that even if all the known topological invariants of two topological spaces coincide, they may not be homeomorphic to each other. Instead, what we can say at most is: *if two topological spaces have different topological invariants they cannot be homeomorphic to each other*.

Example 2.13. (a) A closed line $[-1, 1]$ is not homeomorphic to an open line $(-1, 1)$, since $[-1, 1]$ is compact while $(-1, 1)$ is not.

(b) A circle S^1 is not homeomorphic to $\mathbb{R}$, since S^1 is compact in $\mathbb{R}^2$ while $\mathbb{R}$ is not.

(c) A parabola ($y = x^2$) is not homeomorphic to a hyperbola ($x^2 - y^2 = 1$) although they are both non-compact. A parabola is (arcwise) connected while a hyperbola is not.

(d) A circle S^1 is not homeomorphic to an interval $[-1, 1]$, although they are both compact and (arcwise) connected. $[-1, 1]$ is simply connected while S^1 is not. Alternatively $S^1 - \{p\}$, p being any point in S^1 is connected while $[-1, 1] - \{0\}$ is not, which is more evidence against their equivalence.

(e) Surprisingly, an interval without the endpoints is homeomorphic to a line $\mathbb{R}$. To see this, let us take $X = (-\pi/2, \pi/2)$ and $Y = \mathbb{R}$ and let $f : X \to Y$ be $f(x) = \tan x$. Since $\tan x$ is one to one on X and has an inverse, $\tan^{-1} x$, which is one to one on $\mathbb{R}$, this is indeed a homeomorphism. Thus, *boundedness* is not a topological invariant.

(f) An open disc $D^2 = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 < 1\}$ is homeomorphic to $\mathbb{R}^2$. A homeomorphism $f : D^2 \to \mathbb{R}^2$ may be

$$f(x, y) = \left(\frac{x}{\sqrt{1 - x^2 - y^2}}, \frac{y}{\sqrt{1 - x^2 - y^2}} \right) \tag{2.28}$$

while the inverse $f^{-1} : \mathbb{R}^2 \to D^2$ is

$$f^{-1}(x, y) = \left(\frac{x}{\sqrt{1 + x^2 + y^2}}, \frac{y}{\sqrt{1 + x^2 + y^2}} \right). \tag{2.29}$$

The reader should verify that $f \circ f^{-1} = \mathrm{id}_{\mathbb{R}^2}$, and $f^{-1} \circ f = \mathrm{id}_{D^2}$. As we saw in example 2.5(e), a closed disc whose boundary S^1 corresponds to a point is homeomorphic to S^2. If we take this point away, we have an open disc. The present analysis shows that this open disc is homeomorphic to $\mathbb{R}^2$. By reversing the order of arguments, we find that if we add a point (infinity) to $\mathbb{R}^2$, we obtain a compact space S^2. This procedure is the one-point compactification $S^2 = \mathbb{R}^2 \cup \{\infty\}$ introduced in the previous section. We similarly have $S^n = \mathbb{R}^n \cup \{\infty\}$.

(g) A circle $S^1 = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 = 1\}$ is homeomorphic to a square $I^2 = \{(x, y) \in \mathbb{R}^2 | (|x| = 1, |y| \leq 1), (|x| \leq 1, |y| = 1)\}$. A homeomorphism $f : I^2 \to S^1$ may be given by

$$f(x, y) = \left(\frac{x}{r}, \frac{y}{r}\right) \qquad r = \sqrt{x^2 + y^2}. \tag{2.30}$$

Since r cannot vanish, (2.27) is invertible.

Exercise 2.18. Find a homeomorphism between a circle $S^1 = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 = 1\}$ and an ellipse $E = \{(x, y) \in \mathbb{R}^2 | (x/a)^2 + (y/b)^2 = 1\}$.

