# Elementary Topology

J. B. Cooper
Johannes Kepler Universität Linz

## Contents

# 1 Introduction

The modern theory of topology draws its roots from two main sources. One is the theory of convergence and the related concepts of approximation which play such a central role in modern mathematics and its applications. Since the problems dealt with are of such complexity, the earlier ideal of obtaining exact and explicit solutions in closed forms usually must be replaced by methods which provide successive approximations to the solution. A familiar example of this is the Newton method for obtaining approximate roots of an equation of the form $f(x) = 0$ where $f$ is a suitable function. The second source is that branch of geometry which is often referred to as "rubber sheet geometry" i.e. the study of those properties of a geometrical object which remain unchanged under continuous deformations (in contrast, say, to Euclidean geometry which is concerned with those properties which remain unchanged under congruence). It is remarkable that the same abstract theory—topology—provides the framework for both topics. It is also rather fortunate, since an introduction to the more informal and geometrical aspects of topology is rather more digestible than a dry axiomatic approach which has no intuitive material to draw on. For this reason, we begin with a brief survey of topological notions for subsets of $\mathbf{R}^n$. The usual (euclidean) metric plays an important role here and this leads naturally to the introduction of the general notion of a metric in the second chapter.

With this more intuitive material available, abstract topological spaces are introduced in the next chapter and the various topological concepts are clarified in this context. We also provide a small collection of pathological spaces. The next chapter deals with the various methods of constructing topological spaces. In particular, the construction of the quotient space provides an opportunity for giving a brief survey of classical results on two-dimensional spaces. the next three chapters are devoted to special topological properties which eliminate much of the pathology that can occur in the most general situations. We then treat two variations of the notion of a topological space—uniformities and compactologies. In fact, the concept of a topological space is, in a certain sense, unsatisfactory and owes its pre-eminence more to a historical accident. For many applications, it can be profitably be replaced by one of the above ones. This will be particularly useful in the next section where we consider structures on spaces of mappings between topological spaces. We conclude the first chapter with a discussion of three of the most important special spaces—the Cantor set, the irrationals and the Hilbert cube.

## 2 The topology of subsets of $\mathbf{R}^n$

The familiar objects of study in geometry are subsets of two and three (or higher) dimensional space and have a topology which is intimately related to the euclidean distance between points. Recall that the latter is defined by the formula

$$d_2(x, y) = \sqrt{\sum_{i=1}^{n} (\xi_i - \eta_i)^2}$$

where $x = (\xi_i)$ and $y = (\eta_i)$. (The reason for incorporating the rather mysterious subscript "2" will become apparent below). For future reference we note that this distance satisfies the following natural conditions: 1) $d_2(x, y) \geq 0$ and $d_2(x, y) = 0$ if and only if $x = y$; 2) $d_2(x, y) = d_2(y, x)$; 3) $d_2(x, y) \leq d_2(x, y) + d_2(y, z)$ for points $x, y$, and $z$ in $\mathbf{R}^n$.

Using this distance function we can define the following concepts which will be familiar from an elementary analysis course:

1. A sequence $(x_k)$ in $\mathbf{R}^n$ converges to a point $x$ if and only if $d_2(x_k, y) \to 0$. Of course, this just means that the coordinates of the point $x_k$ converge to those of $x$ i.e. if $x_k = (\xi_i^k)$ and $x = (\xi_i)$, then $\xi_i^k \to \xi_i$ for each $i$.

2. A function $f : A \to B$ (where $A$ and $B$ are subsets of suitable euclidean spaces) is continuous if and only if for each $x_0 \in A$ and for each $\epsilon > 0$, there is a $\delta > 0$ so that $d_2(f(x), f(x_0) \leq \epsilon$ whenever $d_2(x, x_0) \leq \delta$.

3. Two such sets $A$ and $B$ are **homeomorphic** if there is a bijection $f$ from $A$ onto $B$ which is such that $f$ and its inverse $f^{-1}$ are both continuous. Such an $f$ is called a **homeomorphism**.

For example, the circle and the square are homeomorphic. However, the unit interval $[0, 1]$ is not homeomorphic to the unit circle. This is intuitively clear and can be proved by the following informal argument, which can be made precise by using the concepts of chapter I.5: if we remove any point from the unit circle, the latter remains in one piece (i.e. is *connected* in the terminology of I.5). The unit interval does not have this property. Hence the two spaces cannot be homeomorphic.

A further (perhaps surprising) example of two homeomorphic sets are the open interval $]0, 1[$ in $\mathbf{R}$ and $\mathbf{R}$ itself. Figure 1 displays a homeomorphism between these two spaces.

1. An **open ball** in $\mathbf{R}^n$ is a set of the form

$$U(x; \epsilon) = \{y \in \mathbf{R}^n : d_2(x, y) < \epsilon\}.$$

$x$ is the **centre** of the ball, $\epsilon$ the **radius**.

2. A subset $U$ of $\mathbf{R}^n$ is defined to be **open** if it is a union of open balls i.e. if for each $x$ in $U$ there is an $\epsilon > 0$ so that $U(x;\epsilon) \subset U$. The subset $C$ is **closed** if it contains the limit of any converging sequence in the set. This is equivalent to either of the following formulations: a) the complement $\mathbf{R}^n \backslash C$ is open; b) each point $x$ with the property that $U(x;\epsilon) \cap C \neq \emptyset$ for each $\epsilon > 0$ is in $C$. More generally, if $A$ is a subset of $\mathbf{R}^n$, then a set $U$ in $A$ is **relatively open** (or simply open in $A$) if it is of the form $U \cap A$ where $U$ is open in $\mathbf{R}^n$. Similarly, the **relatively closed** subsets of $A$ are those of the form $C \cap A$ where $C$ is closed in $\mathbf{R}^n$.

In terms of these concepts, the above definition of continuity can be reformulated as follows: $f : A \to B$ is continuous if and only if for each relatively open subset $U$ of $B$, $f^{-1}(U)$ is open in $A$.

Owing to the importance of some particular subsets of $\mathbf{R}^n$, we introduce special notations for some of those which occur most frequently:

1. $\mathbf{I}$ denote the unit interval $[0, 1]$ in the real line. Notice that $\mathbf{I}$ is homeomorphic to any closed, bounded interval $[a, b]$, a suitable homeomorphism being provided by the affine mapping $t \mapsto (b-a)t + a$ (cf. figure 2). $\mathbf{I}$ is not homeomorphic to any of the following intervals: $]0, 1[$, $[0, 1[$, $[0, \infty[$ etc.

2. More generally, $\mathbf{I}^n$ is the standard hypercube in $\mathbf{R}^n$ (cf. figure 3). This is the subset $\{x \in \mathbf{R}^n : 0 \leq \xi_i \leq 1 \text{ for each } i\}$ of $n$-space. Of course, it is just the Cartesian product of $n$ copies of $\mathbf{I}$.

3. $\mathbf{S}^1$ is the unit circle in two-space i.e. the set

$$\{x \in \mathbf{R}^2 : |x| = 1\}.$$

4. More generally, $\mathbf{S}^n = \{x \in \mathbf{R}^{n+1} : |x| = 1\}$ is the **$n$-dimensional hypersphere.** (Warning: it is *not* the Cartesian product of copies of the circle. This latter space is the object of the next example).

5. $\mathbf{T}^n$ (the **$n$-dimensional torus** or simply the $n$-torus) is the cartesian product $\mathbf{S}^1 \times \cdots \times \mathbf{S}^1$ of $n$-copies of the circle. As such, it is a subset of $\mathbf{R}^{2n}$. $\mathbf{T}^2 = \mathbf{S}^1 \times \mathbf{S}^1$ is the usual torus. As defined here it is a subset of $\mathbf{R}^4$ but one usually visualises it as a subset of $\mathbf{R}^3$ as in figure 4. The homeomorphism between these two surfaces is obtained as follows: each point in $\mathbf{S}^1$ has the complex representation $e^{i\theta}$ for some $\theta \in [0, 2\pi[$. We

identify the pair $(e^{i\theta}, e^{i\phi})$ in $\mathbf{S}^1 \times \mathbf{S}^1$ with the point on the ring-shaped torus with angular coordinates $\theta$ and $\phi$ as in figure 4.

6. $\mathbf{B}^n$ is the subset $\{x \in \mathbf{R}^n : |x| \leq 1\}$ of $\mathbf{R}^n$—it is called the **unit cell** of the latter space. Its "boundary" is $\mathbf{S}^{n-1}$ (the notion of the boundary of a set is also a topological one which will be introduced rigorously in chapter I.3).

Note that $\mathbf{B}^n$ and $\mathbf{I}^n$ are homeomorphic. It will often be convenient to denote topological spaces which are homeomorphic to standard ones from the above list by the same symbol since they are indistinguishable from the topological point of view. Thus the boundary of the ellipsoid

$$\{x \in \mathbf{R}^3 : \frac{\xi_1^2}{a^2} + \frac{\xi_2^2}{b^2} + \frac{\xi_3^2}{c^2} = 1\}$$

is homeomorphic to $\mathbf{S}^2$ and so may be denoted by this symbol.

Concerning the concept of homeomorphism, a few words of warning are in order. Firstly, in informal discussions on topology the property of being homeomorphic is often paraphrased by stating that the two spaces can be transformed into each other by means of smooth operations (such as stretching or shrinking, but not tearing). However, the two subsets of two-space shown in figure 5 are homeomorphic but cannot be deformed into each other (within the plane) by such operations. Often the appropriate concept for subsets of $\mathbf{R}^n$ is a sharper form of homeomorphism, involving the way in which they live in the latter space. We say that two subsets of $\mathbf{R}^n$ are embedded there **in the same way** (up to topological equivalence) if there is a homeomorphism $h$ of $n$-space onto itself which maps one of the spaces onto the other. Thus the two sets in the plane above are homeomorphic but are not embedded in $\mathbf{R}^2$ in the same way. They *are*, however, embedded in $\mathbf{R}^3$ in the same way.

A more spectacular example of this phenomenon is the so-called "horned sphere" of Alexander which is illustrated in figure 6. It is a subset of three-space which is homeomorphic to $\mathbf{S}^2$ but is not embedded in $\mathbf{R}^3$ in the same way as the latter. (This should be clear from the diagram—the loop around the horned sphere cannot be smoothly extricated from the sphere. There is no such loop which has the same property with respect to the standard sphere in space).

Another example of this phenomenon which has an attractive intuitive basis is the topic of knots. A mathematical knot is defined to be a subset of space which is homeomorphic to $\mathbf{S}^1$. Two such knots are pictured in figure 7 (of course, in everyday parlance, the first one would be regarded as being unknotted). The second is the simplest of all knots, called the trefoil. The reader will note that the difference between them is that they are not embedded in $\mathbf{R}^3$ in the same way. Hence a knot (more precisely, a *mathematical* knot) is an equivalence class of copies of $\mathbf{S}^1$, the equivalence relationship being that two such copies are embedded into space in the same way. There exists a rich mathematical theory of such knots.

In this chapter, we will bring an elementary, non-rigorous approach to some of the more appealing results of topology. We shall consider special cases (for the plane) of results which are, in fact, valid in higher dimensions. (These higher dimensional results will be proved rigorously in chapter III). Our approach will be based on the concept of the winding number of suitable mappings in two dimensions. Although the geometrical meaning of this concept and the properties of it which we shall require for our development are intuitively obvious, it will not be possible to give a completely rigorous treatment at this stage. Our approach will thus be rather informal.

We shall be concerned with continuous mappings from the circle into the punctured plane i.e. the set $\mathbf{R}^2 \setminus \{0\}$. In view of the importance of the latter space in what follows, we introduce the notation **PP** for it. Functions of the type mentioned above can be regarded as closed loops around the origin (cf. figure 8). We can assign to such functions a number — the so-called **winding number** — which describes how often the curve winds around the origin. For example, the above curves have winding numbers $0, 1, -1, -2$ respectively (we are employing the usual convention that the anti-clockwise direction is regarded as positive).

Before sketching how this concept can be given a rigorous topological definition, we bring some preliminaries. Firstly, we note that in the above discussion we can replace the circle by any space which is homeomorphic to it. Thus it will often be convenient to consider the winding number of mappings from $\partial\mathbf{I}^2$, the boundary of $\mathbf{I}^2$ (i.e. the square with vertices $(0,0),(1,0),(1,1),(0,1)$), into **PP**. Now consider the two subsets

$$U_1 = \{x = (\xi_1, \xi_2) : \xi_1 > 0 \text{ or } \xi_2 \neq 0\}$$

and

$$U_2 = \{x : \xi_1 < 0 \text{ or } \xi_2 \neq 0\}$$

of the punctured plane. It is clear that they are open and that their union is all of **PP**. The point of introducing them is that although there is no continuous function on **PP** which measures the angle which a given vector makes with the $x$-axis, there *are* such functions on $U_1$ and $U_2$. More precisely, there are continuous functions

$$\theta_1 : U_1 \to ]-\pi, \pi[ \text{ and } \theta_2 : U_2 \to ]0, 2\pi[$$

so that

$$x = |x|(\cos\theta_1(x), \sin\theta_1(x)) \quad (x \in U_1)$$

resp.

$$x = |x|(\cos\theta_2(x), \sin\theta_2(x)) \quad (x \in U_2).$$

Now suppose that $f$ is a continuous function from $\mathbf{S}^1$ into $\mathbf{PP}$. Then we claim that there is a finite sequence of points $s_1, \dots, s_n$ which traverses the unit circle in the anti-clockwise direction as in figure 9 (i.e. the sequence does not double back) so that $s_0 = s_n$ and on each circular segment from $s_i$ to $s_{i+1}$ the range of $f$ lies either in $U_1$ or in $U_2$. Intuitively, this fact is rather obvious in view of the continuity of $f$ but in order to give an exact proof we shall require the concept of compactness which will be introduced and studied in I.7.

We now define the **winding number** of $f$ by means of the following formula

$$w(f) = \frac{1}{2\pi}\sum_{i=0}^{n-1} \theta_?(f(s_{i+1}) - \theta_?(f(s_i))$$

where $\theta_?$ is $\theta_1$ or $\theta_2$ according as the range of $\phi$ on the arc-segment between $s_i$ and $s_{i+1}$ is in $U_1$ or in $U_2$. Since the jumps in the function obtained at the endpoints of the segments which arise from exchanging the two $\theta$ functions are always multiples of $2\pi$, it is clear that $w(f)$ is a whole number.

We shall require some simple properties of the winding number. Firstly, it is independent of the choice of subdivision. This is proved by showing that if we have two such subdivisions, then we can combine the endpoints to form one which is finer than both as in figure 10. It is not difficult to see that the winding numbers defined by the original subdivisions are the same as those defined by the common refinement.

For the next property, it will at first be more convenient to consider functions from $\partial\mathbf{I}^2$ into $\mathbf{PP}$. We claim that if the function $f$ has a continuous extension to a function from $\mathbf{I}^2$ into $\mathbf{PP}$, then its winding number is zero. The proof of this uses a very common type of argument. Once again, it involves an application of compactness and so we shall merely sketch it.

Firstly, we dissect the unit square into four smaller ones as in figure 11. The restriction of $f$ to the boundaries of these squares (which are also homeomorphic to $\partial\mathbf{S}^1$) defines four mappings $f_{00}, f_{01}, f_{10}$ and $f_{11}$ into $\mathbf{PP}$ and these can also be assigned winding numbers. It is clear that we have the relationship:

$$w(f) = w(f_{00}) + w(f_{01}) + w(f_{10}) + w(f_{11})$$

(for if we consider the definition of the various winding numbers of the four submappings, then we can arrange for the subdivisions which define them to match up at the common borders as in figure 12. Then if we sum the expressions for the winding numbers, the terms which correspond to the common segments cancel and we are left with the winding number of $f$).

We now repeat this process by dissecting the four smaller squares into four new ones. We thus obtain 16 mappings whose winding numbers sum to that of $f$. We can continue this process indefinitely. At the $n$-th stage, we have $4^n$ mappings whose winding numbers sum to that of $f$. Each of these mappings is the restriction of the continuous function to a very tiny square and once again it is intuitively obvious that by choosing $n$ large enough we can arrange for each of these submappings to remain in either $U_1$ or $U_2$, which implies that they have winding number zero. Hence that of $f$ is zero.

Of course, this result means that if we have a continuous function $f$ from $\mathbf{S}^1$ into $\mathbf{PP}$ which has a continuous extension to a mapping from $\mathbf{B}^2$ into $\mathbf{PP}$, then its winding number vanishes.

We shall now use the winding number to prove some results on the topology of the plane:

**Proposition 1** *There is no continuous mapping from $\mathbf{B}^2$ into $\mathbf{S}^1$ which is such that $r(x) = x$ for $x \in \mathbf{S}^1$.*

This follows immediately from the above considerations, since the restriction of $r$ to $\mathbf{S}^1$ is then the identity which, of course, has winding number 1 in contradiction to the above statement since the condition on $r$ means precisely that it is a continuation of the identity function.

At this point, we remark that a continuous mapping $r$ from a subset $X$ of $n$-space onto a subset $X_0$ of $X$ which is such that $r(x) = x$ whenever $x$ is in $X_0$ is called a **retraction** from $X$ onto $X_0$. $X_0$ is then called a **retract** of $X$. This is an important concept in topology since there is then a close relationship between the topological properties of $X$ and those of $X_0$. In this context, the above result means that the circle is *not* a retract of $\mathbf{B}^2$.

The next result is one of *the* famous theorems of topology. It is often called the "crumpled handkerchief" theorem since it can be paraphrased as follows: if one lays a handkerchief flat on a table, picks it up, crumples it and replaces it so that it lies within the region which it previously covered, then there is at least one point which returns to its original position.

The mathematical statement of the theorem is as follows:

**Proposition 2** *(The Brouwer fixed point theorem in two dimensions) If $f$ is a continuous mapping from $\mathbf{B}^2$ into itself, then $f$ has a fixed point i.e. there is a point $x \in \mathbf{B}^2$ with $f(x) = x$.*

PROOF. If $f$ has no fixed points, then figure 13 shows how to construct a retraction $r$ from $\mathbf{B}^2$ onto $\mathbf{S}^1$.

■

The third result concerns the non-existence of certain vector fields on spheres. A **vector field** on $\mathbf{S}^n$ is a continuous mapping $f$ from $\mathbf{S}^n$ into $\mathbf{R}^{n+1}$ so that $f(x)$ is perpendicular to $x$ for each $x$. If we transfer the vector $f(x)$ so that its initial point is at $x$, then we can visualise such a field as a "hairy ball".

We shall be interested in vector fields which do not vanish at any point of the sphere. Figure 14 shows that such vector fields exists on $\mathbf{S}^1$. Our next result shows that such fields do not exist on the sphere. (The general situation is that such a field always exists on spheres of odd dimension and never on those of even dimension).

**Proposition 3** *(the "hairy ball" theorem) Every vector field on the two dimensional sphere has a zero.*

The proof for the two dimensional case is as follows: Suppose that we have a vector field which vanishes nowhere. We consider a point $x_0$ on the sphere. Since the value of the field at $x_0$ is non-zero and since it is continuous there, we can choose a very small circular neighbourhood of the point on which the field is almost constant i.e. looks very much like the picture in figure 15.

We now cut this small disc from the sphere. The remainder is homeomorphic to $\mathbf{B}^2$ so that we can regard the restriction of the vector field to it as a continuous mapping from $\mathbf{B}^2$ into $\mathbf{PP}$. Hence its restriction to the boundary has winding number 0 by the above. However, the following diagram (figure 16) should convince the reader that the winding number is in fact 2.

**The Borsuk-Ulam antipodal theorem** The following result has found its way into popular works on topology as the statement that at any given time there are two antipodal points on the surface of the earth at which the temperature and humidity level coincide. Its mathematical statement is as follows:

**Proposition 4** *Let $f$ and $g$ be continuous functions from $\mathbf{S}^2$ into the real line. Then there is an $x_0$ in $\mathbf{S}^2$ with $f(x_0) = f(-x_0)$ and $g(x_0) = g(-x_0)$.*

PROOF. We consider the continuous function

$$G : x \mapsto (f(x) - f(-x), g(x) - g(-x))$$

from $\mathbf{S}^2$ into $\mathbf{R}^2$. The result to be proved is clearly equivalent to the claim that $G$ has a zero. The proof is by contradiction. We suppose that $G$ has no zeros. Consider then its restriction to the upper hemisphere. The latter is homeomorphic to $\mathbf{B}^2$ and $G$ maps the latter into $\mathbf{PP}$. Hence, by the above, the winding number of its restriction to the equator is zero. We shall deduce a contradiction by using the symmetry of $G$ to show that its winding number cannot be equal to zero. Suppose that we calculate the winding number of $G$ by using a dissection of the

circle which is symmetric as in figure 17. Then it is clear that the angular change on any one arc is the same as that on the opposite one. Hence the winding number is twice the angular change between $s_0$ and $s_n$ (divided by $2\pi$). But the latter must be an odd multiple of $\pi$ and so the winding number of $G$ cannot be zero.

■

# 3 Metric spaces:

We now abstract from the above concrete situation by using a standard method—we use the essential properties of the metric used above to define the concept of an abstract metric spaces. It is then a routine exercise to extend many of the above definitions and properties to this abstract situation. Since much of these (in the concrete situation) will be familiar from a standard Analysis course, we will not spend much time on motivational remarks.

**Definition:** Let $X$ be a set. A **metric** on $X$ is a mapping $d : X \times X \to \mathbf{R}^+$ so that

1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;
2) $d(x, y) = d(y, x)$;
3) $d(x, z) \leq d(x, y) + d(y, z)$ all these for $x, y, z \in X$.

We can then use the metric to define all of the concepts which we have discussed in the context of subsets of $\mathbf{R}^m$. Before doing so, we consider some further examples.

I. The discrete metric: If $X$ is a set, we define a metric $d_D$ on $X$ by setting $d_D(x, y)$ to be equal to 1 if $x$ and $y$ are distinct and to be equal to 0 otherwise.

II More generally, if we have a family $\{(X_\alpha, d_\alpha)\}$ of metric spaces, then we can define a metric $d$ on their disjoint union $X = \amalg X_\alpha$, by putting $d(x, y) = 0$ if $x$ and $y$ come from different $X$'s. If both come from the same one, say $X_\alpha$, we define $d(x, y)$ to be $d_\alpha(x, y)$. The example of I is the case where each $X_\alpha$ is a singleton.

III. The indiscrete semi-metric: If $X$ is a set, we consider the mapping $d_I$ which associates to each pair $(x, y)$ the value zero. This is not a metric since it fails the second part of condition 1) above. However, it does satisfy the other conditions. Such a mapping is called a **semi-metric.** It can be used to define continuity and convergence in exactly the same way as for metrics. However, we shall see later that convergence can then have some rather odd properties.

Using a metric, we can define the following concepts which will provide the basis for our treatment of topological spaces.

A subset $N$ of $X$ is said to be a **neighbourhood** of a point $x$ if there is an $\epsilon > 0$ so that $U(x, \epsilon) \subset N$ where $U(x, \epsilon)$ denotes the set $\{y : d(x, y) < \epsilon\}$ (the **open ball with centre** $x$ **and radius** $\epsilon$).

Note that if $N_1$ and $N_2$ are neighbourhoods of $x$, then so is their intersection $N_1 \cap N_2$ (since if $U(x, \epsilon_1) \subset N_1$ and $U(x, \epsilon_2) \subset N_2$, then $U(x, \epsilon) \subset N_1 \cap N_2$ where $\epsilon = \min(\epsilon_1, \epsilon_2)$). It is trivial that if $N$ is a neighbourhood of $x$, then so is any superset of $N$. Further if we denote by $\mathcal{N}(\S)$ the family of neighbourhoods of $x$, then the intersection of the sets of $\mathcal{N}(\S)$ is the one-point set $\{x\}$. (For if $y \neq x$, then $\epsilon > 0$ where $\epsilon = d(x, y)$ and $U(x, \frac{\epsilon}{2})$ is a neighbourhood of $x$ which does not contain $y$).

A subset $U$ of a metric space $X$ is said to be **open** if it is a neighbourhood of each of its points i.e. for each $x \in U$ there is an $\epsilon > 0$ so that $U(x, \epsilon) \subset U$. This is equivalent to the fact that $U$ is a union of open balls. It is clear that the empty set $\emptyset$ and $X$ are open. Also any union resp. finite intersection of open sets is open. **Continuity of mappings:** Let $(X, d)$ and $(Y, d_1)$ be metric spaces, $f : X \to Y$ a mapping between them. $f$ is **continuous** at $x_0 \in X$ if for each positive $\epsilon$ there is a $\delta > 0$ so that $d(x, y) < \delta$ implies that $d_1(f(x), f(y)) < \epsilon$. It is **continuous on** $X$ if it is continuous at each point of $X$. This can be expressed in terms of neighbourhoods resp. open sets as follows. $f$ is **continuous at** $x_0$ if and only if whenever $N$ is a neighbourhood of $f(x_0)$, $f^{-1}(N)$ is a neighbourhood of $x_0$. (for the original condition means that $f^{-1}(U(f(x_0), \epsilon) \supset U(x_0, \delta))$. Similarly $f$ is **continuous on** $X$ if and only if $f^{-1}(U)$ is open for each $U$ open in $Y$.

A bijection $f : X \to Y$ is a **homeomorphism** if both $f$ and $f^{-1}$ are continuous.

As in the case of functions on $\mathbf{R}$, we can use the metric to introduce two stronger notions of continuity — those of uniform continuity and Lipschitz continuity. A function $f : X \to Y$ as above is **uniformly continuous** if for each positive $\epsilon$ there is a positive $\delta$ so that $d_1(f(x), f(y)) < \epsilon$ whenever $d(x, y) < \delta$. $f$ is **Lipschitz continuous** if there is a positive $K$ so that we have the estimate $d_1(f(x), f(y)) \leq K d(x, y)$ for $x, y$ in $X$. It is clear that Lipschitz continuity implies continuity which in turn implies continuity. Of course, none of these implications is reversible.

Such functions enjoy the expected stability properties. Thus compositions of continuous (uniformly continuous resp. Lipschitz continuous functions) have the same properties. Care is required, however, with products. Thus the product $f.g$ of two Lipschitz continuous functions with values in $\mathbf{R}$ need not be Lipschitz (as the example where both $f$ and $g$ are the identity mapping on the real line shows). However, the product of two bounded, Lipschitz continuous functions is Lipschitz continuous, as the reader can verify. Similar remarks apply to uniformly continuous functions.

## 3.1 Convergence of sequences:

A sequence $(x_n)$ in a metric space $(X, d)$ converges to a point $x$ in $X$ if $d(x_n, x) \to 0$. In terms of neighbourhoods, this can be reformulated as follows: $x_n \to x$ if and only if for each $N \in \mathcal{N}(\S)$ there is a $K \in \mathbf{N}$ so that $x_n \in N$ whenever $n \geq K$. Note that a mapping $f$ from $X$ into $Y$ is then continuous if and only if $f(x_n) \to f(x)$ in $Y$ whenever $x_n \to x$ in $X$.

The **closure** of a set $A$ is the set $\bar{A}$ consisting of those $x$ which have the following property: for each positive $\epsilon$, there is a $y \in A$ with $d(x, y) < \epsilon$ (i.e. for each $\epsilon > 0$, $U(x, \epsilon) \cap A \neq \emptyset$). We say that $A$ is **closed** if $A = \bar{A}$. This can be reformulated as follows: for each $y$ not in $A$ there is a $\delta > 0$ so that $U(y, \delta) \subset X \setminus A$ i.e. $X \setminus A$ is open. Notice that $x$ is in the closure of $A$ if and only if there is a sequence $(x_n)$ in $A$ which converges to $x$ (take $x_n$ to be an element of $A \cap U(x, \frac{1}{n})$). Hence $A$ is closed if and only if it contains the limits of converging sequences with terms in $A$. In the metric space $(X, d_D)$, a sequence converges to $x$ if and only if there is an $N \in \mathbf{N}$ so that $x_n = x$ for $n > N$. (We say then the sequence is **eventually constant**). In a space with the indiscrete semi-metric, *every* sequence converges to every point. (This shows, in particular, that in the absence of the definiteness condition on the metric, limits of sequences need not be unique). Every function from a set $X$ with the discrete metric into a second metric space is continuous. On the other hand, every function from a metric space into a set with the indiscrete semi-metric is continuous.

A subset $A$ of a metric space $X$ is **dense** in $X$ if its closure is equal to $X$. This means that every element of the latter space is the limit of a sequence in $A$. The classical example is $\mathbf{Q}$, the set of rational numbers, which is dense in the real line.

We continue with some more pathological examples of metric spaces:

IV. **The post office metric:** This is a metric on the plane. The distance between two points $x$ and $y$ is defined to be the sum $|x| + |y|$ of the lengths of the vectors if $x$ and $y$ are distinct. Otherwise it is defined to be zero. (The reader should imagine the origin as an operating centre through which every signal from $x$ to $y$ has to pass). One can check that this is a metric. In fact, it is probably easier to verify the following more general fact from which it

follows immediately. If we define the distance between two points $P$ and $Q$ in the plane to be the shortest length of a path from $P$ to $Q$ under suitable restraints, then this will be a metric provided the latter satisfy some natural conditions such as the following:
a) the constant path from $P$ to $P$ satisfies the restraints;
b) if a suitable path from $P$ to $Q$ satisfies the restraints, then the same path taken backwards is a suitable one from $Q$ to $P$;

c) if we have a suitable path from $P$ to $Q$ resp. from $Q$ to $R$, then we can join them together to form a suitable path from $P$ to $R$. In the above case, a path from $P$ to $Q$ is defined to be suitable if it goes through the origin (Except in the trivial case where the endpoints coincide — then there is no restraint).

Using this general fact it is easy to describe two further examples — the taxi-driver's metric and the Washington DC metric.

**Metrics on $\mathbf{R}^n$:** We have already considered the euclidean metric on $n$-space. Two further useful metrics are
VI. $d_1$ where

$$d_1(x, y) = \sum_{i=1}^{n} |\xi_i - \eta_i|.$$

VII. $d_\infty$ where

$$d_\infty(x, y) = \max |\xi_i - \eta_i|.$$

Note that we have the following inequalities:

$$d_\infty \leq d_2 \leq d_1 \leq n d_\infty.$$

From this it follows immediately that each of these metrics induces the same notion of convergence.

We now consider some infinite dimensional spaces.
VIII. $\ell^\infty$. This is the space of bounded sequences in $\mathbf{R}^\mathbf{N}$ i.e.

$$\ell^\infty = \{x = (\xi_n) : \sup |\xi_n| < \infty\}.$$

On this space we define the **supremum metric**

$$d_\infty = \sup\{|\xi_n - \eta_n| : n \in \mathbf{N}\}.$$

IX. $\ell^2$ is the space of sequences $x = (\xi_n)$ which are **square summable** i.e. such that $\sum |\xi_n|^2 < \infty$. The metric $d_2$ on this space is defined by the formula $d_2(x, y) = \sqrt{\sum |\xi_i - \eta_i|^2}$. X. $\ell^1$ is the space of **absolutely**

**summable** sequences i.e. those $x \in \mathbf{R}^{\mathbf{N}}$ with $\sum |\xi_n| < \infty$. On this space we use the metric

$$d_1(x, y) = \sum |\xi_n - \eta_n|.$$

XI. $C([0,1])$ denotes the space of continuous, real-valued functions on the unit interval. On this space we define the following three metrics:

the supremum metric: $d_\infty(x, y) = \sup_t |x(t) - y(t)|$;

the $L^2$-metric: $d_2(x, y) = \sqrt{\int_0^1 |x(t) - y(t)|^2 dt}$;

the $L^1$-metric: $d_1(x, y) = \int_0^1 |x(t) - y(t)| dt$.

We omit the routine and unexciting proofs that the above really are metrics. In contrast to the finite dimensional case, these three metrics define completely different notions of convergence in the space $C([0,1])$.

We now introduce two further, highly important examples of metric spaces:

XII. The **Hilbert cube** i.e. the subset

$$\{x = (\xi_n) \in \ell^2 : |\xi_n| \leq \frac{1}{n}\}$$

of $\ell^2$ — regarded as a metric space with the metric $d_2$. (This is the natural infinite dimensional analogue of the hypercube in $\mathbf{R}^n$).

XIII. The **Cantor set**: Recall the classical definition of the Cantor set via missing thirds. We consider the sequence $(F_n)$ of closed subsets of the unit intervals as in Figure 1 (i.e. those obtained by removing successively the middle thirds of the intervals). The intersection $\bigcap F_n$ is a closed subset of the interval and, in particular, a metric space. It consists of those points in the unit interval which have a tryadic representation of the form $\sum_{n=1}^\infty a_n 3^{-n}$ where the coefficients $a_n$ are either 0 or 2. (For we remove those tryadic numbers whose $n$-th coefficient is 1 at the $n$-th stage of the construction). Because of its importance we denote this topological space by **Can** Note that at the $k$-th level of the construction, we split the interval into $2^k$ sub-intervals. If we denote these by $A_1^k, \dots, A_{2^k}^k$, then $F_k = \bigcup_{i=1}^{2^k} A_i^k$ and the Cantor set is

$$\mathbf{Can} = \bigcap_{k=1}^{\infty} \bigcup_{i=1}^{2^k} A_i^k.$$

Later we shall encounter a less geometric description of the Cantor set, which, however, is often more convenient for analysing its properties.

We now turn to some aspects of the theory of metric spaces which use the metric explicitly. We begin with *completeness*. This is a generalisation of the distinguishing characteristic of the real numbers in contrast to the rational ones — the existence of limits for sequences which *should* converge. This is usually an essential property in applications to analysis.

**Definition:** A sequence $(x_n)$ in a metric space is **Cauchy** if the following condition is satisfied: for each positive $\epsilon$ there is an $N \in \mathbf{N}$ so that $d(x_m, x_n) < \epsilon$ for $m, n > N$. The metric space $(X, d)$ is **complete** if each Cauchy sequence converges there.

Before considering concrete examples, we note that the property of a sequence being Cauchy is not a topological one i.e. it is not preserved under homeomorphisms. For example, consider the mapping $f : n \mapsto \frac{1}{n}$ from $\mathbf{N}$ onto $D = \{\frac{1}{n} : n \in \mathbf{N}\}$. We regard both of these sets as metric spaces (with the metric induced from $\mathbf{R}$). Then $f$ is a homeomorphism. However the sequence $(\frac{1}{n})$ is Cauchy in $D$ but its image in $\mathbf{N}$ under the continuous mapping $f^{-1}$ is not Cauchy there.

**Examples of complete spaces:** I. Any set $X$, when provided with the discrete metric, is complete for the trivial reason that any Cauchy sequence therein is eventually constant and so convergent.
II. The real line is complete. This is a fundamental property which will be familiar from a course on elementary analysis.
III. It follows easily from the completeness of the real line that the higher dimensional spaces $\mathbf{R}^p$ are complete (with respect to any of the three metrics $d_1, d_2, d_\infty$ introduced above). For a sequence in $\mathbf{R}^p$ is Cauchy resp. convergent if and only if the same is true of each of its components.
IV. Any closed subset of $\mathbf{R}^p$ is complete. For if $(x_n)$ is a Cauchy sequence in such a set $B$, then this sequence has a limit in $\mathbf{R}_p$ which, by the closedness, must lie in $B$.
V. An example of a space which is not complete is the open interval $]0, 1[$ in the line. This fails to be complete since the Cauchy sequence $(\frac{1}{n})$ there is not convergent. (Note that this space is homeomorphic to the real line which *is* complete).

The above examples suggest the following simple facts: a) if $X$ is a complete metric space, then each closed subset thereof is complete under the induced metric; b) If a subset $X_1$ of a metric space $X$ is complete for the induced metric, then $X_1$ is closed in $X$.

We now consider the completeness of some infinite dimensional spaces: The space $\ell^\infty$ is complete, as are $\ell^1$ and $\ell^2$. More generally, if $S$ is an arbitrary space, we can define a natural generalisation $\ell^\infty(S)$ of the former space as follows. This space is also complete. The space $C([0, 1])$. This is complete with respect to the supremum norm since it is a closed subspace of the complete space $\ell^\infty(S)$ (this is just a restatement of the well-known fact that the uniform limit of continuous functions is continuous). We remark here that the space $C([0, 1])$ is *not* complete for the metrics $d_1$ and $d_2$.

Recall that a function $f : X \to X_1$ where $(X, d)$ and $(X_1, d_1)$ are metric spaces is **Lipschitz continuous** if there is a $K \geq 0$ so that

$$d_1(f(x), f(y)) \leq K d(x, y)$$

for each $x, y \in X$. The smallest such $K$ is then called the **Lipschitz constant** of $f$. If the latter is strictly smaller than 1, then $f$ is called a **contraction**. Lipschitz continuous functions enjoy an important property which is described in the following Proposition:

**Proposition 5** *Let $(X, d)$ and $(X_1, d_1)$ be metric spaces where $X_1$ is complete and suppose that $X_0$ is a dense subset of $X$. Suppose that $f : X_0 \to X_1$ is a Lipschitz-continuous function. Then it can be extended to a Lipschitz-continuous function $\tilde{f}$ from $X$ into $X_1$. Furthermore, this extension is unique and has the same Lipschitz constant as $f$.*

PROOF. The proof is based on the simple observation that the image of a Cauchy sequence under a Lipschitz mapping is also Cauchy. Suppose that $x$ is a point of $X$. It is then the limit of a sequence $(x_n)$ in $X_0$. The latter is then Cauchy in $X_0$ and hence the image sequence $(f(x_n))$ is Cauchy in $X_1$. By the completeness of the latter there is a $y$ in $X_1$ to which $(f(x_n))$ converges. The remainder of the proof consists of the verification of the following facts:
1) $y$ depends only on $x$ (and not on the choice of the sequence $(x_n)$). For if $(z_n)$ is a second sequence in $X_0$ which tends to $x$, then $d(x_n, z_n)$ tends to zero. It follows from the Lipschitz condition that $d(f(x_n), f(z_n))$ also tends to zero. Hence $(f(x_n))$ and $(f(z_n))$ have the same limit.
2) The function $\tilde{f}$ which maps $x$ to the above $z$ if $x$ is not in $X_0$ (and agrees with $f$ on $X_0$) is Lipschitz. For if $K$ is the Lipschitz constant of $f$ and $(x_n)$ resp. $(y_n)$ are sequences in $X_0$ which tend to $x$ resp. $y$, then the inequality

$$d(f(x_n), f(y_n)) \leq K d(x_n, y_n)$$

tends in the limit to the desired inequality

$$d(f(x), f(y)) \leq K d(x, y).$$

■

**Remark:** In this proof we have tacitly used the fact that if $x_n \to x$ and $y_n \to y$, then $d(x_n, y_n) \to d(x, y)$. The reader should check that this holds.

In the light of the above proposition, it is an important fact that every metric space can be completed in the sense that it can be imbedded as a dense subset of a complete space. We prove this by means of the following Lemma. Before stating it, recall that an **isometry** between two metric spaces $(X_1, d_1)$ and $(X_2, d_2)$ is a bijection $f$ so that $d_2(f(x), f(y)) = d_1(x, y)$ for each pair $x, y$ in $X_1$. If such an $f$ exists, $X_1$ and $X_2$ are said to be **isometric** or **isometrically isomorphic.**

**Proposition 6** *Let $(X, d)$ be a metric space. Then $X$ is isometric to a subspace of the metric space $\ell^\infty(X)$.*

PROOF. Choose a fixed $x_0$ in $X$ (we are assuming that $X$ is non-empty). If $x \in X$, let $f_x$ be the function

$$f_x : y \mapsto (d(x, y) - d(x, x_0)).$$

Then $|f_x(y)| \leq d(x, x_0)$ for each $y$ and so $f_x \in \ell^\infty(X)$. The mapping $x \mapsto f_x$ is the required isometry.

■

**Proposition 7** *Let $(X, d)$ be a metric space. Then $X$ is isometric to a dense subset of a complete metric space.*

PROOF. The closure of the image of $X$ in $\ell^\infty(X)$ as in the above Lemma is the required space.

■

We note the following properties of this embedding which follow immediately from the above results:
I. Consider the embedding $j$ from $X$ into the above complete space which we denote by $Y$. Then if $f$ is a Lipschitz continuous mapping from $X$ into a second complete space $Z$, there is a unique Lipschitz mapping $\tilde{f}$ from $Y$ into $Z$ so that $f = \tilde{f} \circ j$. (This is just a rather abstract and pedantic way of saying that $\tilde{f}$ is an extension of $f$ if we regard $X$ as a subspace of $Y$ via the isometry $j$).
II. Any two complete spaces which contain $X$ as dense subspaces in the above way are isometrically isomorphic. More precisely, if $Y_1$ is a second space into which $X$ is isometrically embedded as a dense subspace via a mapping $j_1$, then there is an isometric isomorphism $I$ from $Y$ onto $Y_1$ so that $I \circ j = j_1$. The proof is simple. The mapping $j_1$ from $X$ into $Y_1$ extends to a mapping $I$ from $Y$ into $Y_1$ by the above extension property. $I$ has the required properties.

This result implies that the space $Y$ constructed above is essentially unique and so independent of the method of construction. Hence we can simply refer to it as *the* completion and denote it by $\hat{X}$.

**Products of metric spaces:** If $(X_1, d_1), \dots, (X_n, d_n)$ are metric space, there are various natural ways of defining a metric on their product $X_1 \times \dots \times X_n$. As in the case of $\mathbf{R}^n$, there are three particularly simple ones:

$$\begin{aligned} x, y) &\mapsto \sum d_i(x_i, y_i) \\ (x, y) &\mapsto \sqrt{\sum d_i(x_i, y_i)^2} \\ (x, y) &\mapsto \max(d_i(x_i, y_i)) \end{aligned}$$

where $x = (x_i), y = (y_i)$.

Since these metrics are equivalent in the sense that they define the same notions of convergent resp. Cauchy sequences, we are free to use any one of them — it will be convenient to use the third one.

In the case of countable products $\prod_{n \in \mathbf{N}} X_n$ of metric spaces, we shall assume at first that each metric is bounded by 1 i.e. that $d_n(x, y) \leq 1$ for each pair $x, y$ in $X_n$. Then we define a metric $\tilde{d}$ on the Cartesian product as follows: if $x = (x_n)$ and $y = (y_n)$, then

$$\tilde{d}(x, y) = \sum \frac{1}{2^n} d_n(x_n, y_n).$$

The properties of the product which we shall require are as follows:

1. a sequence $x_k = (x_n^k)$ in the product is Cauchy if and only if for each $n$ the sequence $(x_n^k)_{k=1}^\infty$ in $X_n$ is Cauchy.

2. a sequence $(x_k)$ as above converges to $x = (x_n)$ if and only if $x_n^k \to x_n$ in $X_n$ for each $n$.

3. $\prod X_n$ is complete if each $X_n$ is.

PROOF. We prove the first statement. The proof of 2) is almost identical and 3) follows from 1) and 2). Firstly, the natural projection from $\Pi X_n$ onto $X_m$ is clearly Lipschitz (with constant $2^m$) for each $m$. This implies the necessity of the above condition.

■

Now suppose that each sequence $(x_n^k)$ (as $k$ varies) is Cauchy. We shall show that $(x^k)$ is Cauchy. Given a positive $\epsilon$, choose $N$ so that $2^{-N} < \frac{\epsilon}{3}$ and a $K \in \mathbf{N}$ so that if $k, l \geq K$, then $d_n(x_n^k, x_n^l) \leq \frac{\epsilon}{3}$ for $n = 1, \dots, N$. Then $d(x^k, x^l) \leq \epsilon$ if $k, l \geq K$.

We remark now that if $(X, d)$ is an arbitrary metric space, then $e : (x, y) \mapsto \min(1, d(x, y))$ is a metric on $X$ which has the same Cauchy sequences resp. convergent sequences as $d$. Hence we can define a countable product of arbitrary metric spaces by replacing their metrics by ones which are bounded as above and then using this construction.

We shall be interested in the following examples of products of the form $X^{\mathbf{N}}$ where $X$ is a fixed metric space. This is the special case of a product where each of the components is the same.

I. $\mathbf{I}^{\mathbf{N}}$ — a product of countably many copies of the unit interval. This space is essentially the Hilbert cube. More precisely, the mapping $(\xi_n) \mapsto (\frac{1}{n}\xi_n)$ is a homeomorphism from $\mathbf{I}^{\mathbf{N}}$ onto the Hilbert cube.

II. $\mathbf{N}^{\mathbf{N}}$ — the product of countably many copies of the natural numbers. (the latter regarded as a metric space with the discrete metric). Strangely enough, this space is homeomorphic to the irrational numbers and, from a topological point of view, is their most convenient representation. The exact proof of this requires some elementary number theory. Since we shall not require this identification directly, we shall be content with the remark that each sequence $(\xi_n, \xi_2, \dots)$ in $\mathbf{N}^{\mathbf{N}}$ defines an irrational number by means of a continued fraction.

This defines a mapping from $\mathbf{N}^{\mathbf{N}}$ into the irrationals which is, in fact, a homeomorphism onto. This fact can also be proved without recourse to number theory as follows. We list the rationals as a sequence $r_1, r_2, \dots$ and construct a countable family $\mathcal{I}_\backslash$ of partitions of the irrationals into countable collections of intervals so that the length of each interval in $\mathcal{I}_\backslash$ is at most $2^{-n}$ and so that $r_n$ is the endpoint of one of the intervals of $\mathcal{I}_\backslash$ but not of any previous partition. Then each irrational number $x$ is in precisely one intersection of the form

$$I^1_{n_1} \cap I^2_{n_2} \cap I^3_{n_3} \cap \dots$$

where $I^1_{n_1}$ denotes the $n_1$-th interval in $\mathcal{I}_\infty$ etc. The mapping

$$x \mapsto (n_1, n_2, \dots)$$

is the required homeomorphism from the irrationals onto $\mathbf{N}^{\mathbf{N}}$.

III. The space $\{0,1\}^{\mathbf{N}}$, i.e. the infinite product of countably many copies of the discrete space $\{0,1\}$. This space is homeomorphic to the Cantor set. In fact, if $x = (\xi_n) \in \{0,1\}^{\mathbf{N}}$, then $f(x) = \sum_{n=1}^\infty \frac{2\xi_n}{3^n}$ is an element of $[0,1]$ — indeed, of the Cantor set. Once again, the mapping $x \mapsto f(x)$ is a homeomorphism from the product space $\{0,1\}^{\mathbf{N}}$ onto **Can**.

For future reference, we note the following fact which follows immediately from this description of the Cantor set. There is a continuous surjection from the Cantor set onto the unit interval **I**. Indeed

$$\{0,1\}^{\mathbf{N}} \in (\xi_n) \mapsto \sum_{n=1}^\infty \frac{\xi_n}{2^n}$$

is such a mapping.

Now this implies that there is a continuous surjection from $\mathbf{Can}^{\mathbf{N}}$ onto $\mathbf{I}^{\mathbf{N}}$. But $\mathbf{I}^{\mathbf{N}}$ is the Hilbert cube and $\mathbf{Can}^{\mathbf{N}}$ is $(\{0,1\}^{\mathbf{N}})^{\mathbf{N}}$ and it should come as no surprise that the latter space is once again the Cantor set. In fact, we have:

**Proposition 8** *If $X$ is a metric space, then $X^{\mathbf{N}}$ and $(X^{\mathbf{N}})^{\mathbf{N}}$ are homeomorphic.*

PROOF. An element of the second set is a sequence of sequences in X and so can be regarded as a double sequence of such elements i.e. $X = (X_n)$ where $X_n = (x_m^n)$. Now the latter double sequence can be straightened out into a single sequence e.g. as the sequence $(x_1^1, x_1^2, x_2^2, x_2^1, x_3^1, \dots)$. This defines a homeomorphism from $(X^{\mathbf{N}})^{\mathbf{N}}$ onto $X^{\mathbf{N}}$.

■

Putting this together, we see that there is a continuous mapping from **Can** *onto* the Hilbert cube.

We now bring some results on metric spaces in which the completeness plays an essential role. Firstly, if $A$ is a subset of a metric space, then its **diameter** is the number

$$\operatorname{diam}(A) = \sup\{d(x,y) : x, y \in A\}.$$

(This can of course be infinite).

**Proposition 9** *Let $(A_n)$ be a sequence of non–empty, closed subsets of a complete metric space $(X,d)$ so that $A_{n+1} \subset A_n$ for each $n$ and oxdiam $A_n \to 0$. Then the intersection $\cap A_n$ is non-empty (and in fact consists of a single point).*

PROOF. We choose for each $n$ an $x_n \in A_n$. Then of course $d(x_m, x_n) \leq \operatorname{diam} A_N$ whenever $m, n \geq N$. Hence $(x_n)$ is Cauchy. Let $x$ denote its limit. Then $x_m \in A_n$ for $m \geq n$ and so in the limit $x \in A_n$ since the latter is closed. Since this holds for each $n$, we have that $x$ is in the required intersection. Suppose now that there were a second element $y$ in this intersection, with $y \neq x$. There is an $n \in \mathbf{N}$ with *oxdiam* $A_n \leq d(x,y)$. $y$ cannot be an element of $A_n$ (since $x$ *is* in the latter). This provides a contradiction.

■

The next result is the basis of countless existence theorems in analysis.

**Proposition 10** *Let $(A_n)$ be a sequence of closed subsets of a complete metric space $(X,d)$ whose union is $X$. Then $\bigcup A_n^o$ is dense in $X$.*

PROOF. Suppose, if possible, that the conclusion is not valid. Then there is a non-empty, open set $U_1$ with $U_1 \cap (\bigcup_n A_n) = \emptyset$. Since $U_1$ is not a subset of $A_1$, the difference $U_1 \setminus A_1$ is non-empty (and open) and so contains an open ball $V_1$ of radius $\epsilon_1 > 0$. Since $V_1$ is not a subset of $A_2$, $V_1 \setminus A_2$ contains an open ball $V_2$ of positive radius $\epsilon_2$. Continuing inductively, we obtain a decreasing sequence $(V_n)$ of open balls, whereby $V_n$ has radius $\epsilon_n$. Furthermore, $V_{n+1} \subset V_n \setminus A_n$. We can arrange for the radii $\epsilon_n$ to converge to zero. Then by Cantor's result the intersection of the $V_n$ contains a point $x$. It is clear that $x$ is in none of the $A_n$ which is a contradiction.

■

**Exercise:** In the above proof we committed the crime of applying the result of Cantor to a decreasing sequence of *open* balls. Use the fact that each open ball contains a smaller, closed one to correct the proof.

**Proposition 11** *If $X$ and $(A_n)$ are as above and $X$ is supposed to be non-empty, then there is an $n_0$ so that $A_{n_0}$ has non-empty interior.*

The above result is known as Baire's theorem. Almost the same proof demonstrates the following version of this result:

**Proposition 12** *If $X$ is as above and $(A_n)$ is a sequence of closed subsets so that the interior $(\bigcup A_n)^o$ of their union is non-empty, then there is an $n_0$ so that the interior of $A_{n_0}$ is non-empty.*

The above results can be usefully reformulated by using the following concepts. A subset $A$ of a metric space is **nowhere dense** if the interior of its closure is empty. It is **of first category** if it is expressible as a countable union of nowhere dense sets. It is **of second category** if it is not of first category i.e. if whenever $A$ is expressed as the union $\bigcup A_n$ of a countable family of sets, then at least one $A_n$ fails to be nowhere dense. Thus the theorem of Baire states that each non-empty, complete metric space is of second category.

**Examples:** $\mathbf{Q}$ is a set of first category in $\mathbf{R}$. In fact, any countable subset of a metric space in which singletons are open is of first category. Of course, $\mathbf{Q}$ is not nowhere dense. The Cantor set, regarded as a subset of $[0, 1]$, is a non-trivial example of a nowhere dense set (i.e. it is nowhere dense and *uncountable*).

**Definition:** A subset of a metric space is a $G_\delta$- set if it is an intersection of a sequence of open sets. Before stating the next version of Baire's theorem, we note that the intersection of two dense subsets of a metric space need not be dense – a typical example is the pair $\mathbf{Q}$ and $\mathbf{R}\setminus\mathbf{Q}$ of dense subsets of the line whose intersection is empty. However, the intersection of two *open*, dense subsets is dense (easy exercise). (In fact, the intersection of an open dense subset with *any* dense subset is dense). Our next version of Baire's theorem is a significant strengthening of this result for complete metric spaces:

**Proposition 13** *Let $(X, d)$ be a complete metric space. Then an intersection of a sequence of open, dense subsets is dense (and so a dense $G_\delta$-set).*

PROOF. The proof is based on the simple fact that a subset of a metric space is open and dense if and only if its complement is a closed, nowhere dense set. Hence if each of the sequence $(U_n)$ is open and dense and if $A$ is their intersection, then the complement of $A$ is the union of the complements of the $U_n$. Hence if the latter set contains a non-empty, open set, then so must one of the sets $(X\setminus U_n)$ by Baire's theorem and this would contradict $U_n$'s status as an open dense subset.

∎

**Proposition 14** *If $(G_n)$ is a sequence of dense $G_\delta$-subsets of a complete metric space, then their intersection is also a dense $G_\delta$-subset.*

As an example of an application of the theorem of Baire, we bring the following result on boundedness of sequences of functions (it is the basis of a number of famous results of Banach in functional analysis - notably of the Banach Steinhaus theorem):

**Proposition 15** *Let $M$ be a subset of the space $C(X)$ of continuous real-valued functions on a complete, non-empty metric space $X$ which is pointwise bounded i.e. such that for each $t \in X$ there is a $K > 0$ so that $|x(t)| < K$ for each $x \in M$. Then there is a non-empty, open subset $U$ of $X$ so that $M$ is uniformly bounded on $U$ (i.e. there is an $L > 0$ so that $|x(t)| < L$ for each $x \in M$ and $t \in U$).*

PROOF. This is a simple application of Baire's theorem using the sets

$$A_n = \{t \in X : |x(t)| \leq n \text{ for each } x \in M\}.$$

∎

Another result which uses the completeness of the metric space is the famous fixed point theorem of Banach. We have already met an example of fixed point theorem — namely that of Brouwer which states that each continuous mapping on $B^2$ has a fixed point. In fact, the same result holds in higher dimensions and we shall prove this in a later chapter. The Banach fixed point theorem has an entirely different character. It is true for a much wider class of spaces (in fact, for any complete metric space) but requires a condition on the mapping which is much stronger than continuity. In return, the fixed point is unique and the proof provides a method of finding it, respectively good approximations to it.

**Proposition 16** *Let $(X, d)$ be a non-empty complete metric space and let $f$ be a contraction on $X$. Then $f$ has a fixed point.*

PROOF. We choose any point $x_0 \in X$ and define a sequence $(x_n)$ recursively as follows:

$$x_1 = f(x_0),\ x_2 = f(x_1), \dots, x_{n+1} = f(x_n), \dots$$

We shall show that this is a Cauchy sequence and hence convergent. Its limit $x$ is then a fixed point since

$$x = \lim x_n = \lim x_{n+1} = \lim f(x_n) = f(x).$$

Firstly, we have the estimate:

$$d(x_n, x_{n+1}) = d(f(x_{n-1}), f(x_n)) \leq \lambda d(x_{n-1}, x_n) \leq \cdots \leq \lambda^n d(x_0, x_1)$$

where $\lambda < 1$ is such that

$$d(f(x), f(y)) \leq \lambda d(x, y)$$

for $x, y \in X$. This implies that

$$\begin{aligned} d(x_n, x_{n+p}) &\leq d(x_n, x_{n+1}) + \cdots + d(x_{n+p-1}, x_{n+p}) \\ &\leq \lambda^n d(x_0, x_1) + \cdots + \lambda^{n+p-1} d(x_0, x_1) \\ &\leq \tfrac{\lambda^n}{1-\lambda} d(x_0, x_1) \end{aligned}$$

which tends to zero as $n$ tends to infinity.

■

The reader will have observed that this proof provides a theoretical basis for the simpler iteration methods used for solving equations numerically.

In the above situation, the fixed point is unique as the reader can easily verify.

One consequence of this theorem is the so-called Lipschitz inverse function theorem, which we state for functions on $\mathbf{R}^n$.

**Proposition 17** *Let $A : \mathbf{R}^n \to \mathbf{R}^n$ be an invertible, linear operator and $f : \mathbf{R}^n \to \mathbf{R}^n$ be a Lipschitz-continuous mapping whose Lispchitz constant satisfies the inequality $Lip\, f < |A^{-1}|^{-1}$. Then $A + f$ is a bijection and its inverse $(A + f)^{-1}$ is also Lipschitz-continuous, with constant at most $(|A^{-1}|^{-1} - Lip\,(f))^{-1}$.*

PROOF. The fact that $A+f$ is injective and that its inverse has the required Lipschitz constant follows from the estimate:

$$\begin{aligned} |(A+f)x - (A+f)y| &= |A(x-y) + f(x-y)| \\ &\geq |A(x-y)| - (\mathrm{Lip}\,((f))|x-y| \\ &\geq |A^{-1}|^{-1}|x-y| - (\mathrm{Lip}\,(f))|x-y| \\ &\geq (|A^{-1}|^{-1} - (\mathrm{Lip}\,(f))|x-y|. \end{aligned}$$

The surjectivity follows from the fact that the equation $(A+f)x = y$ can be rewritten as a fixed point equation as follows:

$$x = A^{-1}y - A^{-1} \circ f(x)$$

It follows from the Banach fixed point theorem that this equation has a solution.

■

We have seen above that a closed subset of a complete metric space is itself complete. On the other hand, a subset of a metric space which is itself complete with respect to the induced metric is automatically closed. A rather more subtle question is the following: which subsets of a given complete metric spaces are topologically complete i.e. such that there is an equivalent metric on the subset for which it is complete? Another way of saying this is that the subset be homeomorphic to a complete metric space. For example, the open subset $]-1,1[$ of the space $[-1,1]$ is not complete but it *is* topologically complete, being homeomorphic to $\mathbf{R}$. This is a special case of the following result:

**Proposition 18** *Let $U$ be an open subset of a complete metric space $(X,d)$. Then $U$ is topogically complete.*

PROOF. We define a new metric $\tilde{d}$ on the space $U$ as follows:

$$\tilde{d}(x,y) = d(x,y) + |\phi(x) - \phi(y)|$$

where $\phi(x) = d(x, X\backslash U)^{-1}$. The proof consists of the verification of the following two facts: a) $\tilde{d}$ is equivalent to the original metric $d$ on $U$; b) $(U, \tilde{d})$ is complete. We prove the second statement and leave the (similar) proof of the first to the reader. Let $(x_n)$ be a $\tilde{d}$-Cauchy sequence. Then it is certainly $d$-Cauchy and so $d$-convergent, say to $x$. It suffices to show that $x$ lies in $U$ and $\tilde{d}(x_n, x) \to 0$. If $x$ does not lie in $U$, then $\phi(x_n) \to \infty$. This implies that $\tilde{d}(x_n, x_1) \to \infty$. But this is obviously incompatible with the fact that the sequence is $\tilde{d}$-Cauchy. Since $x \in U$, then $\delta$, the distance of $x$ from $X\backslash U$, is positive. $x_n \to x$ (for $d$) and so there is an $N$ in $\mathbf{N}$ so that $d(x_n, x) < \frac{\delta}{2}$ if $n \geq N$. Hence $\phi(x_n) \leq \frac{2}{\delta}$ for such $n$. If we now piece together the three pieces of information: $d(x_n, x) \to 0$; $d(x_n, X\backslash U)) \to d(x, X\backslash U)$; $\{\phi(x_n)\}$ is bounded, then we can deduce that $\tilde{d}(x_n, x) \to 0$. ■

Recall the definition of a $G_\delta$-subset of a metric space. These are subsets which are describable as countable unions of open sets. Dually, we define the notion of an $F_\sigma$-set which is a countable union of closed sets.

**Examples:** Each closed subset $C$ of a metric space is trivially an $F_\sigma$-set but it is also a $G_\delta$. For the set $U_n = \{x \in X : d(x, C) < \frac{1}{n}\}$ is open (as the union of the open balls $U(y, \frac{1}{n})$ as $y$ ranges through $C$). $C$ is clearly the intersection of the $U_n$. By taking complements, we see that every open set also has both of these properties simultaneously. A non-example is provided by the rationals which form an $F_\sigma$-subset of the reals but not a $G_\delta$. For suppose that $Q$ has a representation as the intersection of a sequence $(U_n)$ of open (dense) subsets. Then the complement of each of the $U_n$ is nowhere dense. But if $(r_n)$ is an enumeration of the rationals, we can write $\mathbf{R} = \bigcup_n (X\backslash U_n) \cup (\bigcup \{r_n\})$ as a countable union of nowhere dense sets which contradicts Baire's theorem.

In a certain sense, $G_\delta$-subsets of metric spaces are the natural domains of definition of continuous mappings. For suppose that $f$ is a continuous mapping from a subset $A$ of a metric space $X$ into a complete metric space $Y$. If $x$ is in the closure of $A$, we define

$$\operatorname{osc}(f; x) = \inf_n \sup \{d(f(y), f(z)) : y, z \in U(x, \frac{1}{n}) \cap A\}.$$

Then if $\operatorname{osc}(f; x) = 0$, $(f(x_n))$ is a Cauchy sequence whenever $(x_n)$ is a sequence in $A$ which converges to $x$. For if $\epsilon$ is a given positive number, there is a $K \in \mathbf{N}$ so that $d(f(y), f(z)) < \epsilon$ for $y, z$ in $U(x, \frac{1}{K}) \cap A$. Hence if we choose $N \in \mathbf{N}$ so that $d(x, x_n) < \frac{1}{K}$ for $n \geq N$, then $d(f(x_m), f(x_n)) < \epsilon$ for $m, n \geq N$. Hence if we set $A_0 = \{x \in A : oxosc\,(f; x) = 0\}$, we can extend $f$ to a function $\tilde{f}$ from $A_0$ into $Y$ by defining $\tilde{f}(x)$ to be the limit of the sequence $(f(x_n))$. One then checks that this extension is well-defined (i.e. the value of $\tilde{f}(x)$ is independent of the choice of the sequence $(x_n)$) and that the function $\tilde{f}$ is continuous, much as in the proof of ????? Now $A_0$ is the intersection of the sequence of subsets $(A_n)$ where $A_n$ is the set of those $x$ in the closure of $A$ for which $\operatorname{osc}(f; x) < \frac{1}{n}$ and the latter is an open subset of $\bar{A}$ as can easily be checked. Hence $A_0$ is a $G_\delta$ in $\bar{A}$ and so in $X$. (For $\bar{A}$ is a $G_\delta$ and it is easily seen that $G_\delta$-subsets of $G_\delta$-sets are themselves $G_\delta$). We have thus proved the following result:

**Proposition 19** *Let $f$ be a continuous mapping from a subset $A$ of a metric space $X$ into a complete metric space $Y$. Then there is a $G_\delta$-set $A_0$ between $A$ and its closure so that $f$ can be extended to a continuous function $\tilde{f}$ from $A_0$ into $Y$.*

From this we can quickly deduce the following result:

**Proposition 20** *Let $X$ and $Y$ be complete metric spaces and let $f : A \to B$ be a homeomorphism between subsets of $X$ and $Y$. Then there are $G_\delta$-subsets $A_1$ and $B_1$ containing $A$ and $B$ respectively (and contained in their closures) so that $f$ extends to a homeomorphism from $A_1$ onto $B_1$.*

PROOF. We can extend $f$ to a continuous $\tilde{f}$ from $A_0$ into $Y$ and $g = f^{-1}$ to a continuous $\tilde{g}$ from $B_0$ into $X$ where $A_0$ and $B_0$ are suitable $G_\delta$'s. Then $A_1 = A_0 \cap \tilde{f}^{-1}(B_0)$ and $B_1 = B_0 \cap \tilde{g}^{-1}(A_0)$ are the required sets.