2.4.3 Homotopy type

An equivalence class which is somewhat coarser than homeomorphism but which is still quite useful is 'of the **same homotopy type**'. We relax the conditions in definition 2.9 so that the continuous functions f or g need not have inverses. For example, take $X = (0, 1)$ and $Y = \{0\}$ and let $f : X \to Y$, $f(x) = 0$ and $g : Y \to X$, $g(0) = \frac{1}{2}$. Then $f \circ g = \mathrm{id}_Y$, while $g \circ f \neq \mathrm{id}_X$. This shows that an open interval $(0, 1)$ is of the same homotopy type as a point $\{0\}$, although it is not homeomorphic to $\{0\}$. We have more on this topic in section 4.2.

Example 2.14. (a) S^1 is of the same homotopy type as a cylinder, since a cylinder is a direct product $S^1 \times \mathbb{R}$ and we can shrink $\mathbb{R}$ to a point at each point of S^1. By the same reason, the Möbius strip is of the same homotopy type as S^1.

(b) A disc $D^2 = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 < 1\}$ is of the same homotopy type as a point. $D^2 - \{(0, 0)\}$ is of the same homotopy type as S^1. Similarly, $\mathbb{R}^2 - \{\mathbf{0}\}$ is of the same homotopy type as S^1 and $\mathbb{R}^3 - \{\mathbf{0}\}$ as S^2.

2.4.4 Euler characteristic: an example

The Euler characteristic is one of the most useful topological invariants. Moreover, we find the prototype of the algebraic approach to topology in it. To avoid unnecessary complication, we restrict ourselves to points, lines and surfaces in $\mathbb{R}^3$. A **polyhedron** is a geometrical object surrounded by faces. The boundary of two faces is an edge and two edges meet at a vertex. We extend the definition of a polyhedron a bit to include polygons and the boundaries of polygons, lines or points. We call the faces, edges and vertices of a polyhedron **simplexes**. Note that the boundary of two simplexes is either empty or another simplex. (For example, the boundary of two faces is an edge.) Formal definitions of a simplex and a polyhedron in a general number of dimensions will be given in chapter 3. We are now ready to define the Euler characteristic of a figure in $\mathbb{R}^3$.

Definition 2.10. Let X be a subset of $\mathbb{R}^3$, which is homeomorphic to a polyhedron K. Then the **Euler characteristic** $\chi(X)$ of X is defined by

$$\chi(X) = (\text{number of verticies in } K) - (\text{number of edges in } K) + (\text{number of faces in } K). \tag{2.31}$$

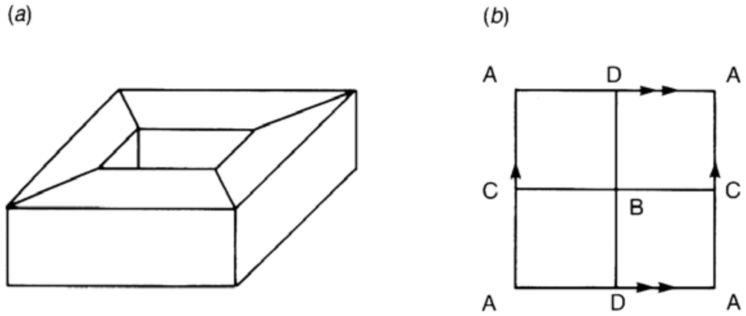

Figure 2.11. Example of a polyhedron which is homeomorphic to a torus.

The reader might wonder if $\chi(X)$ depends on the polyhedron K or not. The following theorem due to Poincaré and Alexander guarantees that it is, in fact, independent of the polyhedron K.

Theorem 2.4. (**Poincaré–Alexander**) The Euler characteristic $\chi(X)$ is independent of the polyhedron K as long as K is homeomorphic to X.

Examples are in order. The Euler characteristic of a point is $\chi(\cdot) = 1$ by definition. The Euler characteristic of a line is $\chi(\text{——}) = 2 - 1 = 1$, since a line has two vertices and an edge. For a triangular disc, we find $\chi(\text{triangle}) = 3 - 3 + 1 = 1$. An example which is a bit non-trivial is the Euler characteristic of S^1. The simplest polyhedron which is homeomorphic to S^1 is made of three edges of a triangle. Then $\chi(S^1) = 3 - 3 = 0$. Similarly, the sphere S^2 is homeomorphic to the surface of a tetrahedron, hence $\chi(S^2) = 4 - 6 + 4 = 2$. It is easily seen that S^2 is also homeomorphic to the surface of a cube. Using a cube to calculate the Euler characteristic of S^2, we have $\chi(S^2) = 8 - 12 + 6 = 2$, in accord with theorem 2.4. Historically this is the conclusion of **Euler's theorem**: if K is any polyhedron homeomorphic to S^2, with v vertices, e edges and f two-dimensional faces, then $v - e + f = 2$.