■

We remark here that in the above proof we used the simple fact that the pre-image of a $G_\delta$-set under a continuous mapping is also a $G_\delta$.

We are now in a position to describe those subsets of a complete metric space which are topologically complete. First we show that $G_\delta$-subsets have this property:

**Proposition 21** *A $G_\delta$-subset of a complete metric space $X$ is topologically complete.*

PROOF. $A$ is the intersection of the sequence $(U_n)$ of open subsets of $X$. We know that each $U_n$ is topologically complete. Hence so is their product $\Pi U_n$ (since products of sequences of complete metric spaces are themselves complete metric). Now $A$ is homeomorphic to a closed subset of the latter product (see the exercise below) and this finishes the proof.

■

**Exercise:** Show that the mapping $x \mapsto (x, x, x, \dots)$ is a homeomorphism from $A$ onto a closed subset of the product of the $U_n$.

We can now complete our characterisation of topologically complete subsets.

**Proposition 22** *A subset $A$ of a complete metric space $(X, d)$ is topologically complete if and only if it is a $G_\delta$.*

PROOF. We have already seen above that this condition is sufficient. We shall now verify the necessity. Suppose that $d_1$ is a metric on $A$ which is equivalent to $d$ there and is such that $(A, d_1)$ is complete. We now apply the above extension result to the identity from $A$ (as a subset of $X$) into the complete metric space $(A, d_1)$. We can extend this to a homeomorphism from a $G_\delta$-set $A_0$ (containing $A$ and contained in $\bar{A}$). But a homeomorphism is a bijection and there is no non-trivial extension of the identity on $A$ to a bijection. Hence $A$ is itself a $G_\delta$.

■

## 3.2 Metric spaces

**Definition:** Let $X$ be a set. A **metric** on $X$ is a mapping $d : X \times X \to \mathbf{R}^+$ so that

1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;

2) $d(x, y) = d(y, x)$;

3) $d(x, z) \leq d(x, y) + d(y, z)$ all these for $x, y, z \in X$.

We can then use the metric to define all of the concepts which we have discussed in the context of subsets of $\mathbf{R}^m$. Before doing so, we consider some further examples.

I. The discrete metric: If $X$ is a set, we define a metric $d_D$ on $X$ by setting $d_D(x, y)$ to be equal to 1 if $x$ and $y$ are distinct and to be equal to 0 otherwise.

II More generally, if we have a family $\{(X_\alpha, d_\alpha)\}$ of metric spaces, then we can define a metric $d$ on their disjoint union $X = \amalg X_\alpha$, by putting $d(x, y) = 0$ if $x$ and $y$ come from different $X$'s. If both come from the same one, say $X_\alpha$, we define $d(x, y)$ to be $d_\alpha(x, y)$. The example of I is the case where each $X_\alpha$ is a singleton.

III. The indiscrete semi-metric: If $X$ is a set, we consider the mapping $d_I$ which associates to each pair $(x, y)$ the value zero. This is not a metric since it fails the second part of condition 1) above. However, it does satisfy the other conditions. Such a mapping is called a **semi-metric.** It can be used to define continuity and convergence in exactly the same way as for metrics.

However, we shall see later that convergence can then have some rather odd properties.

Using a metric, we can define the following concepts which will provide the basis for our treatment of topological spaces. A subset $N$ of $X$ is said to be a **neighbourhood** of a point $x$ if there is an $\epsilon > 0$ so that $U(x, \epsilon) \subset N$ where $U(x, \epsilon)$ denotes the set $\{y : d(x, y) < \epsilon\}$ (the **open ball with centre** $x$ **and radius** $\epsilon$). Note that if $N_1$ and $N_2$ are neighbourhoods of $x$, then so is their intersection $N_1 \cap N_2$ (since if $U(x, \epsilon_1) \subset N_1$ and $U(x, \epsilon_2) \subset N_2$, then $U(x, \epsilon) \subset N_1 \cap N_2$ where $\epsilon = \min(\epsilon_1, \epsilon_2)$). It is trivial that if $N$ is a neighbourhood of $x$, then so is any superset of $N$. Further if we denote by $\mathcal{N}(\S)$ the family of neighbourhoods of $x$, then the intersection of the sets of $\mathcal{N}(\S)$ is the one-point set $\{x\}$. (For if $y \neq x$, then $\epsilon > 0$ where $\epsilon = d(x, y)$ and $U(x, \frac{\epsilon}{2})$ is a neighbourhood of $x$ which does not contain $y$).

A subset $U$ of a metric space $X$ is said to be **open** if it is a neighbourhood of each of its points i.e. for each $x \in U$ there is an $\epsilon > 0$ so that $U(x, \epsilon) \subset U$. This is equivalent to the fact that $U$ is a union of open balls. It is clear that the empty set $\emptyset$ and $X$ are open. Also any union resp. finite intersection of open sets is open.

**Continuity of mappings:** Let $(X, d)$ and $(Y, d_1)$ be metric spaces, $f : X \to Y$ a mapping between them. $f$ is **continuous** at $x_0 \in X$ if for each positive $\epsilon$ there is a $\delta > 0$ so that $d(x, y) < \delta$ implies that $d_1(f(x), f(y)) < \epsilon$. It is **continuous on** $X$ if it is continuous at each point of $X$. This can be expressed in terms of neighbourhoods resp. open sets as follows. $f$ is **continuous at** $x_0$ if and only if whenever $N$ is a neighbourhood of $f(x_0)$, $f^{-1}(N)$ is a neighbourhood of $x_0$. (for the original condition means that $f^{-1}(U(f(x_0), \epsilon) \supset U(x_0, \delta))$. Similarly $f$ is **continuous on** $X$ if and only if $f^{-1}(U)$ is open for each $U$ open in $Y$.

A bijection $f : X \to Y$ is a **homeomorphism** if both $f$ and $f^{-1}$ are continuous.

As in the case of functions on $\mathbf{R}$, we can use the metric to introduce two stronger notions of continuity — those of uniform continuity and Lipschitz continuity. A function $f : X \to Y$ as above is **uniformly continuous** if for each positive $\epsilon$ there is a positive $\delta$ so that $d_1(f(x), f(y)) < \epsilon$ whenever $d(x, y) < \delta$. $f$ is **Lipschitz continuous** if there is a positive $K$ so that we have the estimate $d_1(f(x), f(y)) \leq K d(x, y)$ for $x, y$ in $X$. It is clear that Lipschitz continuity implies continuity which in turn implies continuity. Of course, none of these implications is reversible.

Such functions enjoy the expected stability properties. Thus compositions of continuous (uniformly continuous resp. Lipschitz continuous functions) have the same properties. Care is required, however, with products. Thus the product $f.g$ of two Lipschitz continuous functions with values in $\mathbf{R}$ need not be Lipschitz (as the example where both $f$ and $g$ are the identity mapping on the real line shows). However, the product of two bounded, Lipschitz continuous functions is Lipschitz continuous, as the reader can verify. Similar remarks apply to uniformly continuous functions.

**Convergence of sequences:** A sequence $(x_n)$ in a metric space $(X, d)$ converges to a point $x$ in $X$ if $d(x_n, x) \to 0$. In terms of neighbourhoods, this can be reformulated as follows: $x_n \to x$ if and only if for each $N \in \mathcal{N}(\S)$ there is a $K \in \mathbf{N}$ so that $x_n \in N$ whenever $n \geq K$. Note that a mapping $f$ from $X$ into $Y$ is then continuous if and only if $f(x_n) \to f(x)$ in $Y$ whenever $x_n \to x$ in $X$.

The **closure** of a set $A$ is the set $\bar{A}$ consisting of those $x$ which have the following property: for each positive $\epsilon$, there is a $y \in A$ with $d(x, y) < \epsilon$ (i.e. for each $\epsilon > 0$, $U(x, \epsilon) \cap A \neq \emptyset$). We say that $A$ is **closed** if $A = \bar{A}$. This can be reformulated as follows: for each $y$ not in $A$ there is a $\delta > 0$ so that $U(y, \delta) \subset X \setminus A$ i.e. $X \setminus A$ is open. Notice that $x$ is in the closure of $A$ if and only if there is a sequence $(x_n)$ in $A$ which converges to $x$ (take $x_n$ to be an element of $A \cap U(x, \frac{1}{n})$). Hence $A$ is closed if and only if it contains the limits of converging sequences with terms in $A$. In the metric space $(X, d_D)$, a sequence converges to $x$ if and only if there is an $N \in \mathbf{N}$ so that $x_n = x$ for $n > N$. (We say then the sequence is **eventually constant**). In a space with the indiscrete semi-metric, *every* sequence converges to every point. (This shows, in particular, that in the absence of the definiteness condition on the metric, limits of sequences need not be unique). Every function from a set $X$ with the discrete metric into a second metric space is continuous. On the other hand, every function from a metric space into a set with the indiscrete semi-metric is continuous.

A subset $A$ of a metric space $X$ is **dense** in $X$ if its closure is equal to $X$. This means that every element of the latter space is the limit of a sequence in $A$. The classical example is $\mathbf{Q}$, the set of rational numbers, which is dense in the real line.

We continue with some more pathological examples of metric spaces:

IV. **The post office metric:** This is a metric on the plane. The distance between two points $x$ and $y$ is defined to be the sum $|x| + |y|$ of the lengths of the vectors if $x$ and $y$ are distinct. Otherwise it is defined to be zero. (The reader should imagine the origin as an operating centre through which every signal from $x$ to $y$ has to pass). One can check that this is a metric. In fact, it is probably easier to verify the following more general fact from which it follows immediately. If we define the distance between two points $P$ and $Q$ in the plane to be the shortest length of a path from $P$ to $Q$ under suitable restraints, then this will be a metric provided the latter satisfy some natural conditions such as the following:

a) the constant path from $P$ to $P$ satisfies the restraints;

b) if a suitable path from $P$ to $Q$ satisfies the restraints, then the same path taken backwards is a suitable one from $Q$ to $P$;

c) if we have a suitable path from $P$ to $Q$ resp. from $Q$ to $R$, then we can join them together to form a suitable path from $P$ to $R$. In the above case, a path from $P$ to $Q$ is defined to be suitable if it goes through the origin (Except in the trivial case where the endpoints coincide — then there is no restraint).

Using this general fact it is easy to describe two further examples — the taxi-driver's metric and the Washington DC metric.

**Metrics on $\mathbf{R}^n$:** We have already considered the euclidean metric on $n$-space. Two further useful metrics are

VI. $d_1$ where

$$d_1(x, y) = \sum_{i=1}^{n} |\xi_i - \eta_i|.$$

VII. $d_\infty$ where

$$d_\infty(x, y) = \max |\xi_i - \eta_i|.$$

Note that we have the following inequalities:

$$d_\infty \leq d_2 \leq d_1 \leq n d_\infty.$$

From this it follows immediately that each of these metrics induces the same notion of convergence.

We now consider some infinite dimensional spaces.

VIII. $\ell^\infty$. This is the space of bounded sequences in $\mathbf{R}^{\mathbf{N}}$ i.e.

$$\ell^\infty = \{x = (\xi_n) : \sup |\xi_n| < \infty\}.$$

On this space we define the **supremum metric**

$$d_\infty = \sup\{|\xi_n - \eta_n| : n \in \mathbf{N}\}.$$

IX. $\ell^2$ is the space of sequences $x = (\xi_n)$ which are **square summable** i.e. such that $\sum |\xi_n|^2 < \infty$. The metric $d_2$ on this space is defined by the formula $d_2(x, y) = \sqrt{\sum |\xi_i - \eta_i|^2}$. X. $\ell^1$ is the space of **absolutely**

**summable** sequences i.e. those $x \in \mathbf{R}^{\mathbf{N}}$ with $\sum |\xi_n| < \infty$. On this space we use the metric

$$d_1(x, y) = \sum |\xi_n - \eta_n|.$$

XI. $C([0,1])$ denotes the space of continuous, real-valued functions on the unit interval. On this space we define the following three metrics: the supremum metric: $d_\infty(x, y) = \sup_t |x(t) - y(t)|$; the $L^2$-metric: $d_2(x, y) = \sqrt{\int_0^1 |x(t) - y(t)|^2 dt}$; the $L^1$-metric: $d_1(x, y) = \int_0^1 |x(t) - y(t)| dt$.

We omit the routine and unexciting proofs that the above really are metrics. In contrast to the finite dimensional case, these three metrics define completely different notions of convergence in the space $C([0,1])$.

We now introduce two further, highly important examples of metric spaces:

XII. The **Hilbert cube** i.e. the subset

$$\{x = (\xi_n) \in \ell^2 : |\xi_n| \leq \frac{1}{n}\}$$

of $\ell^2$ — regarded as a metric space with the metric $d_2$. (This is the natural infinite dimensional analogue of the hypercube in $\mathbf{R}^n$).

XIII. The **Cantor set**: Recall the classical definition of the Cantor set via missing thirds. We consider the sequence $(F_n)$ of closed subsets of the unit intervals as in Figure 1 (i.e. those obtained by removing successively the middle thirds of the intervals). The intersection $\bigcap F_n$ is a closed subset of the interval and, in particular, a metric space. It consists of those points in the unit interval which have a tryadic representation of the form $\sum_{n=1}^\infty a_n 3^{-n}$ where the coefficients $a_n$ are either 0 or 2. (For we remove those tryadic numbers whose $n$-th coefficient is 1 at the $n$-th stage of the construction). Because of its importance we denote this topological space by **Can** Note that at the $k$-th level of the construction, we split the interval into $2^k$ sub-intervals. If we denote these by $A_1^k, \dots, A_{2^k}^k$, then $F_k = \bigcup_{i=1}^{2^k} A_i^k$ and the Cantor set is

$$\mathbf{Can} = \bigcap_{k=1}^{\infty} \bigcup_{i=1}^{2^k} A_i^k.$$

Later we shall encounter a less geometric description of the Cantor set, which, however, is often more convenient for analysing its properties.

We now turn to some aspects of the theory of metric spaces which use the metric explicitly. We begin with *completeness*. This is a generalisation of the distinguishing characteristic of the real numbers in contrast to the rational ones — the existence of limits for sequences which *should* converge. This is usually an essential property in applications to analysis.

y

**Definition:** A sequence $(x_n)$ in a metric space is **Cauchy** if the following condition is satisfied: for each positive $\epsilon$ there is an $N \in \mathbf{N}$ so that $d(x_m, x_n) < \epsilon$ for $m, n > N$. The metric space $(X, d)$ is **complete** if each Cauchy sequence converges there.

Before considering concrete examples, we note that the property of a sequence being Cauchy is not a topological one i.e. it is not preserved under homeomorphisms. For example, consider the mapping $f : n \mapsto \frac{1}{n}$ from $\mathbf{N}$ onto $D = \{\frac{1}{n} : n \in \mathbf{N}\}$. We regard both of these sets as metric spaces (with the metric induced from $\mathbf{R}$). Then $f$ is a homeomorphism. However the sequence $(\frac{1}{n})$ is Cauchy in $D$ but its image in $\mathbf{N}$ under the continuous mapping $f^{-1}$ is not Cauchy there.

**Examples of complete spaces:** I. Any set $X$, when provided with the discrete metric, is complete for the trivial reason that any Cauchy sequence therein is eventually constant and so convergent.
II. The real line is complete. This is a fundamental property which will be familiar from a course on elementary analysis.
III. It follows easily from the completeness of the real line that the higher dimensional spaces $\mathbf{R}^p$ are complete (with respect to any of the three metrics $d_1, d_2, d_\infty$ introduced above). For a sequence in $\mathbf{R}^p$ is Cauchy resp. convergent if and only if the same is true of each of its components.
IV. Any closed subset of $\mathbf{R}^p$ is complete. For if $(x_n)$ is a Cauchy sequence in such a set $B$, then this sequence has a limit in $\mathbf{R}_p$ which, by the closedness, must lie in $B$.
V. An example of a space which is not complete is the open interval $]0, 1[$ in the line. This fails to be complete since the Cauchy sequence $(\frac{1}{n})$ there is not convergent. (Note that this space is homeomorphic to the real line which *is* complete).

The above examples suggest the following simple facts:
a) if $X$ is a complete metric space, then each closed subset thereof is complete under the induced metric;
b) If a subset $X_1$ of a metric space $X$ is complete for the induced metric, then $X_1$ is closed in $X$.

We now consider the completeness of some infinite dimensional spaces: The space $\ell^\infty$ is complete, as are $\ell^1$ and $\ell^2$. More generally, if $S$ is an arbitrary space, we can define a natural generalisation $\ell^\infty(S)$ of the former space as follows. This space is also complete.

The space $C([0, 1])$. This is complete with respect to the supremum norm since it is a closed subspace of the complete space $\ell^\infty(S)$ (this is just a restatement of the well-known fact that the uniform limit of continuous functions is continuous). We remark here that the space $C([0,1])$ is *not* complete for the metrics $d_1$ and $d_2$.

Recall that a function $f : X \to X_1$ where $(X, d)$ and $(X_1, d_1)$ are metric spaces is **Lipschitz continuous** if there is a $K \geq 0$ so that

$$d_1(f(x), f(y)) \leq K d(x, y)$$

for each $x, y \in X$. The smallest such $K$ is then called the **Lipschitz constant** of $f$. If the latter is strictly smaller than 1, then $f$ is called a **contraction**. Lipschitz continuous functions enjoy an important property which is described in the following Proposition:

**Proposition 23** *Let $(X, d)$ and $(X_1, d_1)$ be metric spaces where $X_1$ is complete and suppose that $X_0$ is a dense subset of $X$. Suppose that $f : X_0 \to X_1$ is a Lipschitz-continuous function. Then it can be extended to a Lipschitz-continuous function $\tilde{f}$ from $X$ into $X_1$. Furthermore, this extension is unique and has the same Lipschitz constant as $f$.*

PROOF. The proof is based on the simple observation that the image of a Cauchy sequence under a Lipschitz mapping is also Cauchy. Suppose that $x$ is a point of $X$. It is then the limit of a sequence $(x_n)$ in $X_0$. The latter is then Cauchy in $X_0$ and hence the image sequence $(f(x_n))$ is Cauchy in $X_1$. By the completeness of the latter there is a $y$ in $X_1$ to which $(f(x_n))$ converges. The remainder of the proof consists of the verification of the following facts: 1) $y$ depends only on $x$ (and not on the choice of the sequence $(x_n)$). For if $(z_n)$ is a second sequence in $X_0$ which tends to $x$, then $d(x_n, z_n)$ tends to zero. It follows from the Lipschitz condition that $d(f(x_n), f(z_n))$ also tends to zero. Hence $(f(x_n))$ and $(f(z_n))$ have the same limit. 2) The function $\tilde{f}$ which maps $x$ to the above $z$ if $x$ is not in $X_0$ (and agrees with $f$ on $X_0$) is Lipschitz. For if $K$ is the Lipschitz constant of $f$ and $(x_n)$ resp. $(y_n)$ are sequences in $X_0$ which tend to $x$ resp. $y$, then the inequality

$$d(f(x_n), f(y_n)) \leq K d(x_n, y_n)$$

tends in the limit to the desired inequality

$$d(f(x), f(y)) \leq K d(x, y).$$

■

**Remark:** In this proof we have tacitly used the fact that if $x_n \to x$ and $y_n \to y$, then $d(x_n, y_n) \to d(x, y)$. The reader should check that this holds.

In the light of the above proposition, it is an important fact that every metric space can be completed in the sense that it can be imbedded as a dense subset of a complete space. We prove this by means of the following Lemma. Before stating it, recall that an **isometry** between two metric spaces $(X_1, d_1)$ and $(X_2, d_2)$ is a bijection $f$ so that $d_2(f(x), f(y)) = d_1(x, y)$ for each pair $x, y$ in $X_1$. If such an $f$ exists, $X_1$ and $X_2$ are said to be **isometric** or **isometrically isomorphic.**

**Proposition 24** *Let $(X, d)$ be a metric space. Then $X$ is isometric to a subspace of the metric space $\ell^\infty(X)$.*

PROOF. Choose a fixed $x_0$ in $X$ (we are assuming that $X$ is non-empty). If $x \in X$, let $f_x$ be the function

$$f_x : y \mapsto (d(x, y) - d(x, x_0)).$$

Then $|f_x(y)| \leq d(x, x_0)$ for each $y$ and so $f_x \in \ell^\infty(X)$. The mapping $x \mapsto f_x$ is the required isometry.

■

**Proposition 25** *Let $(X, d)$ be a metric space. Then $X$ is isometric to a dense subset of a complete metric space.*

PROOF. The closure of the image of $X$ in $\ell^\infty(X)$ as in the above Lemma is the required space.

■

We note the following properties of this embedding which follow immediately from the above results:

I. Consider the embedding $j$ from $X$ into the above complete space which we denote by $Y$. Then if $f$ is a Lipschitz continuous mapping from $X$ into a second complete space $Z$, there is a unique Lipschitz mapping $\tilde{f}$ from $Y$ into $Z$ so that $f = \tilde{f} \circ j$. (This is just a rather abstract and pedantic way of saying that $\tilde{f}$ is an extension of $f$ if we regard $X$ as a subspace of $Y$ via the isometry $j$).

II. Any two complete spaces which contain $X$ as dense subspaces in the above way are isometrically isomorphic. More precisely, if $Y_1$ is a second space into which $X$ is isometrically embedded as a dense subspace via a mapping $j_1$, then there is an isometric isomorphism $I$ from $Y$ onto $Y_1$ so that $I \circ j = j_1$. The proof is simple. The mapping $j_1$ from $X$ into $Y_1$ extends to a mapping $I$ from $Y$ into $Y_1$ by the above extension property. $I$ has the required properties.

This result implies that the space $Y$ constructed above is essentially unique and so independent of the method of construction. Hence we can simply refer to it as *the* completion and denote it by $\hat{X}$.

## 3.3 Products of metric spaces:

If $(X_1, d_1), \dots, (X_n, d_n)$ are metric space, there are various natural ways of defining a metric on their product $X_1 \times \dots \times X_n$. As in the case of $\mathbf{R}^n$, there are three particularly simple ones:

$$\begin{aligned} x, y) &\mapsto \sum d_i(x_i, y_i) \\ (x, y) &\mapsto \sqrt{\sum d_i(x_i, y_i)^2} \\ (x, y) &\mapsto \max(d_i(x_i, y_i)) \end{aligned}$$

where $x = (x_i), y = (y_i)$.

Since these metrics are equivalent in the sense that they define the same notions of convergent resp. Cauchy sequences, we are free to use any one of them — it will be convenient to use the third one.

In the case of countable products $\prod_{n \in \mathbf{N}} X_n$ of metric spaces, we shall assume at first that each metric is bounded by 1 i.e. that $d_n(x, y) \leq 1$ for each pair $x, y$ in $X_n$. Then we define a metric $\tilde{d}$ on the Cartesian product as follows: if $x = (x_n)$ and $y = (y_n)$, then

$$\tilde{d}(x, y) = \sum \frac{1}{2^n} d_n(x_n, y_n).$$

The properties of the product which we shall require are as follows:

- a sequence $x_k = (x_n^k)$ in the product is Cauchy if and only if for each $n$ the sequence $(x_n^k)_{k=1}^\infty$ in $X_n$ is Cauchy.
- a sequence $(x_k)$ as above converges to $x = (x_n)$ if and only if $x_n^k \to x_n$ in $X_n$ for each $n$.
- $\prod X_n$ is complete if each $X_n$ is.

Proof. We prove the first statement. The proof of 2) is almost identical and 3) follows from 1) and 2). Firstly, the natural projection from $\Pi X_n$ onto $X_m$ is clearly Lipschitz (with constant $2^m$) for each $m$. This implies the necessity of the above condition.

■

Now suppose that each sequence $(x_n^k)$ (as $k$ varies) is Cauchy. We shall show that $(x^k)$ is Cauchy. Given a positive $\epsilon$, choose $N$ so that $2^{-N} < \frac{\epsilon}{3}$ and a $K \in \mathbf{N}$ so that if $k, l \geq K$, then $d_n(x_n^k, x_n^l) \leq \frac{\epsilon}{3}$ for $n = 1, \dots, N$. Then $d(x^k, x^l) \leq \epsilon$ if $k, l \geq K$.

We remark now that if $(X, d)$ is an arbitrary metric space, then $e : (x, y) \mapsto \min(1, d(x, y))$ is a metric on $X$ which has the same Cauchy sequences resp. convergent sequences as $d$. Hence we can define a countable product of arbitrary metric spaces by replacing their metrics by ones which are bounded as above and then using this construction.

We shall be interested in the following examples of products of the form $X^{\mathbf{N}}$ where $X$ is a fixed metric space. This is the special case of a product where each of the components is the same.

I. $\mathbf{I}^{\mathbf{N}}$ — a product of countably many copies of the unit interval. This space is essentially the Hilbert cube. More precisely, the mapping $(\xi_n) \mapsto (\frac{1}{n}\xi_n)$ is a homeomorphism from $\mathbf{I}^{\mathbf{N}}$ onto the Hilbert cube.

II. $\mathbf{N}^{\mathbf{N}}$ — the product of countably many copies of the natural numbers. (the latter regarded as a metric space with the discrete metric). Strangely enough, this space is homeomorphic to the irrational numbers and, from a topological point of view, is their most convenient representation. The exact proof of this requires some elementary number theory. Since we shall not require this identification directly, we shall be content with the remark that each sequence $(\xi_n, \xi_2, \dots)$ in $\mathbf{N}^{\mathbf{N}}$ defines an irrational number by means of a continued fraction.

This defines a mapping from $\mathbf{N}^{\mathbf{N}}$ into the irrationals which is, in fact, a homeomorphism onto. This fact can also be proved without recourse to number theory as follows. We list the rationals as a sequence $r_1, r_2, \dots$ and construct a countable family $\mathcal{I}_\backslash$ of partitions of the irrationals into countable collections of intervals so that the length of each interval in $\mathcal{I}_\backslash$ is at most $2^{-n}$ and so that $r_n$ is the endpoint of one of the intervals of $\mathcal{I}_\backslash$ but not of any previous partition. Then each irrational number $x$ is in precisely one intersection of the form

$$I_{n_1}^1 \cap I_{n_2}^2 \cap I_{n_3}^3 \cap \dots$$

where $I_{n_1}^1$ denotes the $n_1$-th interval in $\mathcal{I}_\infty$ etc. The mapping

$$x \mapsto (n_1, n_2, \dots)$$

is the required homeomorphism from the irrationals onto $\mathbf{N}^{\mathbf{N}}$.

III. The space $\{0, 1\}^{\mathbf{N}}$, i.e. the infinite product of countably many copies of the discrete space $\{0, 1\}$. This space is homeomorphic to the Cantor set. In fact, if $x = (\xi_n) \in \{0, 1\}^{\mathbf{N}}$, then $f(x) = \sum_{n=1}^{\infty} \frac{2\xi_n}{3^n}$ is an element of

$[0,1]$ — indeed, of the Cantor set. Once again, the mapping $x \mapsto f(x)$ is a homeomorphism from the product space $\{0,1\}^{\mathbf{N}}$ onto **Can**.

For future reference, we note the following fact which follows immediately from this description of the Cantor set. There is a continuous surjection from the Cantor set onto the unit interval **I**. Indeed

$$\{0,1\}^{\mathbf{N}} \in (\xi_n) \mapsto \sum_{n=1}^{\infty} \frac{\xi_n}{2^n}$$

is such a mapping.

Now this implies that there is a continuous surjection from $\mathbf{Can}^{\mathbf{N}}$ onto $\mathbf{I}^{\mathbf{N}}$. But $\mathbf{I}^{\mathbf{N}}$ is the Hilbert cube and $\mathbf{Can}^{\mathbf{N}}$ is $(\{0,1\}^{\mathbf{N}})^{\mathbf{N}}$ and it should come as no surprise that the latter space is once again the Cantor set. In fact, we have:

**Proposition 26** *If $X$ is a metric space, then $X^{\mathbf{N}}$ and $(X^{\mathbf{N}})^{\mathbf{N}}$ are homeomorphic.*

PROOF. An element of the second set is a sequence of sequences in X and so can be regarded as a double sequence of such elements i.e. $X = (X_n)$ where $X_n = (x_m^n)$. Now the latter double sequence can be straightened out into a single sequence e.g. as the sequence $(x_1^1, x_1^2, x_2^2, x_2^1, x_3^1, \dots)$. This defines a homeomorphism from $(X^{\mathbf{N}})^{\mathbf{N}}$ onto $X^{\mathbf{N}}$.

■

Putting this together, we see that there is a continuous mapping from **Can** *onto* the Hilbert cube.

We now bring some results on metric spaces in which the completeness plays an essential role. Firstly, if $A$ is a subset of a metric space, then its **diameter** is the number

$$\operatorname{diam}(A) = \sup\{d(x,y) : x, y \in A\}.$$

(This can of course be infinite).

**Proposition 27** *Let $(A_n)$ be a sequence of non–empty, closed subsets of a complete metric space $(X,d)$ so that $A_{n+1} \subset A_n$ for each n and hboxdiam $A_n \to 0$. Then the intersection $\cap A_n$ is non-empty (and in fact consists of a single point).*

Proof. We choose for each $n$ an $x_n \in A_n$. Then of course $d(x_m, x_n) \leq \operatorname{diam} A_N$ whenever $m, n \geq N$. Hence $(x_n)$ is Cauchy. Let $x$ denote its limit. Then $x_m \in A_n$ for $m \geq n$ and so in the limit $x \in A_n$ since the latter is closed. Since this holds for each $n$, we have that $x$ is in the required intersection. Suppose now that there were a second element $y$ in this intersection, with $y \neq x$. There is an $n \in \mathbf{N}$ with $oxdiam\, A_n \leq d(x, y)$. $y$ cannot be an element of $A_n$ (since $x$ *is* in the latter). This provides a contradiction.

■

The next result is the basis of countless existence theorems in analysis.

**Proposition 28** *Let $(A_n)$ be a sequence of closed subsets of a complete metric space $(X, d)$ whose union is $X$. Then $\bigcup A_n^o$ is dense in $X$.*

Proof. Suppose, if possible, that the conclusion is not valid. Then there is a non-empty, open set $U_1$ with $U_1 \cap (\bigcup_n A_n) = \emptyset$. Since $U_1$ is not a subset of $A_1$, the difference $U_1 \backslash A_1$ is non-empty (and open) and so contains an open ball $V_1$ of radius $\epsilon_1 > 0$. Since $V_1$ is not a subset of $A_2$, $V_1 \backslash A_2$ contains an open ball $V_2$ of positive radius $\epsilon_2$. Continuing inductively, we obtain a decreasing sequence $(V_n)$ of open balls, whereby $V_n$ has radius $\epsilon_n$. Furthermore, $V_{n+1} \subset V_n \backslash A_n$. We can arrange for the radii $\epsilon_n$ to converge to zero. Then by Cantor's result the intersection of the $V_n$ contains a point $x$. It is clear that $x$ is in none of the $A_n$ which is a contradiction.

■

**Exercise:** In the above proof we committed the crime of applying the result of Cantor to a decreasing sequence of *open* balls. Use the fact that each open ball contains a smaller, closed one to correct the proof.

**Corollar 1** *If $X$ and $(A_n)$ are as above and $X$ is supposed to be non-empty, then there is an $n_0$ so that $A_{n_0}$ has non-empty interior.*

The above result is known as Baire's theorem. Almost the same proof demonstrates the following version of this result:

**Proposition 29** *If $X$ is as above and $(A_n)$ is a sequence of closed subsets so that the interior $(\bigcup A_n)^o$ of their union is non-empty, then there is an $n_0$ so that the interior of $A_{n_0}$ is non-empty.*

The above results can be usefully reformulated by using the following concepts. A subset $A$ of a metric space is **nowhere dense** if the interior of its closure is empty. It is **of first category** if it is expressible as a countable union of nowhere dense sets. It is **of second category** if it is not of first category i.e. if whenever $A$ is expressed as the union $\bigcup A_n$ of a countable family of sets, then at least one $A_n$ fails to be nowhere dense. Thus the theorem of Baire states that each non-empty, complete metric space is of second category.

**Examples:** $\mathbf{Q}$ is a set of first category in $\mathbf{R}$. In fact, any countable subset of a metric space in which singletons are open is of first category. Of course, $\mathbf{Q}$ is not nowhere dense. The Cantor set, regarded as a subset of $[0, 1]$, is a non-trivial example of a nowhere dense set (i.e. it is nowhere dense and *uncountable*).

**Definition:** A subset of a metric space is a $G_\delta$- set if it is an intersection of a sequence of open sets. Before stating the next version of Baire's theorem, we note that the intersection of two dense subsets of a metric space need not be dense – a typical example is the pair $\mathbf{Q}$ and $\mathbf{R}\setminus\mathbf{Q}$ of dense subsets of the line whose intersection is empty. However, the intersection of two *open*, dense subsets is dense (easy exercise). (In fact, the intersection of an open dense subset with *any* dense subset is dense). Our next version of Baire's theorem is a significant strengthening of this result for complete metric spaces:

**Proposition 30** *Let $(X, d)$ be a complete metric space. Then an intersection of a sequence of open, dense subsets is dense (and so a dense $G_\delta$-set).*

Proof. The proof is based on the simple fact that a subset of a metric space is open and dense if and only if its complement is a closed, nowhere dense set. Hence if each of the sequence $(U_n)$ is open and dense and if $A$ is their intersection, then the complement of $A$ is the union of the complements of the $U_n$. Hence if the latter set contains a non-empty, open set, then so must one of the sets $(X\setminus U_n)$ by Baire's theorem and this would contradict $U_n$'s status as an open dense subset.

■

**Proposition 31** *If $(G_n)$ is a sequence of dense $G_\delta$-subsets of a complete metric space, then their intersection is also a dense $G_\delta$-subset.*

As an example of an application of the theorem of Baire, we bring the following result on boundedness of sequences of functions (it is the basis of a number of famous results of Banach in functional analysis - notably of the Banach Steinhaus theorem):

**Proposition 32** *Let $M$ be a subset of the space $C(X)$ of continuous real-valued functions on a complete, non-empty metric space $X$ which is pointwise bounded i.e. such that for each $t \in X$ there is a $K > 0$ so that $|x(t)| < K$ for each $x \in M$. Then there is a non-empty, open subset $U$ of $X$ so that $M$ is uniformly bounded on $U$ (i.e. there is an $L > 0$ so that $|x(t)| < L$ for each $x \in M$ and $t \in U$).*

PROOF. This is a simple application of Baire's theorem using the sets

$$A_n = \{t \in X : |x(t)| \leq n \text{ for each } x \in M\}.$$

■

Another result which uses the completeness of the metric space is the famous fixed point theorem of Banach. We have already met an example of fixed point theorem — namely that of Brouwer which states that each continuous mapping on $B^2$ has a fixed point. In fact, the same result holds in higher dimensions and we shall prove this in a later chapter. The Banach fixed point theorem has an entirely different character. It is true for a much wider class of spaces (in fact, for any complete metric space) but requires a condition on the mapping which is much stronger than continuity. In return, the fixed point is unique and the proof provides a method of finding it, respectively good approximations to it.

**Proposition 33** *Let $(X, d)$ be a non-empty complete metric space and let $f$ be a contraction on $X$. Then $f$ has a fixed point.*

PROOF. We choose any point $x_0 \in X$ and define a sequence $(x_n)$ recursively as follows:

$$x_1 = f(x_0),\ x_2 = f(x_1), \dots, x_{n+1} = f(x_n), \dots$$

We shall show that this is a Cauchy sequence and hence convergent. Its limit $x$ is then a fixed point since

$$x = \lim x_n = \lim x_{n+1} = \lim f(x_n) = f(x).$$

Firstly, we have the estimate:

$$d(x_n, x_{n+1}) = d(f(x_{n-1}), f(x_n)) \leq \lambda d(x_{n-1}, x_n) \leq \dots \leq \lambda^n d(x_0, x_1)$$

where $\lambda < 1$ is such that

$$d(f(x), f(y)) \leq \lambda d(x, y)$$

for $x, y \in X$. This implies that

$$\begin{aligned} d(x_n, x_{n+p}) &\leq d(x_n, x_{n+1}) + \cdots + d(x_{n+p-1}, x_{n+p}) \\ &\leq l^n d(x_0, x_1) + \cdots + \lambda^{n+p-1} d(x_0, x_1) \\ &\leq \tfrac{\lambda^n}{1-\lambda} d(x_0, x_1) \end{aligned}$$

which tends to zero as $n$ tends to infinity.

∎

The reader will have observed that this proof provides a theoretical basis for the simpler iteration methods used for solving equations numerically.

In the above situation, the fixed point is unique as the reader can easily verify.

One consequence of this theorem is the so-called Lipschitz inverse function theorem, which we state for functions on $\mathbf{R}^n$.

**Proposition 34** *Let $A : \mathbf{R}^n \to \mathbf{R}^n$ be an invertible, linear operator and $f : \mathbf{R}^n \to \mathbf{R}^n$ be a Lipschitz-continuous mapping whose Lispchitz constant satisfies the inequality $oxLip\, f < |A^{-1}|^{-1}$. Then $A + f$ is a bijection and its inverse $(A + f)^{-1}$ is also Lipschitz-continuous, with constant at most $(|A^{-1}|^{-1} - \; Lip\,(f))^{-1}$.*

PROOF. The fact that $A+f$ is injective and that its inverse has the required Lipschitz constant follows from the estimate:

$$\begin{aligned} |(A + f)x - (A + f)y| &= |A(x - y) + f(x - y)| \\ &\geq |A(x - y)| - (\mathrm{Lip}\,((f))|x - y| \\ &\geq |A^{-1}|^{-1}|x - y| - (\mathrm{Lip}\,(f))|x - y| \\ &\geq (|A^{-1}|^{-1} - (\mathrm{Lip}\,(f))|x - y|. \end{aligned}$$

The surjectivity follows from the fact that the equation $(A + f)x = y$ can be rewritten as a fixed point equation as follows:

$$x = A^{-1}y - A^{-1} \circ f(x)$$

It follows from the Banach fixed point theorem that this equation has a solution.

∎

We have seen above that a closed subset of a complete metric space is itself complete. On the other hand, a subset of a metric space which is itself complete with respect to the induced metric is automatically closed. A rather more subtle question is the following: which subsets of a given complete metric spaces are topologically complete i.e. such that there is an equivalent metric on the subset for which it is complete? Another way of saying this is that the subset be homeomorphic to a complete metric space. For example, the open subset $]-1,1[$ of the space $[-1,1]$ is not complete but it *is* topologically complete, being homeomorphic to $\mathbf{R}$. This is a special case of the following result:

**Proposition 35** *Let $U$ be an open subset of a complete metric space $(X,d)$. Then $U$ is topogically complete.*

PROOF. We define a new metric $\tilde{d}$ on the space $U$ as follows:

$$\tilde{d}(x,y) = d(x,y) + |\phi(x) - \phi(y)|$$

where $\phi(x) = d(x, X\backslash U)^{-1}$. The proof consists of the verification of the following two facts:

a) $\tilde{d}$ is equivalent to the original metric $d$ on $U$;

b) $(U,\tilde{d})$ is complete. We prove the second statement and leave the (similar) proof of the first to the reader. Let $(x_n)$ be a $\tilde{d}$-Cauchy sequence. Then it is certainly $d$-Cauchy and so $d$-convergent, say to $x$. It suffices to show that $x$ lies in $U$ and $\tilde{d}(x_n, x) \to 0$. If $x$ does not lie in $U$, then $\phi(x_n) \to \infty$. This implies that $\tilde{d}(x_n, x_1) \to \infty$. But this is obviously incompatible with the fact that the sequence is $\tilde{d}$-Cauchy. Since $x \in U$, then $\delta$, the distance of $x$ from $X\backslash U$, is positive. $x_n \to x$ (for $d$) and so there is an $N$ in $\mathbf{N}$ so that $d(x_n, x) < \frac{\delta}{2}$ if $n \geq N$. Hence $\phi(x_n) \leq \frac{2}{\delta}$ for such $n$. If we now piece together the three pieces of information:

$d(x_n, x) \to 0$;

$d(x_n, X\backslash U)) \to d(x, X\backslash U)$;

$\{\phi(x_n)\}$ is bounded, then we can deduce that $\tilde{d}(x_n, x) \to 0$.

■

Recall the definition of a $G_\delta$-subset of a metric space. These are subsets which are describable as countable unions of open sets. Dually, we define the notion of an $F_\sigma$-set which is a countable union of closed sets.

**Examples:** Each closed subset $C$ of a metric space is trivially an $F_\sigma$-set but it is also a $G_\delta$. For the set $U_n = \{x \in X : d(x,C) < \frac{1}{n}\}$ is open (as the union of the open balls $U(y, \frac{1}{n})$ as $y$ ranges through $C$). $C$ is clearly the intersection of the $U_n$. By taking complements, we see that every open set also has both of these properties simultaneously. A non-example is provided by the rationals which form an $F_\sigma$-subset of the reals but not a $G_\delta$. For suppose that $Q$ has a representation as the intersection of a sequence $(U_n)$ of open (dense) subsets. Then the complement of each of the $U_n$ is nowhere dense. But if $(r_n)$ is an enumeration of the rationals, we can write $\mathbf{R} = \bigcup_n (X \backslash U_n) \cup (\bigcup \{r_n\})$ as a countable union of nowhere dense sets which contradicts Baire's theorem.

In a certain sense, $G_\delta$-subsets of metric spaces are the natural domains of definition of continuous mappings. For suppose that $f$ is a continuous mapping from a subset $A$ of a metric space $X$ into a complete metric space $Y$. If $x$ is in the closure of $A$, we define

$$oxosc\,(f;x) = \inf_n \sup \{d(f(y), f(z)) : y, z \in U(x, \frac{1}{n}) \cap A\}.$$

Then if $hboxosc\,(f;x) = 0$, $(f(x_n))$ is a Cauchy sequence whenever $(x_n)$ is a sequence in $A$ which converges to $x$. For if $\epsilon$ is a given positive number, there is a $K \in \mathbf{N}$ so that $d(f(y), f(z)) < \epsilon$ for $y, z$ in $U(x, \frac{1}{K}) \cap A$. Hence if we choose $N \in \mathbf{N}$ so that $d(x, x_n) < \frac{1}{K}$ for $n \geq N$, then $d(f(x_m), f(x_n)) < \epsilon$ for $m, n \geq N$. Hence if we set $A_0 = \{x \in A : hboxosc\,(f;x) = 0\}$, we can extend $f$ to a function $\tilde{f}$ from $A_0$ into $Y$ by defining $\tilde{f}(x)$ to be the limit of the sequence $(f(x_n))$. One then checks that this extension is well-defined (i.e. the value of $\tilde{f}(x)$ is independent of the choice of the sequence $(x_n)$) and that the function $\tilde{f}$ is continuous, much as in the proof of ????? Now $A_0$ is the intersection of the sequence of subsets $(A_n)$ where $A_n$ is the set of those $x$ in the closure of $A$ for which $hboxosc(f;x) < \frac{1}{n}$ and the latter is an open subset of $\bar{A}$ as can easily be checked. Hence $A_0$ is a $G_\delta$ in $\bar{A}$ and so in $X$. (For $\bar{A}$ is a $G_\delta$ and it is easily seen that $G_\delta$-subsets of $G_\delta$-sets are themselves $G_\delta$). We have thus proved the following result:

**Proposition 36** *Let $f$ be a continuous mapping from a subset $A$ of a metric space $X$ into a complete metric space $Y$. Then there is a $G_\delta$-set $A_0$ between $A$ and its closure so that $f$ can be extended to a continuous function $\tilde{f}$ from $A_0$ into $Y$.*

From this we can quickly deduce the following result:

**Proposition 37** *Let $X$ and $Y$ be complete metric spaces and let $f : A \to B$ be a homeomorphism between subsets of $X$ and $Y$. Then there are $G_\delta$-subsets $A_1$ and $B_1$ containing $A$ and $B$ respectively (and contained in their closures) so that $f$ extends to a homeomorphism from $A_1$ onto $B_1$.*

PROOF. We can extend $f$ to a continuous $\tilde{f}$ from $A_0$ into $Y$ and $g = f^{-1}$ to a continuous $\tilde{g}$ from $B_0$ into $X$ where $A_0$ and $B_0$ are suitable $G_\delta$'s. Then $A_1 = A_0 \cap \tilde{f}^{-1}(B_0)$ and $B_1 = B_0 \cap \tilde{g}^{-1}(A_0)$ are the required sets.

■

We remark here that in the above proof we used the simple fact that the pre-image of a $G_\delta$-set under a continuous mapping is also a $G_\delta$.

We are now in a position to describe those subsets of a complete metric space which are topologically complete. First we show that $G_\delta$-subsets have this property:

**Proposition 38** *A $G_\delta$-subset of a complete metric space $X$ is topologically complete.*

PROOF. $A$ is the intersection of the sequence $(U_n)$ of open subsets of $X$. We know that each $U_n$ is topologically complete. Hence so is their product $\Pi U_n$ (since products of sequences of complete metric spaces are themselves complete metric). Now $A$ is homeomorphic to a closed subset of the latter product (see the exercise below) and this finishes the proof.

■

**Exercise:** Show that the mapping $x \mapsto (x, x, x, \dots)$ is a homeomorphism from $A$ onto a closed subset of the product of the $U_n$.

We can now complete our characterisation of topologically complete subsets.

**Proposition 39** *A subset $A$ of a complete metric space $(X, d)$ is topologically complete if and only if it is a $G_\delta$.*

PROOF. We have already seen above that this condition is sufficient. We shall now verify the necessity. Suppose that $d_1$ is a metric on $A$ which is equivalent to $d$ there and is such that $(A, d_1)$ is complete. We now apply the above extension result to the identity from $A$ (as a subset of $X$) into the complete metric space $(A, d_1)$. We can extend this to a homeomorphism from a $G_\delta$-set $A_0$ (containing $A$ and contained in $\bar{A}$). But a homeomorphism is a bijection and there is no non-trivial extension of the identity on $A$ to a bijection. Hence $A$ is itself a $G_\delta$.

■

# 4 Topological spaces:

As we have seen, in many of the definitions of the last section, the open sets play a more fundamental role than the metric. If we recall the basic stability properties of the family of open sets in a metric space, we are led naturally to the following definition:

**Definition:** Let $X$ be a set. A **topology** on $X$ is a family $\tau$ of subsets which satisfies the following conditions: a) $\emptyset$ and $X$ are in $\tau$; b) the union of a subfamily of $\tau$ is a set of $\tau$; c) the intersection of a finite subfamily of $\tau$ is in $\tau$.
A **topological space** is a set$X$, together with a topology thereon. We will refer then to the topological space $(X, \tau)$ or simply $X$ if it is clear from the context which topology we are dealing with. We now reformulate some of the definitions of the last chapter in this context.

If $(X, \tau)$ is a topological space, we refer to the sets in $\tau$ as the **open sets** (or, more precisely, the $\tau$-open sets). A subset $A$ of $X$ is **closed** if its complement is open.

If $\tau_1$ and $\tau_2$ are topologies on $X$, we say that $\tau_1$ is **finer** than $\tau_2$ (alternatively, that $\tau_2$ is **coarser** than $\tau_2$), if $\tau_1 \supset \tau_2$. Two topological spaces $(X_1, \tau_1)$ and $(X_2, \tau_2)$ are **homeomorphic** if there is a bijection $f$ from $X_1$ onto $X_2$ so that a subset $U$ of $X_1$ is open if and only if $f(U)$ is open in $X_2$. Such an $f$ is called a **homeomorphism**.

A mapping $f : X_1 \to X_2$ is **continuous** if for each open subset $U$ of $Y$ the pre-image $f^{-1}(U)$ is open in $X_1$. Hence a bijection is a homeomorphism if and only if both it and its inverse are continuous.

We note without proof that most of the simple facts about continuous functions on subsets of the line translate without difficulty to the more general situation. These the sum and product of two continuous, real-valued functions are continuous as is the composition of two continuous functions. If a sequence of continuous, real-valued functions on a space $X$ converges uniformly, then the limit function is also continuous.

If $A$ is a subset of a topological space, we define its **interior** $A^o$ to be the union of all open subsets which are contained in $A$. Clear $A^o$ is open and in fact is the largest open set which is contained in $A$. Furthermore, $A$ is open if and only if $A = A^o$.

Similarly, we define the **closure** $\bar{A}$ of $A$ to be the intersection of all closed subsets $C$ which contain $A$. It is the smallest closed set containing $A$ and the latter is closed if and only if $A = \bar{A}$.

**Examples:** If $X$ is a set, then in $\tau_D$, every subset is closed and hence equal to its closure. On the contrary, the closure of every subset of $(X, \tau_I)$ (with the exception of the empty set) is equal to $X$. In $(\mathbf{N}, \tau_{cf}$, every finite subset is closed and so equal to its closure, while the closure of any infinite subset is the whole of $\mathbf{N}$.

We remark here that one can use the concepts of closure and closedness to characterise the continuity of a function. Thus for a function $f : X \to Y$ between topological spaces, the following conditions are equivalent to the fact that $f$ is continuous: a) if $C \subset Y$ is closed, so is $f^{-1}(C)$; b) for each $A \subset X$, $f(\bar{A}) \subset \overline{f(A)}$.

We list some simple properties of interiors and closures: If $A$ and $B$ are subsets of a topological space $X$, then

a) $x \in A^o$ if and only if there is an open subset $U$ with $x \in U$ and $U \subset A$;
b) $x \in \bar{A}$ if and only if every open set $U$ containing $x$ meets $A$.
c) $\bar{A} = X \backslash (X \backslash A)^0$;
d) $(A^o)^o = A^o$, $\bar{\bar{A}} = \bar{A}$;
e) if $A \subset B$, then $A^0 \subset B^0$ and $\bar{A} \subset \bar{B}$;
f) $\overline{A \cup B} = \bar{A} \cup \bar{B}$ and $(A \cap B)^o = A^o \cap B^0$.

Note that it is *not* true that $(A \cup B)^o = A^o \cup B^o$ or that $\overline{A \cap B} = \bar{A} \cap \bar{B}$ as the following examples show. Firstly, we take $A = [0, 1]$, $B = [1, 2]$. Then the equality $(A \cup B)^o = A^o \cup B^o$ is violated. On the other hand, if $A = ]0, 1[$ and $B = ]1, 2[$, then $\overline{A \cap B} \neq \bar{A} \cap \bar{B}$.

Suppose now that $A$ is a subset of a topological space $X$. Then $x \in X$ is called **a limit point** of $A$ if it lies in the closure of $A$ i.e. if every open set $U$ containing $x$ meets $A$; **a cluster point** of $A$ if every open set $U$ containing $x$ meets $A$ in a point other than $x$ (i.e. if $x$ is a limit point of $A \backslash \{x\}$); **an isolated point** of $A$ if it is in $A$ but is not a cluster point thereof i.e. there is an open set $U$ so that $U \cap A = \{x\}$.

The set of all cluster points of $A$ is called the **derived set** of $A$ and denoted by $A^d$. Then $\bar{A} = A \cup A^d$. $A$ is said to be **dense in itself** if $A = A^d$.

A subset $A$ of a topological space $X$ is **dense** in $X$ if $\bar{A} = X$ (i.e. if every non- empty subset of $X$ meets $A$). $X$ is **separable** if it has a countable dense subset. The classical example of a dense subset is the rational numbers as a subset of the real line. This shows that the latter space is separable.

If $A$ is a subset of a topological space, then we define its **boundary** $\partial A$ to be the set $\bar{A} \cap \overline{X \backslash A}$. Thus $x$ is in the boundary of $A$ if and only if each open subset $U$ which contains $x$ meets both $A$ and its complement. The reader will easily check that for domains (i.e. open sets) in $\mathbf{R}^n$, this coincides with the intuitive notion of a boundary.

Note that $x$ is in the boundary of $A$ if and only if it is a limit point of both $A$ and its complement.

The boundary of a set $A$ is closed and we have the inclusion $\partial(\partial A) \subset \partial A$. However, we do not always have equality as the example $A$ the subset of $[0,1]$ consisting of the rational elements of the latter shows. (Here the boundary of $A$ is the interval $[0,1]$ and the boundary of the latter is the two-point set $\{0,1\}$).

The reader can verify the following simple facts:

$$\begin{aligned}
A^o &= A \backslash \partial A \\
\bar{A} &= A \cup \partial A \\
\partial(A \cup B) &\subset \partial A \cup \partial B \\
\partial(A \cap B) &\subset \partial A \cup \partial B \\
\partial(X \backslash A) &= \partial A \\
\partial U &= \bar{U} \backslash U \text{ if } U \text{ is open} \\
\partial A &= \emptyset \text{ if and only if } A \text{ is both open and closed.}
\end{aligned}$$

A subset $U$ of a topological space $X$ is **regularly open** if it is equal to the interior of its closure. Dually, it is **regularly closed** if it is the closure of its interior. The subset $]-1,1[\backslash\{0\}$ of the real line is an example of an open set which is not regularly open. Note that if $A$ is an arbitrary subset of a topological space, then the interior of its closure is regularly open. Also if $U$ and $V$ are regularly open, then so is their intersection (but not necessarily their union as the above example shows). Corresponding results hold for regularly closed sets and are obtained by complementation.

In specifying a topology resp. in interpreting various topological concepts, it is often sufficient to consider only open sets of a special type. For example, for a function $f$ between metric spaces, it suffices for $f$ to be continuous that the pre-images of open *balls* under $f$ be open as the reader can easily verify. The appropriate concept in the case of a general topological space is that of a **basis** i.e. a subfamily $\mathcal{B}$ of a topology $\tau$ so that each open set $U$ is the union of sets in $\mathcal{B}$. More generally $\mathcal{B}$ is a **subbasis** if the family $\tilde{\mathcal{B}}$ consisting of the sets which are intersections of finite collections from $\mathcal{B}$ forms a basis

For example, the family $\{x\}_{x \in X}$ of singletons in $X$ forms a basis for the discrete topology. The set of pairs $\{x,y\}$ $(x, y \in X, x \neq y)$ forms a subbasis (if the cardinality of $X$ is at least three), but not a basis. In the real line, the family of open intervals is a basis for the usual topology, whereas the family of open intervals of length 1 is merely a subbasis. As indicated by the introductory remarks above, the family of open balls in a metric spaces forms a basis for the metric topology.

It is often more convenient to specify a topology by describing a basis rather than *all* open sets and we shall take advantage of this in the following.

Of course, a given family can be a basis for at most *one* topology so there is no ambiguity in specifying a topology in this way.

We recall here that a topological space is defined to be separable if it has a dense countable subset. This condition is satisfied by most of the spaces which arise in applications. However, as we shall see, it has certain disadvantages and it is often useful to replace it by the following stronger condition:

**Definition:** A space $X$ is said to be **countably generated** (or to satisfy the first axiom of countability) if it has a countable basis. Any such space is separable since a sequence with one element from each set of a countable basis is easily seen to be dense. Conversely, if $(X, d)$ is a separable *metric* space, then it has a countable basis. In fact, if $(x_n)$ is a dense sequence, then the countable family $\{U(x_n, 2^{-n})$ forms a basis.

As remarked above, our list of metric spaces supplies us with a variety of topological spaces, including most of those which are useful in analysis or geometry. Here we shall concentrate on more pathological spaces whose topologies cannot be defined by a metric. We begin with a series of topologies which are defined on *any* set $X$ and are independent of any structure which the latter may have.

## 4.1 Examples of a topological spaces

I. The **cofinal topology** $\tau_{cf}$ on $X$ has as open sets the empty set and those subsets of $X$ whose complements are finite. Similarly, the **co-countable topology** $\tau_{cc}$ consists of the empty set, together with those subsets whose complements are countable.
II. If $X$ is a set and $x_0$ is a distinguished point in $X$, then the **particular point topology** has as open sets all subsets of $X$ which contain $x_0$ (plus the empty set of course). The special case where the cardinality of $X$ is 2 is called the **Sierpinski space**. It can be represented schematically as in figure 1.

We now turn to a series of topologies which are related to specific structures on the underlying spaces:
III. **The Niemitsky half-plane:** The underlying space is the upper half-plane

$$H_+ = \{(\xi_1, \xi_2) \in \mathbf{R}^2 : \xi_2 \geq 0\}.$$

It is generated by the set of all open balls in the open half-plane (i.e. those $x$ in $H_+$ with $\xi_2 > 0$) plus sets of the form $U \cup \{x\}$ where $U$ is an open ball in the upper half-plane which touches the $x$-axis at the point $x$ (cf. figure 2).

IV. **Topologies defined by orders** The topology of $\mathbf{R}$ is intimately related to its order structure and it can be conveniently generalised as follows. Suppose that a set $X$ has a partial ordering $<$. Then the open intervals i.e. the sets of the form

$$]x, y[ = \{z \in X : x < z < y\}$$

form a basis for the **order topology** on $X$.

Particularly interesting among the ordered topologies are those on the so-called **ordinal spaces**. Suppose that $\Gamma$ is a limit ordinal. Then we can regard the intervals $[0, \Gamma]$ and $[0, \Gamma[$ as topological spaces in the above way. The most interesting case is where $\Gamma$ is the smallest uncountable ordinal. This is a useful source of counterexamples.

A further example of a topological space which arises from an order structure is:

V. **The Sorgenfrey line:** This is the real line, provided with the topology $\tau_{\text{sorg}}$ generated by the family

$$\{[x, y[ : x, y \in \mathbf{R}, x < y\}$$

of half-open intervals as basis. This topology is finer than the natural one. Closely related are the topologies $\tau_R$ and $\tau_L$ on $\mathbf{R}$ (R and L for right and left) which are generated by the families

$$\{]x, \infty[ : x \in \mathbf{R}\}$$

resp.

$$\{]-\infty, y[ : y \in \mathbf{R}\}.$$

For applications of topology to analysis, two concepts are of crucial importance – convergence (of sequences) and continuity (of functions). That of convergence allows one to provide a framework for the rigorous treatment of questions of approximation. There is a close connection between them. Thus for metric spaces we can define continuity in terms of convergence as follows: a function between metric spaces is continuous if and only if it maps convergent sequences onto convergent sequences, more precisely, if whenever $x_n \to x$ in $X$, then $f(x_n) \to f(x)$ in $Y$. On the other hand, we can characterise convergence in terms of continuity as follows. Let $(x_n)$ be a sequence in a metric space $X$, $x$ a point of $X$. Then $x_n \to x$ if and only if the following function is continuous. We denote by $Y$ the subspace of $\mathbf{R}$ consisting of the origin and all points of the form $\frac{1}{n}$ where $n$ runs through the set $\mathbf{N}_0$. Then the sequence converges to $x$ if and only if the function $f$ from $Y$ into $X$ which maps $\frac{1}{n}$ onto $x_n$ and 0 onto $x$ is continuous.

The definition of convergence for sequences in metric spaces can be given in the following version which carries over to the case of general topological spaces:

**Definition:** A sequence $(x_n)$ in a topological space $(X, \tau)$ converges to a point $x$ if and only if for each open set $U$ containing $x$ there is an $N \in \mathbf{N}$, so that $x_n \in U$ for $n \geq N$.

However, the following examples show that some rather peculiar things can happen with respect to convergence in general topological spaces and this will lead us to generalise the notion of convergence shortly:

This last example shows that two distinct topologies on a set can induce the same notion of convergence of sequences. As mentioned above, this fact makes it necessary to consider convergence for a more general class of objects than sequences. One possibility is the use of **filters**:

**Definition:** A **filter** on a set $X$ is a non-empty collection $\mathcal{F}$ of subsets of $X$ so that

a) each $A \in \mathcal{F}$ is non-empty;

b) if $A$ and $B$ belong to $\mathcal{F}$, then so does their intersection $A \cap B$;

c) if $A \in \mathcal{F}$, then every superset of $A$ is also in $\mathcal{F}$.

The most important example of a filter is the **neighbourhood filter** $\mathcal{N}(\S)$ of a point in a topological space i.e. the set of all subsets $A$ of $X$ which contain an open $U$ with $x \in U$ (such sets $A$ are called **neighbourhoods** of $x$).

Filters are often conveniently specified by the use of so-called **filter bases** which we define as follows: A collection $\mathcal{F}$ of non-empty subsets of a set $X$ is called a **filter basis** if it satisfies condition a) above and, in addition, the condition

b$'$ ) if $A$ and $B$ belong to $\mathcal{F}$, then there is a $C \in \mathcal{F}$ with $C \subset A \cap B$. Then the collection

$$\tilde{\mathcal{F}} = \{B \subset X : \text{there is an}\, A \in \mathcal{F}\, \text{with}\, \mathcal{A} \subset \mathcal{B}\}$$

is a filter on $X$. It is called the **filter** generated by $\mathcal{F}$.

Further examples of filters are:

I. If $A$ is a non-empty subset of a set $X$, then $\mathcal{F}(\mathcal{A})$, the family of subsets of $X$ which contain $A$, is a filter. It is generated by the filter basis $\{A\}$.

II. The following filter on $\mathbf{N}$ is of fundamental importance. It is called the **Fréchet filter**. The family of sets

$$\mathbf{N}_m = \{n \in \mathbf{N} : n \geq m\}$$

forms a filter basis in $\mathbf{N}$. The filter that it generates is the Fréchet filter. In other words, the Fréchet filter consists of those subsets which contain *almost all* positive integers.

III. If $\mathcal{F}$ is a filter on a set $X$ and $f$ is a mapping from $X$ into $Y$, then, as the reader will have no trouble in verifying, the family

$$f(\mathcal{F}) = \{\mathcal{A} \subset \mathcal{Y} : \{^{-\infty}(\mathcal{A}) \in \mathcal{F}\}$$

is a filter on $Y$. It is called the **image** of $\mathcal{F}$ under $f$. Note that this is not, in general, the same thing as the family of images of the sets of $\mathcal{F}$. However, the latter family *is* a filter basis which generates the above filter.