Example 2.15. Let us calculate the Euler characteristic of the torus T^2. Figure 2.11(a) is an example of a polyhedron which is homeomorphic to T^2. From this polyhedron, we find $\chi(T^2) = 16 - 32 + 16 = 0$. As we saw in example 2.5(b), T^2 is equivalent to a rectangle whose edges are identified; see figure 2.4. Taking care of this identification, we find an example of a polyhedron made of rectangular faces as in figure 2.11(b), from which we also have $\chi(T^2) = 0$. This approach is quite useful when the figure cannot be realized (embedded) in $\mathbb{R}^3$. For example, the Klein bottle (figure 2.5(a)) cannot be realized in $\mathbb{R}^3$ without intersecting itself. From the rectangle of figure 2.5(a), we find $\chi(\text{Klein bottle}) = 0$. Similarly, we have $\chi(\text{projective plane}) = 1$.

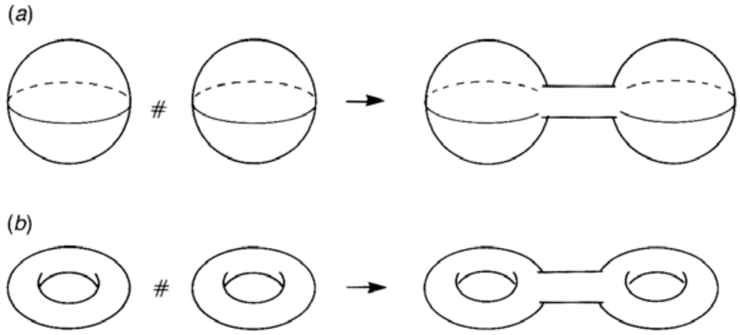

Figure 2.12. The connected sum. (*a*) $S^2 \sharp S^2 = S2$, (*b*) $T^2 \sharp T^2 = \Sigma_2$.

Exercise 2.19. (a) Show that χ(Möbius strip) $= 0$.

(b) Show that $\chi(\Sigma_2) = -2$, where Σ_2 is the torus with two handles (see example 2.5). The reader may either construct a polyhedron homeomorphic to Σ_2 or make use of the octagon in figure 2.6(*a*). We show later that $\chi(\Sigma_g) = 2 - 2g$, where Σ_g is the torus with g handles.

The **connected sum** $X \sharp Y$ of two surfaces X and Y is a surface obtained by removing a small disc from each of X and Y and connecting the resulting holes with a cylinder; see figure 2.12. Let X be an arbitrary surface. Then it is easy to see that

$$S^2 \sharp X = X \tag{2.32}$$

since S^2 and the cylinder may be deformed so that they fill in the hole on X; see figure 2.12(*a*). If we take a connected sum of two tori we get (figure 2.12(*b*))

$$T^2 \sharp T^2 = \Sigma_2. \tag{2.33}$$

Similarly, Σ_g may be given by the connected sum of g tori,

$$\underbrace{T^2 \sharp T^2 \sharp \cdots \sharp T^2}_{g \text{ factors}} = \Sigma_g. \tag{2.34}$$

The connected sum may be used as a trick to calculate an Euler characteristic of a complicated surface from those of known surfaces. Let us prove the following theorem.

Theorem 2.5. Let X and Y be two surfaces. Then the Euler characteristic of the connected sum $X \sharp Y$ is given by

$$\chi(X \sharp Y) = \chi(X) + \chi(Y) - 2.$$

Proof. Take polyhedra K_X and K_Y homeomorphic to X and Y, respectively. We assume, without loss of generality, that each of K_Y and K_Y has a triangle in it. Remove the triangles from them and connect the resulting holes with a trigonal cylinder. Then the number of vertices does not change while the number of edges increases by three. Since we have removed two faces and added three faces, the number of faces increases by $-2 + 3 = 1$. Thus, the change of the Euler characteristic is $0 - 3 + 1 = -2$. □

From the previous theorem and the equality $\chi(T^2) = 0$, we obtain $\chi(\Sigma_2) = 0 + 0 - 2 = -2$ and $\chi(\Sigma_g) = g \times 0 - 2(g - 1) = 2 - 2g$, cf exercise 2.19(b).