One example of this construction is as follows: suppose that we have a sequence $(x_n)$ in a set $X$. Then we can regard this as a mapping from $\mathbf{N}$ into $X$ and so we can define the **Fréchet filter** of the sequence to be the image of the corresponding filter in $\mathbf{N}$ which we defined in II.

We are now ready to introduce the notion of convergence for filters. A filter $\mathcal{F}$ on a topological space $X$ converges to a point $x$ there if it contains the neighbourhood filter of $x$. In terms of a filter basis $\mathcal{G}$ which generates $\mathcal{F}$ we can restate this as follows: for each $U \in \mathcal{N}(\S)$, there exists an $A \in \mathcal{G}$ so that $A$ is contained in $U$.

If we apply this to the Fréchet filter of a sequence $(x_n)$, we see that the latter converges to $x$ in the sense defined above if and only if its Fréchet filter converges.

All topological notions can be expressed in terms of filters. As an example, we consider the closure of a set. In contrast to the case of a metric space, it is not true in general that a subset is closed if it contains the limits of all sequences therein as we have seen above. However, we do have a corresponding and valid version of this result if we replace sequences by filters. Suppose that $A$ is a non-empty subset of $X$ and $\mathcal{F}$ is a filter on $A$. Then $A$, while it need not be a filter on $X$, *is* a filter basis and so *generates* a filter on $X$. (This filter is just the image of $\mathcal{F}$ under the embedding of $A$ in $X$ in the sense of III above).

**Proposition 40** *Let $A$ be a subset of a topological space $X$. Then a point $x$ of $X$ lies in the closure $\bar{A}$ of $A$ if and only if there is a filter on $A$ which is such that the filter it generates in $X$ converges to $x$. Hence $A$ is closed if and only if every filter on $A$ which generates a convergent filter in $X$ is such that the limit is in $A$.*

PROOF. We show that if $x$ is in the closure of $A$, then there is a filter in $A$ which generates one on $X$ which converges to $x$. But it is clear that $\{A \cap N : N \in \mathcal{N}(\S)\}$ has the required property.

■

If $\mathcal{F}$ is a filter, a **cluster point** for $\mathcal{F}$ is a point $x$ so that each neighbourhood of $x$ meets each $A \in \mathcal{F}$ (i.e. the set of cluster points is just the intersection $\bigcap_{A\in\mathcal{F}} \bar{A}$). Cluster points can be characterised as follows:

**Proposition 41** *$x$ is a cluster point of the filter $\mathcal{F}$ if and only if there is a finer filter $\mathcal{G}$ which converges to $x$. ($\mathcal{G}$ is finer than $\mathcal{F}$ means simply that $\mathcal{F} \subset \mathcal{G}$).*

PROOF. It is clear that if $\mathcal{G}$ converges to $x$, then $x$ is a cluster point of $\mathcal{G}$ and so of $\mathcal{F}$ (a limit of a convergent filter is clearly a cluster point). On the other hand, if $x$ is a cluster point, then each neighbourhood of $x$ meets each $A \in \mathcal{F}$. Hence

$$\{C \cap N : C \in \mathcal{F}, \mathcal{N} \in \mathcal{N}(\S)\}$$

is a filter base and so generates a filter $\mathcal{F}$ which is finer than both $\mathcal{F}$ and $\mathcal{N}(\S)$.

■

Recall that a mapping $f$ between topological spaces was defined to be continuous if inverse images of open sets are open. This is equivalent to each of the following conditions (where $f$ maps $X$ into $Y$ and $\mathcal{B}_\infty$ resp. $\mathcal{B}_\in$ are bases for the topologies of $X$ and $Y$ respectively).

for each $U \in \mathcal{B}_\in$, $f^{-1}(U)$ is open in $X$;

for each $x \in X$ and $U \in \mathcal{B}_\in$ containing $f(x)$, there is a $V \in \mathcal{B}_\infty$ containing $x$ with $f(V)$ contained in $U$.

We now display a natural characterisation of continuity which involves convergence of filters:

**Proposition 42** *Let $f : X \to Y$ be a function between topological spaces. Then*

*a) $f$ is continuous at $x_0$ if whenever $\mathcal{F} \to \S_\prime$ in $X$, then $f(\mathcal{F}) \to \{(\S_\prime)$ in $Y$;*

*b) $f$ is continuous on $X$ if whenever a filter $\mathcal{F}$ converges to a point $x$ in $X$, the image $f(\mathcal{F})$ converges to $f(x)$.*

Another possibility for generalising the concept of convergence is the use of nets. A **net** in a set $X$ is a family which is indexed by a directed set $A$ i.e. it is a mapping from $A$ into $X$ (written $(x_\alpha)_{\alpha\in A}$). Such a net converges to a point $x$ if for every open set $U$ containing $x$ there is a $\beta \in A$ so that $x \in U$ for $\alpha \geq \beta$. Then the following characterisations of closedness and continuity hold: $x$ lies in the closure of $A$ if and only if it is the limit of a convergent net in $A$; $f : X \to Y$ is continuous if and only if $x_\alpha \to x$ in $X$ implies $f(x_\alpha) \to f(x)$ in $Y$. This result and similar ones can be proved by using the following correspondence between nets and filters. If $(x_\alpha)$ is a net, then just as in the case of sequences, the set $\{F_\beta : \beta \in A\}$ where $F_\beta = \{x_\alpha : \alpha \geq \beta\}$ is a filter basis and we see that the corresponding filter converges to $x$ if and only if the original net does. On the other hand, if $\mathcal{F}$ is a filter, then we can construct a net as follows: we define the index set $A$ to be the set

$$\{(x, B) : B \in \mathcal{F} \text{ and } \S \in \mathcal{B}\}.$$

This is directed under the ordering defined by specifying that $(x, B) \leq (x_1, B_1)$ if and only if $B \supset B_1$. Then we can define a net $(x_\alpha)$ where $x_\alpha = x$ whenever $\alpha = (x, B)$. This net converges to a point $x_0$ if and only if the original filter does.

## 4.2 Special types of subsets of a topological spaces:

We now introduce some special classes of subsets of topological spaces. The first few are repetitions of definitions which we have already considered within the context of metric spaces. A subset $A$ of a topological space $X$ is

**nowhere dense** if the interior of its closure is empty;

of **first category** if it is the union of countably many nowhere dense sets;

**of second category** if it is not of first category.

We make some simple remarks on these definitions. Thus a set is nowhere dense if and only if its complement contains an open, dense set. Equivalently, if it is equal to its own boundary. Examples of nowhere dense sets are single points (if they are not isolated), the boundary of an open or closed set and the Cantor set (as a subset of the unit interval). Of course, finite unions of nowhere dense sets also enjoy this property.

The family of subsets of a given space which are of first category is closed under the formation of subsets and countable unions. A space is of second category if and only if whenever it is written as a countable union of closed sets, then at least one of these has non-empty interior. A space is said to be a **Baire** space if each non-empty open subset is of second category. Thus Baire's theorem can be interpreted as the statement that each complete metric space is a Baire space. It follows immediately from the definition that open subsets of Baire spaces are also Baire. Also being Baire is a local property i.e. if every $x$ in $X$ has an open neighbourhood which is Baire, then so is $X$. The definition of being a Baire space can usefully be reformulated as follows: the intersection of each countable family of open, dense subsets is itself dense.

The definitions of $F_\sigma$- and $G_\delta$-sets carry over directly from the case of metric spaces i.e. a subset of a topological space is an $F_\sigma$ if it is the union of countably many closed sets. It is a $G_\delta$ if it is an intersection of countably many open sets. Thus the class of $F_\sigma$-sets is closed under the formation of finite intersections and countable unions.

A subset of a topological has the **Baire property** (or is a Baire Property set, abbreviated to BP-set) if it has the form $U\Delta F$ where $U$ is open and $F$ is of first category (Here $\Delta$ stands for the symmetric difference i.e. $U\Delta F = (U\backslash F) \cup (F\backslash U)$. This is equivalent to the fact that it has a representation $C\Delta F$ where $C$ is closed and $F$ is of first category (for if $U$ is open, then $\bar{U}\backslash U$ is nowhere dense and so of first category). From this it follows immediately that if $A$ is a BP-set, then so is its complement.

The family of BP-sets is stable under the following operations:

formation of countable unions;

formation of countable intersections.

In particular, it follows from the above that the BP-sets form a $\sigma$-algebra.

We have the following useful characterisation of sets with the Baire property.

**Proposition 43** *A set has the Baire property if and only it has a representation as the union of a $G_\delta$-set and a set of first category (or, dually, as an $F_\sigma$-set minus one of first category).*

A particularly important $\sigma$-algebra associated with a topological space is the **Borel algebra**. This is, by definition, the $\sigma$-algebra generated by the open sets (i.e. the smallest $\sigma$-algebra which contains the family of open sets). Since each open set is a BP-set and the family of all sets with the Baire property is a $\sigma$-algebra, it follows immediately that each Borel subset of a topological space is a BP-set.

## 4.3 Semi-continuous functions:

In certain arguments, particularly those which arise in optimisation, the property of continuity can be replaced by a weaker one, that of semi-continuity. Since many functionals on infinite dimensional spaces possess the latter property while failing to be continuous, this generalisation of continuity is of some importance.

**Definition:** A real-valued function on a topological space $X$ is **lower semi-continuous** (abbreviated l.s.c.) if it is continuous when regarded as a mapping into the real line, where the latter is provided with the topology $\tau_R$. This means that for each $\alpha \in \mathbf{R}$, the set $\{f > \alpha\}$ is open. Dually, $f$ is **upper semi-continuous** (u.s.c.) if $-f$ is l.s.c. i.e. if $f$ is continuous as a function with values in $(\mathbf{R}, \tau_L)$. For example, if $A$ is a subset of $X$, then the characteristic function $\chi_A$ of $A$ is lower-semi-continuous if and only if $A$ is open and upper semi-continuous if and only if $A$ is closed.

It is easy to see that if $f$ and $g$ are l.s.c. then so are the functions $\min(f,g)$, $\max(f,g)$ and $f+g$. If both function are non-negative, then $fg$ is also l.s.c. More generally, we have:

**Proposition 44** *If $(f_\alpha)$ is a family of l.s.c. functions which is such that for each $x$ in $X$, the set $\{f_\alpha(x)\}$ of real numbers is bounded above, then the function*

$$f : x \mapsto \sup_\alpha (f_\alpha(x))$$

*is also l.s.c.*

In particular, any supremum of a family of continuous functions is l.s.c. (of course, it need not be continuous). Generally the converse is true i.e. a function is l.s.c. if and only if it is the supremum of a family of continuous functions. (*Precisely* when this is the case will be dealt with in the chapter on separation properties). We remark here that a stronger version holds in metric spaces:

**Proposition 45** *Let $f$ be an l.s.c. function on the metric space $X$. Then there is a sequence $(f_n)$ of continuous real-valued functions on $X$ such that $f$ is the supremum of the $f_n$.*

If $f$ is a function on a topological space $X$, we define

$$f^u = \inf\{h : f \leq h \,\text{and}\, h \,\text{is u.s.c.}\}$$

and

$$f_l = \sup\{h : f \geq h \,\text{and}\, h \,\text{is l.s.c.}\}.$$

Then $f_l \leq f \leq f^u$ and all three function agree at a point $x_0$ if and only if $f$ is continuous at $x_0$. This means that the set of discontinuities of $f$ is the union of the sets $A_n$ where

$$A_n = \{f^u - f_l \geq \frac{1}{n}\}$$

and so is an $F_\sigma$.

We conclude this section with a brief mention of some further classes of functions between topological spaces, namely open and closed mappings.

**Definition:** A continuous mapping $f : X \to Y$ between topological spaces is **open** if $f(U)$ is open in $Y$ whenever $U$ is open in $X$; **closed** if $f(C)$ is closed in $Y$ whenever $C$ is closed in $X$.

These two notions are distinct as the following examples show. Firstly, the mapping from $\mathbf{R}$ into $\mathbf{R}$ whose graph is shown in figure ??? is closed, but not open sind the image of the real line is a closed interval. On the other hand, the natural projection of the Niemitsky half-plane onto the $x$-axis (with the natural topology) is open, but not closed. We list some simple properties of open and closed mappings which follow immediately from the definitions:

a) if $f$ and $g$ are open, so is the composition $g \circ f$;

b) if $g \circ f$ is open and $f$ is onto, then $g$ is open;

c) if $g \circ f$ is open and $g$ is injective, then $f$ is open. Corresponding results hold for closed mappings

If $f$ is an open mapping from $X$ into $Y$ and $Y_0$ is a subset of $Y$, then the restriction of $f$ to $f^{-1}(Y_0)$ is also open. The same holds for closed mappings.

# 5 Construction of topological spaces:

In order to enrich our collection of examples of topological spaces, we now describe some simple ways of constructing new spaces from old ones. Firstly we note that any subset of a topological space has itself a natural topology. As open sets in $A$ we take the intersections of $A$ with open subsets of $X$. More formally, the family

$$\tau_A = \{U \cap A : U \in \tau\}$$

is a topology on $A$ – called the **topology induced** on $A$ by $\tau$. $A$ with this topology is called a **topological subspace** of $X$. We make the following simple remarks about induced topologies: I. If $\mathcal{B}$ is a basis (resp. a subbasis) for $\tau$, then

$$\mathcal{B}_\mathcal{A} = \{\mathcal{U} \cap \mathcal{A} : \mathcal{U} \in \mathcal{B}\}$$

is a basis (subbasis) for $\tau_A$.

II. On a subset $A$ of a metric space $(X, d)$, the topology induced by the metric topology coincides with that defined by the restriction of the metric to $A$. This follows easily from the fact that if $x \in A$, then the $\epsilon$-ball in $A$ with $x$ as centre (defined by the restriction of the metric to $A$), is just the intersection of the $\epsilon$-ball in $X$ with $A$.

III. If we regard $\mathbf{R}^m$ as a subspace of $\mathbf{R}^p$ in the natural way (for $m \leq p$), then the usual topology on the latter induces the usual topology on the former.

IV. If $f : X \to X_1$ is a mapping between topological spaces which takes its values in a subset $A$ of $X_1$, then $f$ is continuous (from $X$ into $X_1$) if and only if it is continuous from $X$ into $A$ (the latter with the induced topology).
V. A subset $C$ of the subspace $A$ of $X$ is closed for the induced topology if and only if it has the form $C_1 \cap A$ for a closed subset $C_1$ of $X$.
VI. If we regard the $x$-axis as a subset of the Niemitsky half-plane, then the induced topology is the discrete topology. This provides an example of a separable space with a subspace which is *not* separable.
VII. Suppose that $f$ is a mapping from the space $X$ into $Y$ and that $A$ is a subset of $X$. Then the restriction of $f$ to $A$ is continuous for the induced topology. If $\mathcal{F}$ is a filter in $A$ which converges to a point $x$ in $A$, then the filter generated by $\mathcal{F}$ in $X$ converges to $x$ there (and conversely).

A topological property is called **hereditary** if each subspace of a space with the property also enjoys the property. For example, we have seen that the property of being separable is not hereditary, whereas that of being countably generated is.

## 5.1 Products:

We now show how to regard Cartesian products of topological spaces as topological spaces. For the sake of simplicity, we begin with finite products. Later, we shall show how to deal with infinite products. Suppose that $X$ is the product $\prod_{k=1}^n X_k$ where each $X_k$ is provided with a topology $\tau_k$. Then the sets of the form

$$U_1 \times U_2 \times \cdots \times U_n$$

where $U_i$ is open in $X_i$ form a basis for a topology on $X$. We call it the **product topology**. Its characteristic property is summarised in the next result:

**Proposition 46** *The projection mappings $\pi_k : X \to X_k$ are all continuous and if $f$ maps a topological space $Y$ into $X$, then $f$ is continuous if and only if for each $k$, $\pi_k \circ f$ is continuous from $Y$ into $X_k$.*

A less formal way of stating this result is as follows: a continuous function from $Y$ into $X$ is just a system $(f_1, \dots, f_n)$ of continuous functions where $f_k$ takes its values in $X_k$. ($f_k$ is the mapping $\pi_k \circ f$).

A similar criterium for convergence is valid and can be derived immediately from the above result by using the trick mentioned on p. ????

**Proposition 47** *If $(x_n)$ is a sequence in the product space, then it converges to $x$ there if and only if $\pi_k(x_n) \to \pi_k(x)$ for each $k$.*

In other words, convergence is determined by that of components.

Another simple fact about products which is useful to remember is that if $A_k$ is a subspace of $X_k$, then the product $\prod A_k$ is a *topological* subspace of the product of the $X_k$ (i.e. the product topology on the latter induces the product topology on the former).

We have already met a number of examples of products. For example, the $n$-torus $tT^n$ is just the product of $n$ copies of the circle.

## 5.2 Quotient spaces:

This method of constructing topological spaces can be illustrated by the following simple example: if we take a copy of the unit interval and bend it round so that we can join the end-points, then we obtain a copy of the unit circle. We can say that $S^1$ is the space obtained from a closed interval by identifying the endpoints. The general construction is as follows: $X$ is a topological space and $f$ is a surjection from $X$ onto a *set* $Y$. Then we can define a topology $\tau_1$ on $Y$ as follows. A subset of the latter is defined to be open if its inverse image under $f$ is open in $X$.

In our applications, the function $f$ will arise in the following way. We have an equivalence relationship on $X$ and $Y$ is the set of equivalence classes, $f$ being the natural surjection which maps an element to the equivalence class to which it belongs.

The decisive property of the above topology is as follows: a mapping $g$ from $Y$ into a topological space $Z$ is continuous if and only if the composition $g \circ f$ is continuous.

**Examples:** We begin with three spaces which are useful as counterexamples and which can be most easily defined as quotient spaces:
I. The line with two origins (figure 1). This is the quotient of the union of two copies $X_1$ and $X_2$ of the real line whereby we identify the points in pairs, with the exception of the origin. More formally, consider the following two subsets of the plane:

$$X_1 = \{(\xi_1, 0) : \xi_1 \in \mathbf{R}\} \quad X_2 = \{(\xi_1, 1) : \xi_1 \in \mathbf{R}\}$$

and put $X = X_1 \cup X_2$ and $Y = X \mid_\sim$ where

$$x \sim y \text{ if and only if } \; x = y \text{ or } x = (\xi_1, \eta),\; y = (\xi_2, \eta) \text{ with } \eta \neq 0.$$

II. The interval with three endpoints (figure 2). This is $X \mid_\sim$ where

$$X = \{(\xi_1, 0) : \xi_1 \in [0, 1]\} \cup \{(\xi_2, 1) : \xi_1 \in [0, 1]\}$$

and

$$x \sim y \text{ if and only if } x = y \text{ or } x = (\xi_1, 0) \text{ and } y = (\xi_1, 1) \text{ with } \xi_1 > 0.$$

III. The pinched plane. This is an example of a quotient of a metrisable space which is not metric. It is typical in the sense that many desirable properties of topological spaces can be lost in the passage to a quotient space. The pinched plane is obtained from the plane by shrinking a line down to a point. More formally, it is $\mathbf{R}^2 |_\sim$ where $x \sim y$ if and only if $x = y$ or $x$ and $y$ lie on the real axis. (The quotient space is not metrisable since the special point in the pinched plane fails to have a countable neighbourhood basis).

The above examples illustrate two special types of quotient space which arise frequently, especially in geometric topology, namely spaces obtained by shrinking subsets to points resp. by pasting spaces together along suitable subsets. Because of the frequency of their occurrence, it is worth describing such constructions in their natural generality.

I. Spaces obtained by shrinking subsets to a point. Here $X_0$ is a subset of a topological space $X$ and the equivalence relation is defined as follows:

$$x \sim y \text{ if and only if } x = y \text{ or } x \text{ and } y \text{ are in } X_0.$$

Of course, example III above is a special case of this construction.

II. Spaces obtained by pasting. Here $X$ and $Y$ are topological spaces which we suppose to be disjoint (as sets) and $X_0$ resp. $Y_0$ are homeomorphic subsets, $h_0$ being a suitable homeomorphism from $X_0$ onto $Y_0$. Then we can paste $X$ onto $Y$ along this "common" subset as follows: we consider the quotient space of the union $X \cup Y$ under the following equivalence relationship:

$$x \sim y \text{ if and only if } x = y \text{ or } x \in X_0, y \in Y_0 \text{ and } y = h(x).$$

For example we obtain the **Bretzel** by taking two tori, cutting a hole in each of them and pasting the spaces together along the edges of the hole (see figure 3). Perhaps the simplest examples of this construction are those which involve joining two spaces at a point. Here we have two topological spaces $X$ and $Y$, with distinguished points $x_0$ and $y_0$ resp. The space $X \vee Y$ is then the quotient of the disjoint union $X \amalg Y$ obtained by identifying $x_0$ and $y_0$. Thus $\mathbf{S}^1 \vee \mathbf{S}^1$ is the figure of eight (see figure 4). (Strictly speaking, this wedge product depends on the particular choices of the distinguished points and this should have been incorporated in the notation. However, in the cases we are interested in, the choice of points is irrelevant so that this pedantry is unnecessary). We remark that the wedge product of $X$ and $Y$ can also be identified with the subset $X \times \{y_0\} \cup \{x_0\} \times Y$ of the product space $X \times Y$. For example, when both spaces are the circle, this means that we are identifying the figure of eight with the subset of the torus indicated in the diagram 5.

**Example: The cone over a space**. If $X$ is a topological space, the cone over $X$ is the quotient of the product $X \times I$ of $X$ with the unit interval which is obtained by identifying the points of the form $(x, 1)$ i.e. we pinch the top of the cylinder over $X$ together. Thus the cone over $\mathbf{S}^1$ is just the classical form of a conical dunce's cap (figures 6,7). The following more exotic cones are often used as counterexamples.

I. The cone over $\mathbf{Z}$. II. The cone over the space $\{\frac{1}{n} : n \in\} \cup \{0\}$.

We can picture both of these spaces as subsets of the plane (see figures 8 and 9) but the reader is warned that the quotient topologies described above do not coincide with the topologies induced from the natural one of the plane.

**Example:** The suspension of a space. This is defined as the quotient of the product of $X$ with the unit interval under the equivalence relationship

$$(x, s) \sim (y, t) \quad \text{if and only if} \quad x = y \quad \text{and} \quad s = t \quad \text{or} \quad s = 1 \quad \text{and} \quad t = 1 \quad \text{or} \quad s = 0q \quad \text{and}$$

In other words, we pinch together both the top and the bottom of the cylinder as in figure (10). For example, the suspension of $\mathbf{S}^1$ is $\mathbf{S}^2$. More generally, the suspension of $\mathbf{S}^n$ is $\mathbf{S}^{n+1}$ for each $n$.

## 5.3 Quotient topologies on subspaces

Suppose that $f : X \to Y$ is a surjection and that $Y$ has the corresponding quotient topology. Then this need not apply to subspaces of $Y$. More precisely, if $A$ is a subspace of $Y$ and if $B$ is a subset of $X$ with $f(B) = A$, then the topology induced on $A$ from $Y$ need not coincide with the quotient of the subspace topology on $B$. An example where this fails is the following: we take for $X$ the interval $I$, for $Y$ the circle and for $B$ the interval $]0, 1]$. It is easy to see, however, that the quotient topology is always finer than the subspace topology.

We remark here that if $f : X \to Y$ is a quotient mapping, then it is not necessarily open or closed. On the other hand, any open or closed continuous surjection is automatically a quotient mapping. Just when a quotient mapping is open or closed can be simply characterised as follows:

**Proposition 48** *A surjective mapping $f : X \to Y$ from the topological space $X$ to the set $Y$ is open (resp. closed) when $Y$ is provided with the quotient* *topology if and only the following condition holds: for each open subset $U$, $f^{-1}(f(U)$ is open resp. for each closed subset $C$ of $X$, $f^{-1}(f(C))$ is closed.*

The set $f^{-1}(f(A))$ which occurs in the above formulation is called the **saturation** of $A$. It has the following more intuitive description. $f^{-1}(f(A))$ is the union of all those equivalence classes which have non-empty intersection with $A$.

However, the reader can verify that this *is* the case if any of the following conditions are met:

a) $A$ is open or closed;

b) $f$ is open or closed.

As a further example of pathology which can occur in the case of quotient spaces, we note that if $X$ and $Y$ are product spaces, say $X = \prod X_\lambda$, $Y = \prod Y_\lambda$, and if $f_\lambda$ is a quotient mapping from $X_\lambda$ onto $Y_\lambda$ for each $\lambda$, then this need not be true of the product mapping from $X$ onto $Y$. However, if each $f_\lambda$ is open, then it *is* a quotient mapping (in fact an open mapping).

We now consider two very general constructions which contain all of the previous methods as special cases:

## 5.4 Initial and final topologies:

Let $X$ be a set, $\{(X_\alpha, \tau_\alpha)\}_{\alpha \in A}$ be a family of topological spaces and, for each $\alpha \in A$, let $f_\alpha$ be a mapping from $X$ into $X_\alpha$. Then $\mathcal{B}$, the set of those subsets of the form $f_\alpha^{-1}(V)$ for some $\alpha \in A$ and some open set $V$ in $X_\alpha$, is a subbasis for a topology $\tau$ on $X$. The latter is called the **initial topology on $X$ induced** by the $f_\alpha$. It can be characterised as the coarsest topology on $X$ for which each $f_\alpha$ is continuous.

Dual to the above is the construction of final topologies. Here $X$ and the $X_\alpha$ are as above but now the $f_\alpha$ are mappings from $X_\alpha$ into $X$. Then the set of those subsets $U$ of $X$ which have the property that $f_\alpha^{-1}(U)$ is in $\tau_\alpha$ for each $\alpha$ forms a topology on $X$ – the **final topology** induced by the $f_\alpha$. It is the finest topology so that each $f_\alpha$ is continuous.

An important aspect of the above topologies is the following description of continuous mappings: if $X$ is a topological space with the initial topology induced by the mappings $f_\alpha : X \to X_\alpha$, then a mapping $f$ from a further space $Y$ into $X$ is continuous if and only if $f_\alpha \circ f$ from $Y$ into $X_\alpha$ is continuous for each $\alpha$; if $X$ has the final topology induced by the mappings $f_\alpha$ from $X_\alpha$ into $X$, then a mapping $f$ from $X$ into a further space $Y$ is continuous if and only if $f \circ f_\alpha$ is continuous for each $\alpha$.

These characterisations follow immediately from the descriptions of the open subsets of $X$.

**Examples:** Examples of initial topologies are subspaces and products. More precisely, if $A$ is a subset of a topological space $X$, then the topology induced on $A$ from $X$ is precisely the initial topology corresponding to the inclusion mapping from $A$ into $X$. Similarly, if $X$ is the product $\prod_{k=1}^{n} X_k$ of a finite number of topological space, then the product topology on $X$ is just the initial topology induced by the mappings $\pi_1, \dots, \pi_n$.

Examples of final topologies are quotient spaces. A further example is that of disjoint unions. Let $(X_\alpha, \tau_\alpha)$ be a family of topological spaces and let $X$ be the (set-theoretical) disjoint union of the $X_\alpha$. Then we can regard $X$ as a topological space, by providing it with the final topology induced by the natural embeddings of the $X_\alpha$ in $X$. In other words, a subset $U$ of $X$ is open if and only its intersection with each $X_\alpha$ is open for $\tau_\alpha$. Thus a mapping $f$ from $X$ into a second topological space is continuous if and only if its restrictions to the $X_\alpha$ are continuous. $X$ with this topology is called the **topological disjoint union** of the $X_\alpha$. We note in passing that each $X_\alpha$ is then a clopen subset of $X$ and that a subset $C$ is closed if and only if $C \cap X_\alpha$ is closed in $X_\alpha$ for each $\alpha$.

**Infinite products:** The above description of finite products as initial structures should make it clear how we shall define infinite products. Suppose that we have an arbitrary family $(X_\alpha, \tau_a)_{\alpha \in A}$ of topological spaces and that $X$ is their set-theoretical product. Then for each $\alpha$ there is a natural projection $\pi_\alpha$ from $X$ into $X_\alpha$. Then we regard $X$ as a topological space with the corresponding initial topology. Suppose that $U_\beta$ is open in $X_\beta$. Then the pre-image of this set under $\pi_\beta$ is the product $U_\beta \times \prod_{\alpha \neq \beta} X_\alpha$. Hence the sets of this form are a sub-basis for the product topology. From this it follows that a basis consists of all sets of the form $\prod_{\alpha \in A} U_\alpha$ where each $U_\alpha$ is open in $X_\alpha$ and all but finitely many of the $U_\alpha$ are equal to the whole space $X_\alpha$. Using this description of the open sets it is not difficult to prove that the product topology has the natural properties, of which we mention explicitly two:

a) a function $f$ from a space $Y$ into $X$ is continuous if and only if $\pi_\alpha \circ f$ is continuous into $X_\alpha$ for each $\alpha$;

b) a filter $\mathcal{F}$ in $X$ converges to $x$ if and only if $\pi_\alpha(\mathcal{F})$ converges to $\pi_\alpha(x)$ for each $\alpha$.

**Projective limits:** One of the most ubiquitous constructions in mathematics is that of projective limits. It is based on the following set-theoretical construction. Suppose that we have a family $(X_\alpha)_{\alpha \in A}$ of sets which are indexed by a directed set $A$. In addition we have, for each pair $\alpha, \beta$ in $A$ with $\alpha \leq \beta$, a mapping $\pi_{\beta\alpha}$ from $X_\beta$ into $X_\alpha$ so that the following compatibility conditions are satisfied:

a) $\pi_{\alpha\alpha} =$ id;

b) $\pi_{\beta\alpha} \circ \pi_{\gamma\beta} = \pi_{\gamma\alpha}$ $(\alpha \leq \beta \leq \gamma)$. Such a family is called a **projective spectrum**. From now on we shall assume that the directed set is $\mathbf{N}$ in order to simplify the notation but the reader will have no difficulty in seeing that most of our considerations are equally valid in the general situation. In the case of spectra indexed by $\mathbf{N}$, it suffices to know the linking mappings $\pi_{n+1,n}$ for each $n$ since condition b) implies that

$$\pi_{mn} = \pi_{n+1,n} \circ \pi_{n+2,n+1} \circ \cdots \circ \pi_{m,m-1}.$$

Suppose then that we have a projective spectrum

$$\{\pi_{mn} : X_m \to X_n : n \leq m\}.$$

Its **projective limit** is then defined to be the family of **threads** in the Cartesian product $\prod X_n$ i.e. the set of those sequences $(x_n)$ for which $\pi_{n+1,n}(x_n + 1) = x_n$ for each $n$.

We begin with an example to display the motivation for this construction. We consider the space of continuous functions on the real line which we display as a projective limit in the following way. For each $n \in \mathbf{N}$, we let $X_n$ denote the space of continuous real-valued functions on the interval $[-n, n]$. and define $\pi_{n+1,n}$ to be the natural restriction mapping. Then the $(X_n)$ form a projective spectrum and we identify its limit as follows: a typical element of the latter is a thread $(x_n)$ of continuous function where $x_n$ is defined on $[-n, n]$. Further these functions are compatible in the sense that if $n < m$, then $x_n$ is just the restriction of $x_m$. Clearly, this implies that the functions can be combined to define a continuous function on the whole line. Conversely, the restrictions of a continuous function on the line to the appropriate intervals defines a thread. Thus we can identify the projective limit with the space of continuous functions on the line.

This example can be generalised to arbitrary topological spaces, in which case we are forced to use directed sets of arbitrary cardinality. We shall return to this topic later.

We now turn to the situation where each of the $X_n$ is a topological space. We then provide the projective limit of the spectrum, which we denote by $\lim X_n$, with a topology as follows. By definition, the limit is a subspace of the Cartesian product and so we simply regard it as a topological space with the induced topology.

We remark that there is a natural mapping from the projective limit $X$ into $X_n$ which we denote by $\pi_n$. It is simply the restriction of the projection from the product into the component $X_n$. Of course, this mapping is continuous and the reader will recognise from the above description of the topology that it is precisely the initial topology induced on $X$ by the family of mappings $(\pi_n)$.

**Examples of projective limits:** I. Intersections: Suppose that $(X_n)$ is a decreasing sequence of subsets of a topological space $X$. Then the intersection $\bigcap X_n$ can be identified in a natural way with the projective limit of the spectrum

$$\{i_{mn} : X_m \to X_n, m \leq n\}$$

(where $i_{mn}$ is the natural inclusion).
II. The following type of space, which is most naturally defined as a projective limit, is of importance in descriptive topology. These are spaces which have a representation as the projective limit of a spectrum $\{\pi_{mn} : X_m \to X_n, n \leq m\}$ where each $X_n$ is countable and discrete (i.e. is homeomorphic to $\mathbf{N}$). Such spaces are then complete metric spaces. In fact, they are homeomorphic to closed subspaces of $\mathbf{N}^{\mathbf{N}}$ by the very definition of the projective limit.

Projective limits can degenerate into triviality as the following examples show.

**Examples:** I. We consider the sequence $([n, \infty[)$ of subsets of the real line. If we regard them as a projective spectrum as in I above, then their projective limit is equal to their intersection, which is, of course, the empty set.,
II. A rather more interesting example is the following. Let $S$ be an uncountable set and consider the system $\{X_A : A \in \mathcal{F}(\mathcal{S})\}$ which is indexed by the family $\mathcal{F}(\mathcal{S})$ of finite subsets of $S$ (of course, this is an uncountable indexing set — it is ordered by inclusion). $X_A$ denotes the family of all injective mappings from $A$ into $\mathbf{N}$. This forms a projective system (with the restriction mapping from $X_B$ into $X_A$ where $A \subset B$ as linking mapping). Since the only possible members of the projective limit are injective mappings from $S$ into $\mathbf{N}$ and there are *no* such mappings, the projective limit is empty. This example is interesting because, in contrast to example I above, the linking mappings are all surjective but the mappings from the projective limit to the components are far from being surjective. We shall see below that this can only happen for indexing sets which are more complicated than $\mathbf{N}$.

It is important in some applications to know *a priori* that a given projective limit is non-empty. We cite here a result that ensures that this will be the case under certain special conditions. Its proof can be regarded as an abstract version of that of the classical theorem of Mittag-Leffler on the existence of meromorphic functions with pre-ascribed principal parts at its poles and for this reason the result is often referred to as the abstract Mittag-Leffler theorem.

**Proposition 49** *Let $(X_n)$ be a projective system of complete metric spaces, whereby the linking mapping $\pi_{n+1,n}$ are supposed to be Lipschitz continuous. If for each $n$ $\pi_{n+1,n}$ is dense in $X_n$, then the image of the projective limit $X$ in $X_n$ under $\pi_n$ is also dense.*

This is a case of a theorem which has no useful generalisation to the case where the index set is uncountable.

For the sake of completeness, we include the much more shallow result referred to above. (It is of purely set-theoretical nature):

**Proposition 50** *If $\{\pi_{mn} : X_m \to X_n, n \leq m\}$ is a projective system of sets indexed by $\mathbf{N}$ whereby the linking mappings $\pi_{mn}$ are all surjective, then the corresponding mappings from the projective limit to each $X_n$ are also surjective.*

We leave the simple proof to the reader.

Dual to the construction of projective limits is that of inductive limits. In this case, we have a sequence $(X_n)$ of topological spaces and, for each $n$, a mapping $i_n$ from $X_n$ into $X_{n+1}$. We can then define, for $m \leq n$ a mapping $i_{mn}$ from $X_m$ into $X_n$. The inductive limit $X$ of this spectrum is defined to be the quotient of the disjoint union $X_n$ under the following equivalence relation: $x \sim y$ if and only if there are $m, n, p$ with $p$ larger than both $m$ and $n$ so that $x$ is in $X_m$, $y$ is in $X_n$ and $i_{mp}(x) = i_{np}(y)$.

The main property of the space constructed in this manner is that a continuous mapping from $X$ into a topological space $Y$ is defined by a sequence $(f_n)$ of continuous mappings, where $f_n$ maps $X_n$ into $Y$ and $f_n \circ i_{mn} = f_m$ whenever $n < m$.

**Examples:** We begin with the following remarks about this construction. In most of the application, the mappings $i_n$ are homeomorphisms from $X_n$ into $X_{n+1}$. Thus we have the case of unions.

We close this chapter by bringing applications of the quotient structure to two branches of geometrical topology — graphs and surfaces.

## 5.5 Graphs:

A graph is a topological space $G$ which is a quotient of a disjoint union $X$ of a finite number of copies of the unit interval under an equivalence relationship of the following simple type. Suppose that the intervals $I_1, \dots, I_n$ have endpoints $(P_i, Q_i)$ and further that we have a partition $S_1, \dots, S_r$ of the finite set $\{P_1, \dots, P_n, Q_1, \dots, Q_n\}$. Then points $x$ and $y$ from $X$ are defined to be equivalent if they are equal or if they are endpoints which belong to the same element $S_k$ of the partition.

**Examples:** see figure 11.

The equivalence class of an endpoint is called a **vertex** of the graph and the copies of the interval are called its **edges**. The **order** of a vertex is defined to be the number of edges which have this vertex as an endpoint i.e. it is the cardinality of the set $S_k$ of the partition to which the endpoint belongs. The vertex is **odd** or **even** according as its order is odd or even. A basic fact about graphs is that the number of odd endpoints is always even.
PROOF. Let $n_i$ denote the number of vertices of order $i$. Then the number of vertices is clearly

$$N = n_1 + n_2 + \dots$$

while the number of odd vertices is

$$N_{\text{odd}} = n_1 + n_3 + \dots$$

Now the number of original endpoints in $X$ (before identification) is

$$N_{\text{tot}} = n_1 + 2n_2 + 3n_3 + \dots$$

Of course, the latter number is even. Now the difference $N_{\text{tot}} - N_{\text{odd}}$ is

$$2n_2 + 2n_3 + 4n_4 + \dots$$

which is also even. Hence $N_{\text{odd}}$ is even.

■

A **path** in a graph is a finite sequence $(v_0, \dots, v_n)$ of vertices which are such that for each $i$ there is an edge from $v_i$ to $v_{i+1}$. $v_0$ is called the **initial point** of the path, $v_n$ the **endpoint**. If these coincide, then the path is **closed**. A **cycle** is a path with $v_0 = v_n$. A **circuit** is a trail with $v_0 = v_n$. A **Hamiltonian path** is one which contains all vertices. A **Hamiltonian cycle** is a cycle which contains all vertices. An **Eulerian circuit** is a circuit which contains all edges and a **Eulerian trail** is one which contains all edges.

Then we have the following result:

**Proposition 51** *A non-trivial connected graph has a Eulerian circuit if and only if each vertex has even degree. A connected graph has an Eulerian trail from a vertex $v$ to a vertex $w$ (whereby these vertices are distinct) if and only if $v$ and $w$ are the only vertices with odd degree.*

**Example:** The question whether the graph (figure ????) has an Eulerian path is the question whether there is a closed walk in the above configuration (figure ???) which involves crossing each bridge exactly once (the Koenigsberg bridge problem).

## 5.6 Surfaces as quotient spaces:

In the first section we defined the torus as the product of two circles. It can also be described as a quotient of a square in the following simple manner. Consider the following equivalence relationship on $\mathbf{I}^2$.

$$\begin{aligned}(s,t) \sim (s',t') \text{if and only if} \quad & s = s' \text{and} t = t' \text{or} \\ & s = 0, s' = 1 \text{and} t = t' \text{or} \\ & s = 1, s' = 0 \text{and} t = t' \text{or} \\ & s = s', t = 0 \text{and} t' = 1 \text{or} \\ & s = s', t = 1 \text{and} t' = 0.\end{aligned}$$

Instead of this hopelessly unwieldy description, the above relationship can be displayed graphically as in figure (13). Here the fact that opposite sides are labelled with the same letter means that they are to be identified, while the arrows indicate that points are to be identified with their mirror images in the lines parallel to the appropriate sides which bisect the square. In the following examples, we shall use this more informal method of describing such equivalence relationships.

**The cylinder:** This is the quotient of $\mathbf{I}^2$ displayed in figure 14.

**The Möbius band:** Figure 15. Here the fact that the arrows point in different directions indicates that the points on the appropriate sides are identified with their mirror images in the centre.

**The Klein bottle:** Figure 16

Note that in the above representation of the Klein bottle as a subset of three-dimensional we have cheated in the sense that the bottle is self-intersecting. This is unavoidable since there is no subset of $\mathbf{R}^3$ which is homeomorphic to the Klein bottle. One has to go into the fourth dimension to obtain a representation of it.

**The projective plane:** Figure 17

As in the case of the Klein bottle, the projective plane cannot be realised in space without introducing a self-intersection. We can visualise it as a "sphere with cross-cap".

We now introduce a convenient algebraic notation for describing such representations of surfaces as quotients of squares. Consider the concrete case of the torus with the above representation (figure 13).

We denote this by the symbol $ab^{-1}a^{-1}b$ which is obtained as follows. We start at an arbitrary side (in this case the top $a$) and traverse the circumference of the square (say in the clockwise direction, although this does not matter), writing down successively the symbols for the sides. If the arrow on a side points in the direction in which we are travelling, we write down the letter of the alphabet with which it is labelled. If the arrow points backwards, we add the index $-1$. Thus the Möbius band has symbol ??? the Klein bottle ????? and the projective plane ?????

Exactly the same method can be used if we replace the square by *any* regular polygon. This allows us to construct more intricate surfaces. For example, the **handle** is the quotient of the regular pentagon indicated in figure 18. It has symbol $acb^{-1}a^{-1}b$. (Here the letter $c$ occurs only once to indicate that this side of the pentagon is not identified with any other one. It then forms part of the boundary of the resulting surface. By the way, we are now using the word "boundary" in its everyday sense, not in the technical sense introduced in I.3)

With this convention any string of symbols of the form

$$a_1^{\epsilon_1} \dots a_n^{\epsilon_n}$$

where the indices $\epsilon_i$ are either 1 (in which case we do not reproduce it) or $-1$ represents a surface, provided that each letter appears at most twice. The surface is uniquely determined by this string of symbols (although a given surface can have several such representations, depending, for example, on the edge at which one starts).

Using the symbolic representation, we can distinguish between certain types of surfaces. A surface is **closed** or **without boundary** if each letter appears exactly twice. Otherwise, we have a surface **with boundary**. For example, the torus, the Klein bottle and the projective plane are closed surfaces, whereas the Möbius band and the cylinder are with boundary.

We also distinguish between **two-sided** (or oriented) surfaces and **one-sided** (or non-oriented) surfaces. Examples of the former are the torus and the cylinder and of the latter the Möbius band, the Klein bottle and the projective plane. The nomenclature one- or two-sided is self-explanatory (try to find a second side on a Möbius band!). The term oriented or non-oriented comes from the fact that on a surface like the Möbius it is possible to change the orientation of a coordinate frame by traversing the surface. This is impossible to do on an oriented surface such as a cylinder or the sphere.

One can determine whether a surface is orientable or not from its algebraic symbol as follows. The surface is non-orientable if one symbol occurs twice with the same index. Otherwise it is orientable.

Representations of surfaces as quotients of polygons can be used to decide the results of simple experiments involving cutting up surfaces which often crop as party games. Consider, for example, the well-known parlour trick of cutting a Möbius band along a central line. This can be carried out without scissors and paper as in figure 19. Hence the result is a cylinder. In fact, if actually carried out with a real Möbius band, what one obtains is a Möbius band with two twists. This is homeomorphic to the cylinder (but is not embedded in space in the same way).

As a further example, consider figure 20 which shows that the projective plane is obtained by "closing" the Möbius band with a disc. In other words, if we cut a disc out of the projective plane, we get a Möbius band.

We can construct new surfaces by means of the exotically named construction of the **smash product**. This is defined as follows. We cut discs out of each of the surfaces $X$ and $Y$ and paste them together along the edges of the wholes (which are homeomorphic to $\mathbf{S}^1$). (see figure 21) The resulting manifold is denoted by $X\sharp Y$. As in the case of the wedge product, for those surfaces which we consider the resulting surface is independent of exactly where and how we cut out these discs.

For example, the smash product of two copies of the sphere is again the sphere. More interestingly, the smash product of two tori is the **Bretzel** (figure 23)

We can determine the algebraic symbol of a smash product from the symbols of the components. We do this for the Bretzel as the smash product of two tori but the method is completely general.

The smash product of two projective planes is the Klein bottle. This is rather difficult to visualise directly, but we can also see it formal manipulations with the polygon representations as follows. We begin with two projective planes as in figure 24.

The spaces which we have been discussing are examples of what are called two-dimensional manifolds (or just 2-manifolds). The fact that they are two-dimensional can be expressed mathematically by noting that each point on them has a neighbourhood which is homeomorphic either to

a) the open disc in $\mathbf{R}^2$; or

b) half of this disc (together with the bounding diameter. (figure 25) The handle, for example, has points of both types (figure 26). Points of the second type are just those on the boundary as described above i.e. they lie on edges of the polygon which are not identified with a second edge.

The surfaces which we obtain by our constructions have two topological properties which will be discussed in detail in later chapters. They are compact and connected. For our present purposes, the first conditions means that they are homeomorphic to closed, bounded sets in some euclidean space. The second means that each pair of points can be joined by a continuous curve *which lies on the surface* (figure 27). The first condition excludes surfaces such as that in figure 28, the second surfaces which are composed of several pieces.

Using the concept of smash products, we can describe *all* compact, connected surfaces as follows:

a) each oriented compact, connected surface is homeomorphic the sphere or a smash product of tori;

b) each non-oriented compact, connected surface is a smash product of a finite number of tori with a copy of the projective plane or of the Klein bottle. (This is a deep result of topology which we shall not prove here).

For this reason it is convenient to introduce the following notations:

$nT$ denotes a space which is homeomorphic to the smash product of $n$ tori;

$(nT, P)$ is a smash product of $n$ tori with a copy of the projective plane;

$(nT, K)$ is the smash product of $n$ tori with a copy of the Klein bottle.

**The Euler characteristic:** The Euler characteristic of a surface is an example of a so-called **topological invariant** and is calculated as follows. Consider a closed surface, such as a sphere. On the sphere we draw a network as in figure 29 and calculate the number $V - E + F$ where $V$ denotes the number of vertices, $E$ the number of edges and $F$ the number of faces. It turns out that for a given surface (up to homeomorphism) this number is invariant i.e. independent of the network. It is called the Euler characteristic of the surface and denoted by the symbol $\chi(X)$. Such topological invariants will be discussed in some detail in Chapter III. Here we note that it can be calculated in terms of the surface's polygonal representation as the number $m - n + 1$ where the surface is represented by a $2n$-gon (the number of faces must be even since we are considering *closed* surfaces) and $m$ is the number of *distinct* points represented by the vertices of the polygon.

**Examples:** The method of representing the smash product of two surfaces can be used to show that we have the following formula for the Euler characteristic of $X\sharp Y$:

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

Using this formula, we can calculate very simply the characteristics of the surfaces listed above. We have the following result:

$\chi(n, T) = 2 - 2n$;

$\chi(nT, K) = -2n$;

$\chi(nT, P) = 1 - 2n$. From these we can deduce the following criterium for the equivalence of compact, connected surfaces. Two surfaces $X$ and $Y$ are homeomorphic if and only if

a) they have the same orientability properties. and

b) they have the same Euler characteristic.

# 6 Connectedness

In the introductory chapter on geometrical topology, we noted that the topological difference between the unit interval $I$ and the circle $S^1$ could be pinpointed by using the fact that the removal of a point from $I$ splits the space into two parts (provided that we do not remove an endpoint). We shall study the corresponding topological notion in this chapter.

**Definition:** A topological space $(X, \tau)$ is **connected** if it has no representation $X = A \cup B$ where $A$ and $B$ are open and disjoint. An equivalent condition is that the only subsets of $X$ with are clopen are $X$ and $\emptyset$. More generally, a subset $A$ of $X$ is connected (in $X$) if it is connected in the induced topology. This means that if $U$ and $V$ are disjoint open subsets of $X$ whose union contains $A$, then either $A \subset U$ or $A \subset V$.

For example, a space with the indiscrete topology is connected (since $X$ and  are the only *open* sets). On the contrary, no space with the discrete topology is connected (with the trivial exceptions of the empty set or a one-point set). An infinite set with the co-finite topology is connected. The set of irrationals (with the natural topology) or the real line with the Sorgenfrey topology are not connected. For example, if $\alpha$ is an irrational number then

$$]-\infty, \alpha[ \cap \mathbf{Q} = ]-\infty, \alpha] \cap \mathbf{Q}$$

is clopen in $\mathbf{Q}$. On the other hand, $\mathbf{R}$ with the natural topology, is connected. In fact, the connected subsets of the real line are just the intervals as we shall see shortly. We begin with a simple Lemma:

**Proposition 52** *Lemma Suppose that $A$ is a connected subset of a topological space $X$. Then the closure $\bar{A}$ of $A$ is also connected. (More generally, any set which lies between $A$ and $\bar{A}$ is connected).*

PROOF. We suppose that $\bar{A}$ is contained in the union $U \cup V$ of two disjoint, open subsets of $X$. Then since $A$ is connected, we have either $A \subset U$ or $A \subset V$. In the first case, $A$ is a subset of the closed set $X \backslash V$ and hence so is $\bar{A}$. i.e. $\bar{A} \subset U$.

■

**Proposition 53** *A subset of the real line is connected if and only if it is an interval.*

PROOF. We show firstly that intervals are connected, beginning with the line itself. Let $U$ be a clopen subset of $\mathbf{R}$, which we suppose to be neither the whole space or the empty set. Then it is the disjoint union of at most countably many disjoint intervals. (This is a standard result from an elementary Analysis course). At least one of the endpoints of these intervals is finite (otherwise $U$ is $\mathbf{R}$). Such an endpoint is clearly in the closure of $U$ but not in $U$ which contradicts the fact that $U$ is closed.

■

Once we know that the line is connected it follows that any open interval has the same property since it is homeomorphic to $\mathbf{R}$ and connectedness is clearly a topological property. We can then deduce the connectedness of any interval, since each interval is the closure (in itself) of its interior (in $\mathbf{R}$).

We now show that the only connected subsets of the line are intervals. Suppose that $A$ is connected and put

$$\beta = \sup \{x : x \in A\}, \quad \alpha = \inf \{x : x \in A\}$$

(of course, these can be infinite). Now use the simple remark that if $a$ and $b$ are in $A$, then the interval $[a, b]$ is a subset of $A$ (for if $\xi \in ]a, b[$ were not in $A$, then the disjoint sets $U = ]-\infty, \xi[$ and $V = ]\xi, \infty[$ cover $A$). In order to be concrete, we suppose that $\alpha$ and $\beta$ are not in $A$ and that both are finite. Then we claim that $A$ is the open interval $]\alpha, \beta[$. For if $x \in ]\alpha, \beta[$, there are $a, b$ in $A$ with $a \leq x \leq b$ and so $x \in A$. Hence $]\alpha, \beta[ \subset A$ and so is equal to $A$. (For the closed interval contains $A$ and neither endpoint is in $A$). The remaining cases are dealt with in a similar fashion.

In order to extend our list of examples of connected spaces, we establish some simple stability properties. Firstly, it is clear that the union of connected sets need not be connected. However, if we can pin them down with a fixed set as in the next proposition, then we do obtain connectedness:

**Proposition 54** *Let $A$ be a non-empty connected subset of $X$ and $\mathcal{A}$ a family of connected subsets, each of which intersects $A$. Then $A \cup (\bigcup \mathcal{A})$ is connected.*

PROOF. Suppose that $U$ and $V$ are disjoint open sets whose union covers the relevant set. Then it also covers $A$ and so we can assume, without loss of generality, that $A$ is contained in $U$ and disjoint from $V$. Now each $B \in \mathcal{A}$ must, for the same reason, also lie either in $U$ or in $V$. But the fact that $B$ intersects $A$ rules out the second possibility. Hence all of the relevant sets are contained in $U$ and therefore so is their union.

■

In most applications, $A$ is a singleton. For example, we can deduce immediately from this result that $\mathbf{R}^n$ is connected as the union of the family of all lines which pass through the origin. More generally we can prove that products of connected spaces are connected, but in this case we have to use the general form of the result. We prove this for the product $X \times Y$ of two spaces but the result holds for arbitrary products. Without loss of generality we can assume that $X$ and $Y$ are non-empty (the case where one of them is empty is trivially true). Suppose then that $b$ is an element of $Y$. The subset $A = X \times \{b\}$ of the product is homeomorphic to $X$ and so is connected. Now $X \times Y$ is the union of $A$ and the sets of the form $\{x\} \times Y$ $(x \in X)$ and the conditions of the above result are fulfilled.

We remark that it follows from a result which we shall prove below (Proposition ???) that the converse is valid i.e. a product can only be connected if each component is connected. (However, in this case we must assume that all of the spaces are non-empty).

A further consequence of the above result is that if $X$ has the property that for any pair $x, y$ of points therein, there exists a connected subset which contains $x$ and $y$, then $X$ is connected., For if we fix $x$ and choose for each $y$ in $X$ a connected set $C_y$ which contains both $x$ and $y$, then $X = \bigcup_{y \in X} C_y$ is connected. This implies, for example, that convex subsets of $\mathbf{R}^n$ are connected. For any two points in such a set lie on the segment joining them and this is connected (being homeomorphic to an interval).

Another consequence of the Proposition is the following: suppose that we have a sequence $(A_n)$ of connected subsets so that for each $n$, $A_n \cap A_{n+1}$ is non-empty. Then their union is connected.

If $x$ is a point in a topological space, then we define the **component** $C(x)$ of $x$ to be the union of those connected subsets which contain $x$. Note that this is connected and so is the *largest* connected set containing $x$. It is closed since its closure is connected and so cannot be larger.

A further simple stability result is the

**Proposition 55** *Suppose that $Y$ is the continuous image of a connected space. Then $Y$ is also connected.*

PROOF. The hypothesis means that there is a connected space $X$ and a continuous $f : X \to Y$ with $f(X) = Y$. Suppose then that $Y = U \cup V$ where $U$ and $V$ are open and disjoint. Then $X = f^{-1}(U) \cup f^{-1}(V)$ and these pre-images are clearly open and disjoint. Hence one of the latter sets, say the first one, is the whole of $X$. Then $U = Y$ since $f$ is surjective.

■

If we apply this result to a real-valued, continuous function on a connected space $X$, we see that its image $f(X)$ must also be connected and so is an interval. This can be restated as the following abstract version of the intermediate-value theorem from elementary calculus: if $x$ and $y$ are in $X$ and $c$ is a real number which lies between the values of $f$ at $x$ and $y$, then there is a $z$ in $X$ for which $f(z) = c$.

We remark further that it follows from the above proposition that a quotient of a connected space is itself connected. Thus the torus, Klein bottle and the Möbius band (more generally, all of the surfaces which we described as quotients of polygons) are connected.

We now consider some variants of the concept of connectedness:

**Definition:** A space $X$ is **locally connected** if each point has a neighbourhood basis consisting of open, connected sets. The properties of being connected or locally connected are not comparable, as the following examples show:

**Examples:** Firstly it is easy to find a space which is locally connected but not connected. For example, a disjoint union of two copies of the interval is locally connected, but not connected. The following is an example of a space which is connected, but not locally connected. it is called the topologist's **sine curve**. One takes the subset of the plane which consists of the closure of the graph of the function $x \mapsto \sin \frac{1}{x}$ (defined on the interval $]0, 1[$). (in other words, it is the union of this graph with the interval $[-1, 1]$ in the $y$-axis). This space is connected (since the above graph is homeomorphic to the real line. Hence so is its closure). However, a glance at figure ??? shows that it is not locally connected. We remark that if $X$ is locally connected, then $C(x)$ is open for each $x$ (and hence clopen). The converse is also true i.e. if $C(x)$ is always open, then $X$ is locally connected.

**Definition:** $X$ is **arcwise connected** if any two points can be connected by a continuous curve i.e. for each $x$,$y$ in $X$, there is a continuous function

$c : [0,1] \to X$ with $c(0) = x, c(1) = y$. It follows from the criterium for connectedness in ??? that this implies connectedness, but the converse is not true. For example, the topologist's sine curve is connected but not arcwise connected since it is impossible to joint any point on the graph of $\sin \frac{1}{x}$ to the point $(0,1)$ say.

There is also a local version of the above property, called **local arcwise connectedness**. This means that each point has an open neighbourhood basis consisting of arcwise connected sets. The following is an example of a space which is arcwise connected, but not locally arcwise connected.

**Example:** Consider the subset of the plane indicated in figure ??? i.e. the union of the collection of segments from $(0,1)$ to the points $\{0, 1, \frac{1}{2}, \frac{1}{3}, \dots\}$ on the $x$-axis. Then this is arcwise connected, but not locally arcwise connected).

Typical examples of locally arcwise connected spaces are open subsets of $\mathbf{R}^n$. For such spaces, the difference between connectedness and arcwise connectedness vanishes:

**Proposition 56** *If $X$ is locally arcwise connected, then it is connected if and only if it is arcwise connected.*

PROOF. We need only show that if $X$ is connected and locally arcwise connected, then it is arcwise connected. In order to do this, we fix a point $x$ in $X$ and define $U$ to be the set of points which can be joined to $x$ by an arc in $X$, resp. $V$ to be the complement of $U$. We wish to show that $X = U$. Since $X$ is connected and $U$ is non-empty ($x$ is a member), it suffices to show that $U$ and $V$ are open. But this follows from the fact that $X$ is locally arcwise connected.

■

We remark that this result implies that for open subsets of $\mathbf{R}^n$, the concepts of connectedness and arcwise connectedness coincide.

Paradoxical as it may seem, there are situations where *bad* connectedness properties are useful. We discuss some of these briefly:

**Definition:** A topological space $X$ is **totally disconnected** if for each $x \in X$, $C(x) = \{x\}$ (i.e. the only connected subsets of $X$ are the singletons. $X$ is **zero-dimensional** if it has a basis of clopen sets. Finally it is **extremally disconnected** if the closure of each open set is open.

Examples of totally disconnected spaces are

a) any set with the discrete topology;

b) $\mathbf{Q}$ with the usual topology;

c) the Sorgenfrey line.

Products of totally disconnected spaces are easily seen to be totally disconnected. Hence the Cantor set, for example, is totally disconnected.

Since intervals with irrational endpoints are both open and closed, $\mathbf{Q}$ is zero-dimensional.

Any discrete space is extremally disconnected. We shall give less trivial examples later.

We shall see later that there is a close connection between the concepts of total disconnectedness and zero dimensionality. Nevertheless, the two concepts are distinct. For instance, we give an example of a totally-disconnected metric space which is not zero-dimensional.

**Example:** Consider the Cantor set as a subset of $[0,1]$, together with the auxiliary point $A = (\frac{1}{2}, \frac{1}{2})$ in the plane. For each $x$ in the Cantor set, we write $L_x$ for the line from $x$ to $A$ and we define $M_x$ to be the set of points in $L_x$ whose $y$- coordinate is rational provided that $x$ is an endpoint of one of the intervals eliminated in the construction of the Cantor set, resp. to he the set of points in $L_x$ whose $y$-coordinate is irrational otherwise. Then if $X$ is defined to be the union of the $M_x$ ($x$ from the Cantor set minus $A$), it is totally disconnected, but not zero-dimensional.

Disjoint unions of totally disconnected spaces, resp. products and projective limits of such spaces are also totally disconnected. The property of being zero-dimensional enjoys the same stability properties.

# 7 Separation properties

As we saw above, in a general topological space a sequence can have more than one limit. We now discuss a number of conditions (which are known as **separation** properties) which ensure that such pathological behaviour cannot occur. Topologies are very general structures and this generality is paid for by the lack of depth of the results which can be proved and by the number of pathologies which can arise. The conditions which we shall introduce in this chapter can be seen as a means of reducing the possible types of space which we consider in order to arrive at a more coherent and richer theory.

We start with three simple separation axioms. In fact, for our purposes only the third will be of any importance but we bring the first two for the sake of completeness. **Definition:** A topological space $(X, \tau)$ is

- $T_0$ if whenever $x$ and $y$ are distinct points, then there is an $N \in \mathcal{N}(\S)$ with $y \notin N$ or an $N \in \mathcal{N}(\dagger)$ with $x \notin N$; $T_1$ if whenever $x$ and $y$ are distinct points, there is an $N \in \mathcal{N}(\S)$ with $y \notin N$; $T_2$ if whenever $x$ and $y$ are distinct points there is an $N_1 \in \mathcal{N}(\S)$ and an $N_2 \in \mathcal{N}(\S)$ with $N_1 \cap N_2 = \emptyset$.

(See figure 1). Of course, these three conditions are in increasing order of restrictiveness.

The following examples show that they are in fact distinct. An indiscrete space (with more than two elements) is not $T_0$. The Sierpinski space is $T_0$ but not $T_1$. If $X$ is an infinite set, then $(X, \tau_{cf})$ is $T_1$ but not $T_2$. Finally, any metric space is $T_2$.

Another trivial but useful fact is that if $X$ is a topological space with one of the above $T$-properties, then any finer topology on $X$ also possesses it. They are also preserved by subspaces and products. For example to see that the product of a family of $T_2$ spaces is again $T_2$, we proceed as follows. if $x = (x_\alpha)$ and $y = (y_\alpha)$ are distinct elements of the product $\prod X_\alpha$, then there is a $\beta \in A$ with $x_\beta \neq y_\beta$. Hence there are disjoint $N_1 \in \mathcal{N}(\S_\beta)$ and $N_2 \in \mathcal{N}(\dagger_\beta)$. Then $\tilde{N}_1 = (\prod_{\alpha \neq \beta} X_\alpha) \times N_1$ resp. $\tilde{N}_2 = (\prod_{\alpha \neq \beta} X_\alpha) \times N_2$ resp. are the required neighbourhoods.

The $T_1$-property can be usefully characterised in the following ways:

**Proposition 57** *Let $X$ be a topological space. Then the following are equivalent: a) $X$ is $T_1$; b) for each $x \in X$, the singleton $\{x\}$ is closed; c) $\tau$ is finer than $\tau_{cf}$; d) for each $x \in X$, $\{x\} = \bigcap_{N \in \mathcal{N}(\S)} N$.*

These are simple reformulations of the definition. It follows immediately from b) or c) above that there is precisely one $T_1$ topology on a finite set – the discrete topology.

If $x$ is an element of the $T_1$-space $X$ and $y_1, \dots, y_n$ are points of $X$ which are distinct from $x$, then it is clear that there is a neighbourhood $N$ of $x$ which does not contain any of the $y$'s. From this it follows that if $x$ is a cluster point of a subset $M$ of a $T_1$-space, then every neighbourhood of $x$ contains infinitely many points of $M$. For suppose if possible that $U$ is a neighbourhood of $x$ which contains only finitely many points of $M$ – say $\{y_1, \dots, y_n\}$. Then there is a neighbourhood $V$ of $x$ which fails to contain any of the $y$'s. $U \cap V$ is a neighbourhood of $x$ and so contains a point in $M$ which is distinct from $x$. This is a contradiction.