The significance of the Euler characteristic is that it is a topological invariant, which is calculated relatively easily. We accept, without proof, the following theorem.

Theorem 2.6. Let X and Y be two figures in $\mathbb{R}^3$. If X is homeomorphic to Y, then $\chi(X) = \chi(Y)$. In other words, if $\chi(X) \neq \chi(Y)$, X cannot be homeomorphic to Y.

Example 2.16. (a) S^1 is not homeomorphic to S^2, since $\chi(S^1) = 0$ while $\chi(S^2) = 2$.

(b) Two figures, which are not homeomorphic to each other, may have the same Euler characteristic. A point $(\cdot)$ is not homeomorphic to a line (——) but $\chi(\cdot) = \chi(\text{——}) = 1$. This is a general consequence of the following fact: *if a figure X is of the same homotopy type as a figure Y, then $\chi(X) = \chi(Y)$.*

The reader might have noticed that the Euler characteristic is different from other topological invariants such as compactness or connectedness in character. Compactness and connectedness are geometrical properties of a figure or a space while the Euler characteristic is an *integer* $\chi(X) \in \mathbb{Z}$. Note that $\mathbb{Z}$ is an algebraic object rather than a geometrical one. Since the work of Euler, many mathematicians have worked out the relation between geometry and algebra and elaborated this idea, in the last century, to establish combinatorial topology and algebraic topology. We may compute the Euler characteristic of a smooth surface by the celebrated Gauss–Bonnet theorem, which relates the integral of the Gauss curvature of the surface with the Euler characteristic calculated from the corresponding polyhedron. We will give the generalized form of the Gauss–Bonnet theorem in chapter 12.

Problems

2.1 Show that the $4g$-gon in figure 2.13(a), with the boundary identified, represents the torus with genus g of figure 2.13(b). The reader may use equation (2.34).

2.2 Let $X = \{1, 1/2, \ldots, 1/n, \ldots\}$ be a subset of $\mathbb{R}$. Show that X is not closed in $\mathbb{R}$. Show that $Y = \{1, 1/2, \ldots, 1/n, \ldots, 0\}$ is closed in $\mathbb{R}$, hence compact.

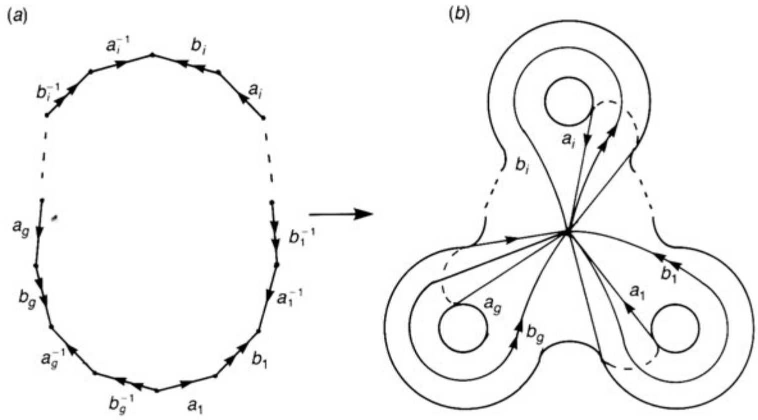

Figure 2.13. The polygon (*a*) whose edges are identified is the torus Σ_g with genus g.

2.3 Show that two figures in figure 2.109(*b*) are homeomorphic to each other. Find how to unlink the right figure in $\mathbb{R}^4$.

2.4 Show that there are only five regular polyhedra: a tetrahedron, a hexahedron, an octahedron, a dodecahedron and an icosahedron. [*Hint*: Use Euler's theorem.]