We now turn to the $T_2$-spaces and show that these are precisely those ones in which convergent filters have unique limits. Almost all of the interesting topological spaces have this property.

**Proposition 58** *The following conditions on a topological space $X$ are equivalent: a) $X$ is $T_2$; b) for each $x$ in $X$, $\{x\} = \bigcap_{N \in \mathcal{N}(\S)} \bar{N}$; c) if a filter $\mathcal{F}$ on*

*X converges simultaneously to two points $x$ and $y$, then $x$ and $y$ coincide; d) the diagonal set $\Delta = \{(x,x) : x \in X\}$ is closed in the product $X \times X$.*

PROOF. Once again, these are all simple manipulations of the definition. We prove the equivalence of a) and c). a) implies c): Suppose that $\mathcal{F} \to \S$ and $\mathcal{F} \to \dagger$. If these two points were distinct, there would be disjoint neighbourhoods $N_1$ of $x$ and $N_2$ of $y$. But this contradicts the fact that both of these sets are in $\mathcal{F}$. c) implies a): Suppose that $X$ is not $T_2$. Then there are distinct points $x$ and $y$ so that $N_1 \cap N_2 \neq \emptyset$ for each pair $N_1$, $N_2$ where the first set is a neighbourhood of $x$ and the second one a neighbourhood of $y$. Then the union of the families $\mathcal{N}(\S)$ and $\mathcal{N}(\dagger)$ forms a filter basis and the filter which it generates clearly converges to both $x$ and $y$.

■

Condition d) above implies the following simple fact which is the basis of many uniqueness proofs: if two continuous functions agree on a dense set, then they agree *everywhere* (i.e. are identical functions). We remark that it is rather easy to forget that the range space must satisfy the $T_2$-condition for it to hold):

**Proposition 59** *Let $f$ and $g$ be continuous mappings from a space $X$ into a $T_2$-space $Y$. Then the set $\{x \in X : f(x) = g(x)\}$ where $f$ and $g$ coincide is closed. In particular, if $f$ and $g$ coincide on a dense subset of $X$, then they are equal.*

PROOF. The above set is the pre-image of the diagonal set in $Y \times Y$ (which we know to be closed) under the continuous mapping $x \mapsto (f(x), g(x))$ from $X$ into $Y \times Y$ and hence is closed.

■

A typical example of an application of this theorem is the following: suppose that we have a projective spectrum $\{\pi_{m,n} : X_m \to X_n, n \leq m\}$ of topological spaces. Then, by the very definition, the projective limit is homeomorphic to a subspace of the product $\prod X_n$. Then we claim that in the case where the $X_n$ are $T_2$-spaces, then it is actually homeomorphic to a *closed* subspace of the product. This is because the image of the projective limit under the homeomorphism is the set $\bigcap_{n \leq m} V_{mn}$ where

$$V_{mn} = \{(x_n) \in \prod X_n : \pi_{mn}(x_m) = x_n\}$$

and this is the coincidence set of two continuous functions with values in the $T_2$-space $X_n$.

## 7.1 Further separation axioms:

We now discuss some more sophisticated separation axioms.

**Definition:** A topological space $(X, \tau)$ is
**regular** if whenever $x \in X$ and $A$ is a closed subset of $X$ which does not contain $x$, there are disjoint open sets $U$ and $V$ with $x$ in $U$ and $A \subset V$;
**completely regular** if whenever $x$ and $A$ are as above, there is a continuous function $f$ from $X$ into $[0,1]$ so that $f(x) = 1$ and $f = 0$ on $A$;
**normal** if whenever $A$ and $B$ are disjoint, closed subsets of $X$, there are open, disjoint sets $U$ and $V$ so that $A \subset U$ and $B \subset V$.

Further we say that $X$ is $T_3$ if it is $T_1$ and regular, $T_{3\frac{1}{2}}$ if it is $T_1$ and completely regular and $T_4$ if it is $T_1$ and normal. As the notation implies, these conditions become successively more restrictive. Indeed it is trivial that a $T_{3\frac{1}{2}}$-space is $T_3$. The implication: $T_4$ implies $T_{3\frac{1}{2}}$ is true but less trivial and will be proved below. We remark that no such simple implications exist between the properties of being regular, completely regular and normal. For example, a set $X$ with the indiscrete topology is normal but not even $T_0$ (if $X$ has at least two points). The Sierpinski space is normal but not completely regular (it is normal for the rather trivial reason that there *are* no disjoint pairs of closed sets other than the obvious one $(X, \emptyset)$).

It is clear that $T_3$ implies $T_2$ and that complete regularity implies regularity. We now bring an example of a $T_2$-space which is not $T_3$. Consider the real line with the following topology. We write $K$ for the set $\{\frac{1}{n} : n \in \mathbf{Z} \setminus \{0\}\}$. Then $\tau$ is the topology generated by the sets

$$\{]x - \frac{1}{n}, x + \frac{1}{n}[\}_{x \neq 0, n \in \mathbf{N}} \cup \{] = \frac{1}{n}, \frac{1}{n}[\setminus K\}_{n \in \mathbf{N}}.$$

This is $T_2$ since it is finer than the natural topology but it is not $T_3$ since the closed set $K$ cannot be separated from the origin.

In order to simplify the statements of the next results we introduce some more terminology. We say that two subsets $A$ and $B$ of a topological space are
**weakly separated** if there exist disjoint open sets $U$ and $V$ with $A \subset U$ and $B \subset V$;
**strongly separated** if there is a continuous function $f$ from $X$ into $[0,1]$ which vanishes on $A$ and takes on the value 1 on $B$.

Of course, if $A$ and $B$ are strongly separated (by $f$), they are weakly separated (take $U = f^{-1}([0, \frac{1}{2}[)$ and $V = f^{-1}(]\frac{1}{2}, 1])$).

We can then formulate the separation axioms as follows:

$X$ is $T_2$ if distinct points can be weakly separated, regular if closed sets can be weakly separated from disjoint points, completely regular if closed sets can be weakly separated from disjoint points and normal if disjoint closed sets can be weakly separated.

In the spirit of the reformulation of the lower $T$-conditions, we can recast the definitions of regular and completely regular spaces as follows:

**Proposition 60** *A topological space $X$ is regular if and only if every neighbourhood of a given point $x$ of $X$ contains a closed neighbourhood (in other words, the collection of closed neighbourhoods of $x$ is a basis for $\mathcal{N}(\S)$).*

PROOF. Suppose that $X$ is regular and that $U$ is an open neighbourhood of $x$. Then $X \backslash U$ is closed and so we can find disjoint sets $W$ and $V$ with $x$ contained in $V$ and $X \backslash U \subset W$. Then $\bar{V} \subset U$.

■

In order to formulate the next proposition, we introduce some notation. If $X$ is a topological space, then $C(X)$ denotes the space of continuous, real-valued functions on $X$. If $x \in C(X)$, put

$$Z(f) = \{x \in X : f(x) = 0\}$$
$$C(f) = \{x \in X : f(x) \neq 0\}$$

(the **zero-set** and **cozero-set** of $f$ respectively). Note that it follows from the definition of initial topologies that the sets $\{C(f) : f \in C(X)\}$ form a basis for the initial topology on $X$ generated by the functions $f$ in $C(X)$.

**Proposition 61** *$X$ is completely regular if and only if its topology coincides with the initial topology induced by $C(X)$ (i.e. if and only if the sets $\{C(f) : f \in C(X)\}$ generate the topology).*

Once again, this is just a restatement of the definition.

In the same way we can recast the definition of normality. $X$ is normal if and only if whenever a closed set $C$ is contained in an open set $U$, there is an open set $V$ with $C \subset V \subset \bar{V} \subset U$.

Our main result will be the statement that in the definition of $T_4$-spaces the notion of weak separation can be relaced by that of strong separation. In the proof , we shall use some simple facts about rational numbers which we now recall. A **dyadic rational** is one of the form $k.2^{-n}$ with $k \in \mathbf{Z}$ and $n \in \mathbf{N}$. We denote the set of such numbers by $\mathbf{Q}_d$. Then of course, $\mathbf{Q}_d$ is dense in the real numbers and the set of half-intervals of the form $]x, \infty[$ resp. $]-\infty, y[$ $(x, y \in \mathbf{Q}_d)$ is a subbasis for the natural topology on the real line.

**Proposition 62** *(Urysohn's Lemma) Let $A$ and $B$ be closed, disjoint subsets of a normal space. Then $A$ and $B$ are strongly separated.*

Proof. For each dyadic rational $r = \frac{k}{2^n}$, we define an open set $U(r)$ so that these have the following properties:

$$A \subset U(0) \subset U(r) \subset U(1) \subset X \backslash B$$

for each $r$ and furthermore

$$\overline{U(r)} \subset U(s)$$

whenever $r < s$. We shall show how to do this below. Before doing so, we proceed with the proof by defining the function $f$ as follows:

$$f(x) = \sup_r \{x \in U(r)\}.$$

Then $f = 0$ on $A$ and $f = 1$ on $B$. Also $f$ is continuous since $\{f(x) < a\} = \bigcup_{r<a} U(r)$ and $\{f(x) > a\} = \bigcup_{r>a} X \backslash \overline{U(r)}$ and both are open.

We now turn to the construction of the $U$'s where we proceed as follows. We first choose an open set $U(\frac{1}{2})$ so that

$$A \subset U(\frac{1}{2}) \subset \overline{U(\frac{1}{2})} \subset X \backslash B.$$

We then choose $U(\frac{1}{4})$ and $U(\frac{3}{4})$ so that

$$A \subset U(\frac{1}{4}) \subset \overline{U(\frac{1}{4})} \subset U(\frac{1}{2}) \subset \overline{U(\frac{1}{2})} \subset U(\frac{3}{4}) \subset \overline{U(\frac{3}{4})} \subset X \backslash B$$

and continue in the obvious way.

■

It follows immediately from this Lemma that a $T_4$-space is $T_{3\frac{1}{2}}$.

We can refine the above proof to get the following result which is often useful.

**Proposition 63** *Suppose that $A$ and $B$ are subsets of a normal space $X$, whereby $A$ is an $F_\sigma$ and $B$ is a $G_\delta$. Suppose further that $\bar{A}$ is contained in $B$ and the interior of $B$ contains $A$. Then there is an open $F_\sigma$-set $W$ so that $A \subset W \subset \bar{W} \subset B$.*

Proof. Suppose that $A$ is the union of the sequence

■

Using this separation theorem, we can prove the following result which contains Urysohn's Lemma and a second deep result – the Tietze extension theorem – as special cases.

**Proposition 64** *Let $X$ be a normal topological space and let $g$ and $h$ be real-valued functions on $X$ whereby $g$ is u.s.c., $h$ is l.s.c. and $g \leq h$. Then there is a continuous function $f$ on $X$ which lies between $g$ and $h$.*

Notice that if we apply this result with $g = 1 - \chi_A$ and $f = \chi_B$ in the situation of Urysohn's Lemma, then we obtain another proof of the latter. In addition, it immediately implies the following result:

**Proposition 65** *(The Tietze extension theorem) Let $C$ be a closed subset of the normal space $X$ and let $f$ be a continuous function on $C$ with values in $[-1, 1]$. Then there is a continuous extension $\tilde{f}$ of $f$ to a function from $X$ into $[-1, 1]$.*

Of course there is nothing sacred about the interval $[-1, 1]$ and the result applies to any bounded, continuous real-valued function on $C$.

We now consider hereditary behaviour of the separation properties.

It is clear that *any* subset of a regular (resp. completely regular) space has the same property. Hence the $T_3$- and $T_{3\frac{1}{2}}$-properties are hereditary. *Closed* subspaces of $T_4$ spaces are clearly also $T_4$ but an arbitrary subspace of a normal space need not be normal as the following example shows.

As regards products, we have the following result: products of regular and completely regular spaces have the same properties. Hence this is also true of $T_3$- and $T_{3\frac{1}{2}}$-spaces.

Examples of $T_4$-spaces are provided by metric spaces. For if $A$ is a subspace of a metric space, the mapping

$$d_A : x \mapsto \inf\{d(x, y) : y \in A\}$$

is continuous (in fact, Lipschitz continuous with constant 1). Also $x \in \bar{A}$ if and only if $d(x, A) = 0$. Hence if $A$ and $B$ are disjoint, closed sets, then the mapping

$$f : x \mapsto \frac{d_A(x)}{d_A(x) + d_B(x)}$$

is well-defined, continuous and separates $A$ and $B$. The same argument shows that pseudo-metric spaces are normal. We conclude this section with a so-called embedding theorem. This characterises $T_{3\frac{1}{2}}$-spaces as those which can be embedded into products of the real line.

**Proposition 66** *A topological space $(X, \tau)$ is $T_{3\frac{1}{2}}$ if and only if it is homeomorphic to a subset of a product $\mathbf{R}^A$ for some index set $A$.*

Proof. We note first that the condition is sufficient since **R** is $T_{3\frac{1}{2}}$ and this property is preserved by products and subspaces. Conversely,

■

We remark that since $I$ is a subset of the real line while the latter is homeomorphic to a subset of $I$ (for instance, the *open* unit interval), we can replace **R** by $I$ in the above result i.e. $T_{3\frac{1}{2}}$-spaces can be characterised as subspaces of products of the unit interval.

A further useful remark is that if the $T_{3\frac{1}{2}}$ is countably generated, then one can refine the above proof to show that we can embed $X$ into the Hilbert cube (i.e. we can take **N** as the indexing set).

Once again, the converse to this result is trivially true — subspaces of the Hilbert cube are countably generated. Later we shall see that a countably generated $T_3$-space is automatically $T_{3\frac{1}{2}}$ (even $T_4$) so that we can relax the separation condition to $T_3$. This provides the following purely internal topological characterisation of separable metrisable spaces:

**Proposition 67** *A $T_3$-space is separable metrisable if and only if it is regular and countably generated. It is then homeomorphic to a subspace of the Hilbert cube.*

## 7.2 Still more separation properties:

The list of separations can be extended to still higher forms. We conclude with some brief remarks on the $T_5$- and $T_6$-properties.

As we have seen above, not every subspace of a normal space is normal. A normal space is defined to be **hereditarily normal** if each subspace is normal. (In fact, it suffices to demand that each open subspace be normal). This can be reformulated as follows: if $A$ and $B$ are disjoint subsets of $X$ so that

$$A \cap \bar{B} = \bar{A} \cap B = \emptyset,$$

then $A$ and $B$ are weakly separated.

A $T_5$**-space** is a hereditarily normal space which is also $T_4$.

In order to motivate the next definition, we remark that not every closed subset $C$ of a topological space is the zero-set of a real-valued function. A *necessary* condition for this to be the case is that $C$ be a $G_\delta$-set. For if $C$ is the zero-set of $f$ we have the representation $C = \bigcap f^{-1}([0, \frac{1}{n}])$. In a normal space, the converse is true i.e. every closed $G_\delta$-subset is a zero-set. For in this case we can write its complement $X \setminus C$ as a union $\bigcup C_n$ of closed sets. Then for each $n$ we can find a continuous function $f_n$ with values in $[0, 1]$ so that $f_n$ vanishes on $C$ and takes on the value 1 on $C_n$. Then $f = \sum 2^{-n} f_n$ clearly has $C$ as its zero-set. This is the motivation for the following definition:

**Definition:** A space is **perfectly normal** if every closed subset is a $G_\delta$. This condition can be restated in the form that the following sharper version of Urysohn's Lemma hold: if $A$ and $B$ are disjoint closed sets, then there is a continuous function $f$ from $X$ into $[0,1]$ so that $A = f^{-1}(0)$ and $B = f^{-1}(1)$. This definition is also equivalent to the conditions that each open set be a co-zero set or that each closed set be a zero-set.

A perfectly normal space which is $T_1$ is called a $T_6$-set.

# 8 Compactness

This is one of *the* most fundamental notions of analysis. The original example of its use is in the proof of the fact that a continuous function on a bounded, closed interval is bounded and attains its supremum and infimum. The abstract definition may at first glance hardly seem connected with this proof but the relationship will become clear in the course of the chapter.
A **covering** of $A$ (in $X$) is a family $\mathcal{U}$ of subsets of $X$ whose union contains $A$. A **subcovering** of $\mathcal{U}$ is a family $\mathcal{V} \subset \mathcal{U}$ which also covers $A$. If, in addition, $X$ has a topology, then the covering is said to be **open** if each set therein is open.

A subset $A$ of a topological space $X$ is **quasi-compact** if each open covering of $A$ has a finite sub-cover. In particular, $X$ is quasicompact if it is quasi-compact as a subset of itself. We remark that quasi-compactness is an intrinsic property of a space i.e. if $A$ is a subspace of $X$, then $A$ is quasi-compact in $X$ if and only if the space $(A, \tau_A)$ is quasi-compact.

$X$ is **compact** if it is quasi-compact and $T_2$.

We note some simple properties of these notions. The definition can be reformulated as follows: $X$ is quasi-compact if and only if each family $\mathcal{C}$ of closed subsets of $X$ with the finite intersection property has non-empty intersection. (The finite intersection property means that each finite subfamily of $\mathcal{C}$ has non-empty intersection). This follows immediately from the original definition by taking complements.

The union of finitely many quasi-compact sets is quasi-compact. Hence finite sets, for example, are quasi-compact. It follows that in a $T_2$-space, finite unions of compact sets are compact. This is *not* true in the absence of the separation property as the example of the interval with three endpoints shows (it is not $T_2$ and therefore not compact but it *is* the union of two copies of the unit interval).

Examples of compact sets are closed, bounded subsets of $\mathbf{R}^n$ (this is essentially the Heine-Borel theorem of elementary Analysis). We shall prove below a more general result from which this follows. A space with the indiscrete topology is always quasi-compact but it is not compact if it has more than one point. A discrete space is compact if and only if it is finite.

In order to produce less trivial examples, we shall require some theory.

**Proposition 68** *Let $A$ be a compact subset of a $T_2$-space $X$, $x$ a point of $X$ which does not lie in $A$. Then we can separate $x$ from $A$ in the sense that there are disjoint open sets $U$ and $V$ with $x \in U, A \subset V$.*

PROOF. Since $X$ is $T_2$, we can find, for every $y \in A$, disjoint open sets $U_y$ and $V_y$ with $x \in U_y$ and $y \in V_y$. Then $\{V_y : y \in A\}$ is an open cover of $A$ and so we can find a finite subcover $\{V_{y_1}, \dots, V_{y_n}\}$.

Then

$$U = \bigcap_{n=1}^{n} U_{y_i} \quad \text{and} \quad V = \bigcup_{i=1}^{n} V_{y_i}$$

are the required sets.

**Proposition 69** *Let $A$ and $B$ be disjoint, compact subsets of a $T_2$-space $X$. Then there are disjoint open sets $U$ and $V$ in $X$ with $A \subset B$ and $B \subset V$.*

PROOF. We repeat the method of the proof of the Lemma. For each $x \in A$, we find open disjoint sets $U_x$ and $V_x$ with $x \in U_x$ and $B \subset V_x$. The proof then proceeds in the same way, using the cover $\{U_x : x \in A\}$ of $A$.

■

**Proposition 70** *Let $A$ be a subset of a topological space $X$. Then a) if $X$ is $T_2$ and $A$ is compact, $A$ is closed; b) if $X$ is quasi-compact and $A$ is closed, $A$ is quasi-compact.*

PROOF. a) follows immediately from the Lemma above since if $x$ is not in $A$, we can find an open neighbourhood of $x$ which is disjoint from $A$. b) If $\mathcal{U}$ is an open cover of $A$ in $X$, then $\mathcal{U} \cup \{\mathcal{X} \backslash \mathcal{A}\}$ is an open covering of $X$ and so has a finite sub-covering. This clearly provides a finite subcovering of $\mathcal{U}$ for $A$.

■

Of course, these results imply that quasi-compact spaces are normal and that compact spaces are $T_4$. They also imply that a subset of a compact space is closed if and only if it is compact (this is important since it provides an *intrinsic* characterisation of closedness in this situation).

We now investigate the behaviour of quasi-compactness under continuous mappings.

**Proposition 71** *If $f : X \to Y$ is a continuous, surjective mapping and $X$ is quasi-compact, then so is $Y$.*

PROOF. If $\mathcal{U}$ is an open covering of $Y$, then $f^{-1}(\mathcal{U}) = \{\{^{-\infty}(\mathcal{U}) : \mathcal{U} \in \mathcal{U}\}$ is an open covering of $X$ and so has a finite subcovering. The images of these form the required subcovering of $\mathcal{U}$.

■

This simple result has a number of interesting corollaries. Firstly note that it implies that quotients of quasi-compact spaces are quasi-compact. Thus the Klein bottle and other surfaces, which we described as quotients of polygons, are quasi-compact (and hence compact).

A further consequence of this result is that a continuous mapping from a compact space $X$ into a $T_2$-space $Y$ is closed. For if $A$ is a closed subset of $X$, then it is compact and hence so is its image. This implies that $f(A)$ is closed. The following special case of this result is important enough to be quoted as a Proposition:

**Proposition 72** *If $f : X \to Y$ is a continuous bijection from a compact space onto a $T_2$-space, then $f$ is a homeomorphism.*

In other words, it is impossible to weak a compact topology on a set without losing the $T_2$-property.

The connection between our abstract definition and more classical concepts of compactness is provided by a rather deep characterisation of compactness in metric spaces which we now consider.

## 8.1 Characterisations of compactness in a metric spaces:

**Definition** A metric space $(X, d))$ is **precompact** (or totally bounded) if for every positive $\epsilon$ there is a finite subset $\{x_1, \dots, x_n\}$ so that $X \subset \bigcup_{i=1}^n U(x_i, \epsilon)$.

It is clear that any compact metric space is precompact (consider the open covering $\{U(x, \epsilon) : x \in X\}$). The converse is not true since, for example, the open interval $]0, 1[$ is precompact but not compact.

We note that precompactness is not a topological concept. The following *are*:

**Definition:** A topological space $(X, \tau)$ is **sequentially compact** if every sequence in $X$ has a convergent subsequence. It is **countably compact** if every sequence has a cluster point i.e. a point $x$ so that every neighbourhood of $x$ contains infinitely many elements of the sequence (more precisely, for each open neighbourhood $U$ of $x$ and each $N \in \mathbf{N}$, there is an $n > N$ with $x_n \in U$).

We note that a compact topological space is countably compact and that a sequentially compact space is also countably compact. In general, there is no other relationship between these notions. (In order to prove that a compact space is countably compact, consider the closed sets $B_m = \overline{\{x_m, x_{m+1}, \dots\}}$. This family has the finite intersection property and so its intersection is non-empty. But this intersection is precisely the set of cluster points of the sequence).

Our main result shows that the situation is quite different for metric spaces:

**Proposition 73** *Let $(X, d)$ be a metric space. Then the following conditions are equivalent: a) $(X, \tau_d)$ is compact; b) $(X, d)$ is precompact and complete; c) $(X, \tau_d)$ is sequentially compact; d) $(X, \tau_d)$ is countably compact.*

The proof will be divided up into a series of Lemmata, several of which are of interest in their own right. At this point we note that the above result contains the fact that a subset of $\mathbf{R}^n$ is compact if and only if it is closed and bounded. For we already know that completeness is equivalent to the fact that it is closed and the reader can verify for himself that for subsets of $\mathbf{R}^n$ total boundedness is equivalent to boundedness. (Needless to say, this fact is not valid for general metric spaces. The typical example is provided by any infinite dimensional Banach space).

**Proposition 74** *If $X$ is a sequentially compact metric space, then it is separable.*

PROOF. For each positive $\epsilon$ we find a maximal set $A_\epsilon$ so that for each pair $x, y$ of distinct points in $A_\epsilon$, $d(x, y) \geq \epsilon$. (??????) Now each $A_\epsilon$ is finite (otherwise we could extract a sequence $(x_n)$ of distinct elements from $A_\epsilon$ — this has the property that $d(x_m, x_n) \geq \epsilon$ if $m \neq n$. Of course, any subsequence has the same property and so cannot be Cauchy). We claim that the union $B$ of the $A_{\frac{1}{n}}$ (which is of course countable) is dense in $X$. For if $x$ were a point of $X$ which does not lie in the closure of $B$, there is an $n \in \mathbf{N}$ so that $U(x, \frac{1}{n})$ is disjoint from $B$ and so from $A_{\frac{1}{n}}$. This contradicts the maximality of the latter.

■

**Proposition 75** *If $X$ is sequentially compact, then it is compact.*

Proof. We know that $X$ is separable. Hence it has a countable basis $\mathcal{B} = \{\mathcal{U}_\backslash : \backslash \in \mathbf{N}\}$. Now suppose that $\mathcal{U}$ is an open covering. Then we can reduce it to a countable cover as follows: for each $x \in X$, there is a $U \in \mathcal{U}$ with $x \in U$. Since $\mathcal{B}$ is a basis, there is an $n_x$ so that $x \in U_{n_x}$ and the latter is a subset of $U$. The family $\mathcal{V}$ of all $U_{n_x}$ which arise in this way is a basis and, of course, countable. Hence it suffices to show that every countable open cover $\mathcal{V} = (\mathcal{V}_\backslash)$ has a finite subcover. Suppose that this is not the case. Then for each positive integer $n$ there is an $x_n$ which is not in the union of the first $n$ $V_n$'s. Consider the sequence $(x_n)$. By the hypothesis, this has a convergent subsequence $(x_{n_k})$. Let $x$ be the limit of this sequence. There is an $N \in \mathbf{N}$ so that $x \in V_N$. Then almost all of the $x_{n_k}$ are in $V_N$ and this clearly contradicts the construction of the sequence $(x_n)$.

■

**Proposition 76** *If $X$ is compact, then it is sequentially compact.*

Proof. Notice first that if $(x_n)$ is a sequence in $X$, then it has a limit point as we remarked above. We can now find a subsequence of $(x_n)$ which converges to $x$ as follows. We choose $n_1$ so that $x_{n_1}$ is in $U(x, 1)$. We then choose $n_2$ larger than $n_1$ so that $x_{n_2}$ is in $U(x, \frac{1}{2})$. Continuing in this manner, we can construct a subsequence $(x_{n_k})$ with $x_{n_k} \in U(x, \frac{1}{k})$ and this subsequence converges to $x$.

■

**Proposition 77** *If $X$ is a compact metric space, then it is precompact and complete.*

Proof. We already know that it is precompact. Suppose now that $(x_n)$ is a Cauchy sequence. Then there is a subsequence $(x_{n_k})$ which converges, say to $x$, since $X$ is sequentially compact. We now show that $x_n \to x$. For $\epsilon$ positive there is an integer $N$ so that $d(x_m, x_n) \leq \epsilon$ if $m, n \geq N$. Hence $d(x_m, x_{n_k}) \leq \epsilon$ if $m \geq N, n_k \geq N$. If we now let $n_k$ tend to infinity in this inequality, we have that $d(x_m, x) \leq \epsilon$ if $m \geq N$ which shows that the original sequence converges to $x$.

■

**Proposition 78** *If the metric space $X$ is precompact and complete, then it is compact.*

PROOF. First note that if $X$ is precompact, then every sequence $(x_n)$ has a Cauchy subsequence. For we can cover $X$ by finitely many balls of radius at most 1. Then there must be one of these balls, say $U_1$, which contains infinitely many terms of the sequence. Hence we can find a subsequence in $U_1$ which, for reasons which will soon be apparent, we denote by $(x_n^1)$. Proceeding in exactly the same way we find a subsequence $(x_n^2)$ of this sequence which lies in a ball $U_2$ of radius at most $\frac{1}{2}$ which is contained in $U_1$. In this manner, we can construct a decreasing sequence $(U_k)$ of balls, where $U_k$ has radius at most $2^{-k}$, and, for each $k$, a subsequence $(x_n^k)$ of the original sequence which lies in $U_k$. Further $(x_n^k)$ is a subsequence of $(x_n^{k-1})$. We can display this sequence of sequences as a square array and form the diagonal sequence $(x_n^n)$ which consists of the circled terms. This has the property that it is a subsequence of each of the subsequences constructed above, up to the first $k$ terms. In particular, this sequence lies in each $U_k$ (again up to a finite number of terms). This clearly implies that it is Cauchy as claimed. Now this Cauchy sequence converges since $X$ is complete and we have thus shown that $X$ is sequentially compact and hence compact by the previous Lemma.

■

The proof of this Lemma completes that of the main result as the reader can verify.

As a Corollary to the final Lemma, we have the following:

**Proposition 79** *Let $X$ be a metric space. Then it is precompact if and only if its completion is compact.*

PROOF. If the latter is compact, then it is precompact and hence so is $X$ (as a subspace of a precompact space). On the other hand, the precompactness of $X$ implies that of the completion as can easily be verified and so the latter is compact by the last Lemma.

■

As a final contribution to this circle of ideas we mention the following property of compact spaces which is exactly the one required to make rigorous the compactness arguments employed in the first Chapter.

**Proposition 80** *Let $\mathcal{U}$ be an open covering of a compact metric space $X$. Then there is a positive $\eta$ so that each subset of $X$ of diameter less than $\eta$ is contained in a set of $\mathcal{U}$.*

PROOF. It suffices to prove the result for a finite covering $U_1, \ldots, U_k$ (since $\mathcal{U}$ possesses a finite subcovering). If the claim were false, we could find for each $n$ a set $A_n$ of diameter at most $\frac{1}{n}$ which is not contained in any $U_k$. Choose $x_n \in A_n$. By going over to a subsequence, we can assume that $(x_n)$ converges, say to $x$. Then $x$ belongs to one of the $U$'s, say $U_i$. Choose a positive $\epsilon$ so that $U(x, \epsilon)$ is a subset of $U_i$ and $n$ so large that $d(x, x_n) \leq \frac{\epsilon}{3}$ resp. diam $A_n < \frac{\epsilon}{3}$. Then it is clear that $A_n$ is contained in $U_i$ which is a contradiction.

■

The positive number $\eta$ whose existence is ensured by the above Lemma is called **a Lebesgue number** for the covering $\mathcal{U}$.

We indicate briefly how this result can be used to stop the holes in the proofs given in Chapter 1.

We conclude this section with the remark that the above lemma can be restated as follows:

**Proposition 81** *Let $C_1, \ldots, C_n$ be closed, non-empty subsets of a compact metric space and suppose that their intersection is empty. Then there exists a positive $\epsilon$ so that any subset of $X$ which meets each $C_i$ has diameter at least $\epsilon$.*

As a Corollary, we have that if $C_1, \ldots, C_n$ is a closed covering of a compact metric space, then there is a positive $\epsilon$ which is such that if any set $A$ of diameter $\epsilon$ meets the sets $C_{i_1}, \ldots, C_{i_r}$, the intersection of the latter is non-empty.

## 8.2 Tychonov's theorem:

We now turn to one of the most important results on compact space, the theorem mentioned in the paragraph title which states that products of compact spaces are compact. In order to prove this, we introduce the concept of an ultrafilter:

**Definition:** An **ultrafilter** on a set $X$ is a filter $\mathcal{F}$ which is maximal in the sense that if $\mathcal{G}$ is a second filter which is finer than $\mathcal{F}$, then $\mathcal{F}$ and $\mathcal{G}$ coincide.

It follows immediately from Zorn's Lemma, applied to the family of all filters on a set which are finer than a given one (ordered by inclusion), that every filter can be refined to an ultrafilter.

Ultrafilters can be characterised as follows:

**Proposition 82** *A filter $\mathcal{F}$ on a set $X$ is an ultrafilter if and only if the following condition is satisfied: for each subset $A$ of $X$, either $A$ or its complement is in $\mathcal{F}$.*

PROOF. Suppose firstly that the filter is an ultrafilter and that there is a subset $A$ for which the above condition fails i.e. neither $A$ nor $X \setminus A$ are in $\mathcal{F}$. Then we can define a filter $\mathcal{G}$ which is finer than $\mathcal{F}$ as follows:

$$\mathcal{G} = \{\mathcal{B} \subset \mathcal{X} : \mathcal{A} \cup \mathcal{B} \in \mathcal{F}\}.$$

This is *strictly* finer since $X \setminus A \in \mathcal{G}$. This contradiction shows that ultrafilters have the above property.

Suppose on the other hand, that $\mathcal{F}$ is not an ultrafilter. Let $\mathcal{G}$ be a filter which is strictly finer than $\mathcal{F}$. Then there is a subset $A$ which is in $\mathcal{G}$ but not in $\mathcal{F}$. This $A$ fails the above condition. For if its complement were in $\mathcal{F}$ it would also be in $\mathcal{G}$. Then both $A$ and its complement would belong to the same filter.

■

If $a$ is a point in a set $X$, then the filter generated by the one point set $\{a\}$ is easily seen to be an ultrafilter,. Such filter are called **fixed ultrafilters.** Of course, an ultrafilter is fixed if and only if the intersection of its elements is non-empty. Such ultrafilters are not very interesting. Filters whose intersections are empty are called **free**. Free ultrafilters are "constructed" by applying the above existence statement to free filters, the typical example being an ultrafilter which is finer than the Fréchet filter on the integers. The relevance of ultrafilters for the proof of Tychonov's theorem is based on the following facts:

I. if $\mathcal{F}$ is an ultrafilter on a topological space and $x$ is a cluster point of $\mathcal{F}$, then $\mathcal{F} \to \S$. For we know that the fact that $x$ is a cluster pint of $\mathcal{F}$ means that a finer filter converges to $x$. But, apart from $\mathcal{F}$ itself, there *is no* finer filter than $\mathcal{F}$.

II. A topological space $X$ is quasi-compact if and only if every ultrafilter on $X$ converges. This is essentially a restatement of the definition of quasi-compactness (in terms of the finite intersection property). For example, we shall show here that if $X$ is quasi-compact then every ultrafilter $\mathcal{F}$ converges. Note that the family $\{\bar{A} : A \in \mathcal{F}\}$ has the finite intersection property and so its intersection is non-empty. Hence $\mathcal{F}$ has a cluster point to which it must converge by I.

III. The same proof provides the following characterisation: a space $X$ is quasi-compact if and only if every filter on $X$ has a cluster point.

We are now in a position to state and prove Tychonov's theorem. Due to its importance in analysis, it is perhaps worth mentioning that its proof uses the Axiom of Choice in an essential way (in the form of the existence of ultrafilters). In fact, it is known that the result stated here implies this Axiom and so its use is unavoidable.

**Proposition 83** *Let* $(X_\alpha)_{\alpha \in A}$ *be a family of quasi-compact spaces. Then their Cartesian product is quasi-compact. (Of course, the corresponding result for compact spaces holds also).*

PROOF. Let $\mathcal{F}$ be an ultrafilter on the product. Then for each $\alpha$ the image filter $\pi_\alpha(\mathcal{F})$ has a cluster point $x_\alpha$. The reader will have no difficulty in verifying that $x = (x_\alpha)$ is then a cluster point for $\mathcal{F}$.

■

**Projective limits of compacta:** We mentioned previously that general projective limits of sets can be trivial. However, as we shall now see, in the case of compact components such pathologies cannot occur. A typical example is the fact that the intersection of a decreasing sequence of non-empty, compact subsets of a given space is non-empty. This follows from the characterisation of compactness using the finite intersection property. The latter is a special case of the following more general result which we state for a general projective limit indexed by an directed set $A$ which need not be the integers:

**Proposition 84** *Let*

$$\{\pi_{\alpha\beta} : K_\beta \to K_a, \alpha \leq \beta, \alpha, \beta \in A\}$$

*be a projective spectrum of compact sets. Then the projective limit* $K$ *is compact and we have the formula*

$$\pi_\alpha(K) = \bigcap_{\beta \geq \alpha} \pi_{\beta\alpha}(K_\beta)$$

*for each* $\alpha \in A$*. In particular, if the* $K_\alpha$ *are non-empty, then so is* $K$ *and if each of the* $\pi_{\beta\alpha}$ *is surjective, then so are the* $\pi_\alpha$*.*

PROOF.

■

We remark that the above theorem is also true under the weaker assumption that for each $\alpha \leq \beta$ and each $x_\alpha \in K_\alpha$, the pre-image $\pi_{\beta\alpha}^{-1}(x_\alpha)$ of $x_\alpha$ in $K_\beta$ is compact.

We can also give a version of this theorem which relates to suitable families of mappings between the components of projective spectra: Suppose that we have two such spectra:

$$\{\pi_{\beta\alpha} : X_\beta \to X_\alpha, \alpha \leq \beta, \alpha, \beta \in A\}$$

and

$$\{\pi'_{\beta\alpha} : Y_\beta \to Y_\alpha, \alpha \leq \beta, \alpha, \beta \in A\}.$$

Suppose further that we have a collection $(f_\alpha)$ of mappings, where $f_\alpha$ is continuous from $X_\alpha$ into $Y_\alpha$ and where the $f_\alpha$ are compatible with the linking mappings in the sense that $f_\alpha \circ \pi_{\beta\alpha} = \pi'_{\beta\alpha} \circ f_\beta$ whenever $\alpha \leq \beta$ (see the commutative diagram ?????). Then we can define in a natural way a mapping $f$ from the projective limit $X$ of the first spectrum into $Y$, the limit of the second one, by defining the image of a thread $(x_\alpha)$ in $X$ to be the thread $(f_\alpha(x_\alpha))$. (That this *is* a thread follows from the compatibility condition).

In general, we cannot deduce interesting properties of $f$ from those of $f_\alpha$ for the simple reason that, as we have seen, one or both of the limits can trivialise. However, in the presence of compactness, we have, for example, the following result:

**Proposition 85** *If the spaces $X_\alpha$ and $Y_\alpha$ are all compact and the $f_\alpha$ are surjective, then so is $f$.*

## 8.3 Locally compact spaces:

Despite the importance of the concept of compactness in analysis, the basic space for the latter (namely, the real line) does not enjoy this property. However, it does satisfy the condition that every point has a compact neighbourhood and this suffices to allow one to use compactness arguments for many purposes. This is formalised in the following

**Definition:** A $T_2$-space is **locally compact** if each point has a compact neighbourhood.

This is equivalent to either of the following

- every point has a closed, compact neighbourhood (for $X$ then is automatically $T_2$);

- $X$ is $T_2$ and every neighbourhood of a point $x$ of $X$ contains a compact neighbourhood of $x$.

Of course $\mathbf{R}^n$, with its usual topology, is locally compact, as is any discrete space. $\mathbf{Q}$ is clearly not locally compact. The interval with two endpoints is a space which is not locally compact despite the fact that every point has a compact neighbourhood (this shows that the closedness condition in (1) above is essential).

Subsets of locally compact spaces need not be locally compact (as the rationals as a subset of the reals shows). However, we de have the following result:

**Proposition 86** *Let $U$ be an open (resp. $C$ a closed subset) of the locally compact space $X$. Then $U$ and $C$ are themselves locally compact.*

PROOF. We prove this for $U$ and leave the (easier) case of closed subsets to the reader. If $x$ is a point of $U$, then $U$ is a neighbourhood of $x$ in $X$ and so there is a closed neighbourhood $V$ of $x$ contained in $U$. There is also, by definition, a compact neighbourhood $W$ of $x$ in $X$. Then $W \cap V$ is a compact neighbourhood of $x$ in $U$.

■

As a Corollary, we note the fact the every subset of the form $C \cap U$ (i.e. the intersection of a closed and an open set) is locally compact, being an open subset of the locally compact space $C$. A subset of the above form is called **locally closed**. In fact, if a subset of a locally compact space is itself locally compact, it must be locally closed (Exercise).

It is clear that a disjoint union of locally compact spaces is locally compact. On the other hand, a product $\prod X_\alpha$ of non-trivial locally compact spaces can only be locally compact if all of the constituent spaces are locally compact and all but a finite number are compact.

Another stability property is that if a continuous function maps a locally compact space onto a $T_2$-space $Y$ and the mapping is open, then $Y$ is locally compact.

A locally compact space $X$ is said to be **$\sigma$-compact** if it can be expressed as a union of countably many compact subsets. We can then find a sequence $(K_n)$ of compacta in $X$ whose union is $X$ and which is such that each $K_n$ is contained in the interior of its successor $K_{n+1}$. In this case, any compact subset of $X$ is contained in some $K_n$.
PROOF.

■

We remark that a closed subspace of a $\sigma$-compact locally compact space is of the same type. However, this is not true of open subsets since any locally compact set is an open subset of a *compact* one, as we shall now show.

## 8.4 The Alexandrov compactification:

If $X$ is a locally compact space, there is a standard method of embedding it into a compact space by adding one point. As a model for this construction, recall the usual method of "compactifying" the real line by adding a point at infinity. The resulting space is then homeomorphic to $S^1$. The abstract construction is as follows. Let $(X, \tau)$ be a topological space (at this point we shall not assume that it is locally compact). We introduce a new set $X_\infty$ which is $X$ together with a point which we shall denote by $\infty$ for obvious reasons. On $X_\infty$ we define a topology $\tau_\infty$ as follows: $U \in \tau_\infty$ if and only if $U \subset X$ and $U \in \tau$ or $\infty \in U$ and $X \setminus U$ is compact in $X$. It is clear that $(X_\infty, \tau_\infty)$ is a topological space which contains $X$ as a subspace. It is quasi-compact and is $T_1$ provided that $X$ is. The important point to note is that it is $T_2$ (and so compact) precisely when $X$ is locally compact. This construction shows that each locally compact space is homeomorphic to an open subspace of a compact space. In fact, the latter condition is a characterisation of local compactness since we saw above that it is necessary. A simple consequence is that each locally compact space is $T_{3\frac{1}{2}}$ (since each compact space is $T_4$ and the $T_{3\frac{1}{2}}$-property is hereditary).

## 8.5 Further compacness properties:

We now turn to some weaker properties which are related to compactness in that they are defined by covering properties: In order to avoid pathologies, we will incorporate the $T_2$-property in the definition.

**Definition:** A $T_2$-space $X$ is called

**countably compact** if each countable open covering of X has a finite subcover;

**Lindelöf** if each open covering has a countable subcovering.

It is immediately clear that compactness is equivalent to countable compactness plus the property of being Lindelöf. In other words, we have split compactness into the combination of two weaker properties.

We note some simple properties of such spaces:

- a closed subspace of a countably compact space (resp. a Lindelöf space) is countably compact (resp. Lindelöf);

- a countably generated space is Lindelöf;
- the following condition on a $T_2$-space is equivalent to countable compactness: countable families of closed sets with the finite intersection property have non-empty intersections;
- continuous images of countably compact (resp. Lindelöf) spaces are countably compact (resp. Lindelöf).
- a continuous, real-valued function on a countably compact spaces is bounded and attains its supremum;
- the property of being Lindelöf for a $T_2$-space is equivalent to the following: every family of closed subsets with the countable intersection property has a non-empty intersection.

We now turn to an important result on countably generated spaces. As we have remarked above, these are automatically Lindelöf and we shall now show that we can lift the $T_3$ property to $T_4$.

**Proposition 87** *A countably generated $T_3$-space is $T_4$.*

PROOF. Let $A$ and $B$ be disjoint, closed subsets of $X$. For each $x$ in $A$ there is an open neighbourhood $U_x$ of $x$ whose closure has empty intersection with $B$. The family of all such $U_x$ is an open covering of $A$ and the latter is Lindelöf as a closed subset of $X$. Hence we can find a countable subcovering which we denote by $(U_n)$. Similarly, we can find a countable open covering $(V_n)$ of $B$ which is such that the closure of each element has empty intersection with $A$. We now replace the $U$'s and $V$'s with new sequences $(U'_n)$ and $(V'_n)$ which have the same properties and, in addition, are such that each $U'_n$ is disjoint from $V'_1, \dots, V'_{n-1}$ and each $V'_n$ is disjoint from $U'_n, \dots, U'_n$. These sets are defined recursively as follows

$$\begin{array}{llll} U'_1 & = U_1 & V'_1 = V_1 \setminus (V_1 \cap \overline{U_1}) \\ U'_2 & = U_2 \setminus (U_2 \cap \overline{V_1}) & V'_2 = V_2 \setminus (V_2 \cap (\overline{U_1} \cup \overline{U_2})) \\ U'_n & = U_n \setminus (U_n \cap (\bigcup_i^{n-1} \overline{(V_i)}) & V'_n = V_n \setminus (V_n \cap (\bigcup_1^n \overline{(U_i)}) \end{array}$$

By their very definition, the $U'_n$ and the $V'_n$ are open and hence so are $U' = \bigcup U'_n$ and $V' = \bigcup V'_n$. By the disjointness condition, $U'$ and $V'$ are disjoint. Also $A \subset U'$ and $B \subset V'$ since at no point of the construction have we removed an element of $A$ from the $U_n$'s.

■

From this we can easily deduce that a countably generated $T_3$-space is homeomorphic to a subspace of the Hilbert cube and so is metrisable.

A further weakening of the notion of compactness is contained in the next definition:

**Definition:** A $T_2$-topological space $X$ is said to be **paracompact** if each open covering $\mathcal{U}$ of $X$ has a locally finite refinement i.e. a refinement $\mathcal{V}$ such that each point in $X$ has a neighbourhood which meets only finitely many sets of $\mathcal{V}$.

Exactly as in the proof of the fact that closed subsets of compact spaces are compact, one can show that closed subsets of paracompact spaces are paracompact. The disjoint union of a family of paracompact spaces is clearly paracompact. This immediately provides a large supply of paracompact spaces which are not compact. Of course the converse is always true i.e. compact spaces are paracompact.

It is not true in general that products of paracompact spaces are paracompact. However, we *do* have the result:

**Proposition 88** *The product $K \times X$ of a compact space and a paracompact space is paracompact.*

PROOF.

■

An example of a space which is *not* paracompact is the ordinal space $[0, ø[$. This can be seen as follows: for each $x$ we define a neighbourhood $U_x$ as follows. If $x$ is not a limit ordinal, we define $U_x$ to be $\{x\}$. If $x$ *is* a limit ordinal, then $U_x$ is $]y, x+1[$ where $y$ is any ordinal strictly less than $x$. Then the open covering $\{U_x\}$ has no locally finite refinement.

In many cases we can replace the definition of paracompactness by one which is formally weaker. This will be useful in verifying that certain types of spaces (notably metric spaces) are paracompact.

**Definition:** A cover of a space $X$ is said to be **$\sigma$-locally finite** if it can be expressed as a countable union of a sequence $(\mathcal{U}_\backslash)$ of subfamilies, each of which is locally finite. (Of course, the $\mathcal{U}_\backslash$ will not, in general, be coverings.

**Proposition 89** *A $T_3$-space is paracompact, provided that each open cover has a $\sigma$-locally finite refinement.*

.

Proof.

∎

The following result describes two classes of spaces which are automatically paracompact:

**Proposition 90** *Each metric space is paracompact as is each $\sigma$- compact locally compact space*

It follows immediately form the second part of the above result that disjoint unions of $\sigma$-compact, locally compact spaces are paracompact. In fact this is a characterisation of locally compact, paracompact spaces:

**Proposition 91** *A locally compact space is paracompact if and only if it is the disjoint union of $\sigma$-compact, locally compact spaces.*

Proof. It suffices to show that a paracompact, locally compact space is a disjoint union of $\sigma$-compact spaces. For each $x$ in $X$, we choose an open, relatively compact neighbourhood $U_x$. Let $\mathcal{V}$ be a locally finite, open refinement of the covering $\{U_x\}$. Of course, each $V$ in $\mathcal{V}$ is relatively compact. It is also clear that each relatively compact subset of $X$ meets at most finitely many sets of $\mathcal{V}$. We now define an equivalence relationship on $X$ as follows: $x \sim y$ if and only if there is a finite collection $U_1, \dots, U_n$ of sets of $\mathcal{V}$ so that $x \in U_1$, $y \in U_n$ and for each $i$, $U_i$ and $U_{i+1}$ intersect.

It is clear that this is an equivalence relationship. It is also clear that the corresponding equivalence classes are open (for if $x$ is in an equivalence class, then so is $V$ where $V$ is a member of $\mathcal{V}$ which contains $x$). Hence we can complete the proof by showing that the equivalence class which contains a given point $x$ is $\sigma$-compact. We do this by defining a sequence $(V_n)$ of relatively compact open subsets of $X$ as follows:

$V_1$ is the union of those sets of $CalV$ which contain $x$;

$V_2$ is the union of those sets of $\mathcal{V}$ which meet $V_1$ and so on (note that at each step we only have a finite union by the local finiteness of $\mathcal{V}$). It is clear that the equivalence class which contains $x$ is the union of the $V_n$.

∎

The main importance of paracompact spaces lies in the fact that they possess so-called partitions of unity which are defined as follows: if $X$ is a topological space, then a **partition of unity** for $X$ is a family $(f_\alpha)_{\alpha \in A}$ of continuous functions from $X$ into the unit interval which sum to one and are such that the open covering $(U_\alpha)$ of $X$ is locally finite, where $U_\alpha = \{t \in X : f_\alpha(t) > 0\}$. Such a partition is said to be **subordinate** to a covering $\mathcal{U}$ of $X$ if $(U_\alpha)$ is a refinement of $\mathcal{U}$.

It is clear that if a space $X$ has the property that each open covering has a partition of unity subordinate to it, then it is paracompact. We shall now show that, conversely, each paracompact space has this apparently stronger property.

We begin with the result that paracompact spaces are $T_{3\frac{1}{2}}$. (Recall that they are $T_2$ by definition).

**Proposition 92** *A paracompact space is normal (and so $T_4$).*

PROOF. This is very similar to the proof of the normality of compact spaces, the finiteness argument being replaced by local finiteness which suffices to carry through the proof. We shall only prove the first step, that the above hypothesis implies regularity. The second step is almost identical.

Suppose then that $C$ is closed and $x$ lies outside of $C$. For each $y$ in $C$ we find an open set $V_y$ which contains $y$ but does not contain $x$ in its closure. The covering $(V_y)_{y \in C}$ of $C$ has a locally finite refinement $\mathcal{U}$. It is then routine to check that the union of $\mathcal{U}$ is a neighbourhood of $C$ whose closure does not contain $x$.

■

We now show that locally finite open coverings of normal spaces admit subordinate partitions of unity.

**Proposition 93** *If $(U_i)$ is a locally finite open covering of a normal space $X$, then there is a partition of unity subordinate to $\mathcal{U}$.*

PROOF. We know that we can find a second open covering $(V_i)$ so that $\bar{V}_i \subset U_i$ for each $i$.

By the normality we can find further open sets $W_i$ with $\bar{V}_i \subset W_i \subset \bar{W}_i \subset U_i$. Now we use Urysohn's Lemma to construct continuous functions $g_i$ from $X$ into the unit interval so that $g_i$ takes on the value 1 on $\bar{V}_i$ and 0 outside of $W_i$. Consider the sum $g = \sum g_i$. This is positive on $X$ (since $(V_i)$ covers the latter) and is continuous since $(W_i)$ is locally finite. We now define the functions $(f_i)$ by specifying

$$f_i(t) = g_i(t)/g(t).$$

■

If we combine the last two result we have the promised sharpening of the definition of paracompactness:

**Proposition 94** *A space is paracompact if and only each open covering admits a subordinate partition of unity.*

# 9 Uniform spaces

We now investigate a structure which occupies an intermediate position between a metric and a topology. Spaces with this structure are called **uniform spaces** because they are the natural context for the concept of uniform continuity. Roughly speaking, they are spaces whose structure is induced by a *family* of metrics (more precisely , of pseudo-metrics).

**Definition:** A **pseudo-metric** on a set $X$ is a mapping $d$ from $X \times X$ into $\mathbf{R}^+$ so that

a) $d(x, x) = 0$ $(x \in X)$;

b) $d(x, y) = d(y, x)$ $(x, y \in X)$;

c) $d(x, z) \leq d(x, y) + d(y, z)$ $(x, y, z \in X)$. In other words we omit the condition of positive definiteness in the definition of a metric.

If $X$ is a set, a collection $\mathcal{D}$ of pseudometrics on $X$ is called a) a **cone** if for each pair $d_1, d_2$ of elements of $\mathcal{D}$ and each positive $l$, $ld_1 \in \mathcal{D}$ and $\max(d_1, d_2) \in \mathcal{D}$. b) **separating** if for each $x \neq y$ in $X$, there is a $d \in \mathcal{D}$ with $d(x, y) \neq 0$.

If $\mathcal{D}$ is a separating cone on $X$, then a pseudo-metric $d$ on $X$ is **uniformly continuous with respect to** $\mathcal{D}$ if for every positive $\epsilon$ there is a $d_1 \in \mathcal{D}$ so that $d(x, y) \leq \epsilon$ whenever $d_1(x, y) \leq 1$. We denote by $\tilde{\mathcal{D}}$ the set of all such pseudo-metrics. A separating cone $\mathcal{D}$ is said to be **saturated** if $\tilde{\mathcal{D}} = \mathcal{D}$. A **uniformity** on a set $X$ is a saturated family of pseudo-metrics thereon.

If $\mathcal{D}$ is a separating family of pseudo-metrics on $X$, there is a smallest saturated family containing $\mathcal{D}$. This consists of the following pseudo-metrics

This is called the **uniform structure generated by** $\mathcal{D}$ and $\mathcal{D}$ is a **subbasis** for it. If $\mathcal{D}$ is in addition a cone, it is called **a basis** for the uniformity.

If $(X, d)$ is a metric space, then it has a natural uniform structure, induced by the separating family $\{d\}$. Thus each metric space is a uniform space. We shall now show that each $T_{3\frac{1}{2}}$-space carries a natural uniformity. Recall that if $X$ is such a space, then $C(X)$ resp. $C^b(X)$ denotes the space of continuous, complex-valued functions (resp. bounded, continuous, complex-valued functions) on $X$. If $f$ is an element of $C(X)$, then

$$d_f : (x, y) \mapsto |f(x) - f(y)|$$

is a pseudo-metric on $X$. The families $\{d_f : f \in C(X)\}$ resp. $\{d_f : f \in C^b(X)\}$ define uniform structures on $X$ which are called the $C$**-uniformity** resp. the $C^b$**-uniformity.**

If $(X, \mathcal{D})$ and $(Y, \mathcal{D}_\infty)$ are uniform spaces, a mapping $f$ from $X$ into $Y$ is said to be **uniformly continuous** if for each $d \in \mathcal{D}_\infty$, $\lceil \circ (\{ \times \{) \in \mathcal{D}$. In terms of bases or subbases, this definition must be reformulated as follows:

A **uniform isomorphism** between two space is a bijection which, together with its inverse, is uniformly continuous.

If $\mathcal{D}$ and $\mathcal{D}_\infty$ are two uniformities on a set $X$, we say that $\mathcal{D}_\infty$ is **finer than** $\mathcal{D}$ if $\mathcal{D} \subset \mathcal{D}_\infty$. (This means that the identity on $X$ is uniformly continuous as a mapping from $(X, \mathcal{D}_\infty)$ into $(X, \mathcal{D})$).

For example if $X$ is a discrete uniform space i.e. is provided with the uniformity induced by the discrete metric, then any mapping from $X$ into a second uniform space is uniformly continuous. If $X$ and $X_1$ are two $T_{3\frac{1}{2}}$-spaces, then any continuous mapping $f$ from $X$ into $X_1$ is uniformly continuous both for the $C$- uniformity and for the $C^b$-uniformity. However, the identity function on $X$ is not uniformly continuous from the $C$-uniformity into the $C^b$-uniformity unless $C(X) = C^b(X)$ i.e. every continuous real-valued function on $X$ is bounded (spaces with this property are called **pseudo-compact**).

**The topology of a uniform space:** Analogous to the case of metric space, uniform spaces carry a natural topology which is defined as follows: for each pseudo-metric $d$, we define

$$U_d(x) = \{y \in X : d(x, y) < 1\}$$

$(x \in X)$. The family of all such subsets forms a basis for a topology on $X$ which we denote by $\tau_{\mathcal{D}}$ (the topology **associated with** $\mathcal{D}$). Naturally, every uniformly continuous mapping between uniform spaces is continuous for the associated topologies.

The topology described above is always $T_{3\frac{1}{2}}$ as can be seen as follows:

If we combine this with the fact noted above we see that a topological space is **uniformisable** (i.e. its topology is induced by a uniformity) if and only if it is $T_{3\frac{1}{2}}$.

**Completeness:** We now discuss briefly the concept of completeness for a uniform space. The definition of a Cauchy sequence can be carried over to uniform spaces ion the natural way. However, the more general situation requires a more general concept to define a suitable form of completeness. We shall cast our definition within the framework of nets. A net $(x_\alpha)_{\alpha \in A}$ in a uniform space $(X, \mathcal{D})$ is **Cauchy** if for each $d$ in $\mathcal{D}$ there is a $\gamma$ so that if $\alpha, \beta \geq \gamma$, then $d(x_\alpha, x_\beta) \leq 1$. The space is **complete** if each Cauchy net there converges. We remark that in the above definition, it suffices to verify the condition for each $d$ from a basis for the uniformity. In fact, it suffices to use a subbasis $\mathcal{D}_\infty$, but in this case the condition must be modified as follows: for each $\epsilon > 0$ and each $d \in \mathcal{D}_\infty$, there exists a $\gamma \in A$ so that $d(x_\alpha, x_\beta) \leq \epsilon$ if $\alpha, \beta \geq \gamma$.

## 9.1 Constructions of uniform spaces:

Not unexpectedly, we can carry over the constructions which we discussed within the context of topological spaces to uniform spaces. We begin with initial and final structures:

**Definition:** It follows almost immediately from the definition that if $\mathcal{D}$ is the initial uniformity defined by the family $(\mathcal{D}_\alpha)$, then the corresponding topology $\tau_\mathcal{D}$ is the initial topology defined by the $\tau_{\mathcal{D}_\alpha}$. Also a mapping $f$ from a uniform space $Y$ into $X$ is uniformly continuous if and only if $f_\alpha \circ f$ is uniformly continuous for each $\alpha$.

**Examples of initial uniformities:** I. The $C$- and $C^b$-uniformities on a topological space are the initial uniformities defined by the mappings of $C(X)$ resp. $C^b(X)$.
II. **Subspaces:** if $(X, \mathcal{D})$ is a uniform space and $X_1$ is a subset, then the restrictions of the elements of $\mathcal{D}$ to $X_1$ generate a uniform structure. This is just the initial uniformity induced by the natural embedding of $X_1$ in $X$. We denote it by $\mathcal{D}_{\mathcal{X}_\infty}$. We remark that $\tau_\mathcal{D}$ induces on $X_1$ the topology which is defined by the induced uniformity.
III. **Products:** On a product $\prod X_\alpha$ of a family of uniform spaces, we define a uniformity simply by taking the initial uniformity induced by the projection mappings $\pi_\alpha$ from $X$ onto $X_\alpha$. Once again, this construction is compatible with the corresponding one for topologies.
IV. **Projective limits:** Let

$$\{\pi_{\beta\alpha} : X_\beta \to X_\alpha, \alpha, \beta \in A, \alpha \leq \beta\}$$

be a projective spectrum, where the $X_\alpha$'s are uniform spaces and the $\pi_{\beta\alpha}$'s are uniformly continuous.

Just as for metric spaces, we can prove the following simple facts about completeness in relation to these constructions. A subset $X_1$ of a complete uniform space $(X, \mathcal{D})$ is complete for the induced structure if and only if it is closed. A product of a family of complete spaces is itself complete. It follows from these two facts that a projective limit of a spectrum of complete spaces is complete (as a closed subspace of their product).

## 9.2 The spectrum of a uniform space:

We shall now show that each uniform space has a natural representation as a projective limit of metric spaces (more precisely, a *complete* uniform space has such a representation.) This allows a reduction of many results on uniform spaces to the case of a metric space. We begin with the simple remark that each pseudo-metric $d$ on a set $X$ defines a metric on a quotient space in a natural way. We simply define the equivalence relationship $\sim_d$ by setting $x \sim_d y$ if and only if $d(x, y) = 0$. We denote by $X_d$ the corresponding quotient space of $X$. Then it is clear that $d$ induces a *metric* on this quotient space which, by an abuse of notation, we also denote by $d$. Finally, $(\hat{X}_d, \hat{d})$ denotes the metric completion of this space.

We now suppose that $(X, \mathcal{D})$ is a uniform space. If $d$ and $d_1$ are elements of $\mathcal{D}$ so that $d \leq d_1$, then there is a natural mapping (with Lipschitz constant $\leq 1$) from $\hat{X}_{d_1}$ into $\hat{X}_d$ which we denote by $\pi_{d_1,d}$ and which is constructed as follows.

The system

$$\{\pi_{d_1,d} : \hat{X}_{d_1} \to \hat{X}_d, d, d_1 \in \mathcal{D}, \lceil \leq \lceil_\infty\}$$

is a projective spectrum. We denote by $(\hat{X}, \hat{\mathcal{D}})$ its projective limit. Then there is a natural mapping $i_X$ from $X$ into $\hat{X}$ defined by mapping $x$ onto the thread $(\pi_d(x))_{d \in \mathcal{D}}$.

**Proposition 95** *$i_X$ is a uniform isomorphism from $X$ onto a $\tau_{\hat{\mathcal{D}}}$-dense subspace of $(\hat{X}, \hat{D})$. If $X$ is complete, then $i_X$ is onto.*

Now we know from the above that the space $(\hat{X}, \hat{\mathcal{D}})$ is complete. Hence we have shown that every uniform space is uniformly isomorphic to a dense subspace of a complete one. The latter enjoys the following extension property: if $f$ is a uniformly continuous mapping from $X$ into a complete uniform space $Y$, then $f$ has a (unique) extension to a uniformly continuous mapping $\tilde{f}$ from $\hat{X}$ to $Y$. The proof is almost exactly as in the metric case. In view of this analogy it is natural to call the space $\hat{X}$ the **completion** of $X$.

We remark that in defining the natural projective representation of $X$ and hence of its completion, it suffices to use a basis for the uniformity.

We turn now to the topic of compactness for uniform spaces: $(X, \mathcal{D})$ is said to be **compact** if it is compact for the associated topology. It is **totally bounded** if for each $d \in \mathcal{D}$, there is a finite subset $A$ of $X$ so that $X \subset \bigcup_{x \in A} U_d(x)$. It is **precompact** if its completion $\hat{X}$ is compact.

**Proposition 96** *For a uniform space, $(X, \mathcal{D})$, the following are equivalent:*
*a) $X$ is totally bounded;* *b) for each $d \in \mathcal{D}$, $(X_d, d)$ is a totally bounded*

*metric space; c) for each $d \in \mathcal{D}$, $(\hat{X}_d, d)$ is compact; d) $X$ is precompact.*

Proof.

■

**Proposition 97** *If $K$ is a compact space, then there is a unique uniformity on $X$ which induces the topology.*

Proof. We already know that such a uniformity exists since $X$ is $T_4$. We now demonstrate its uniqueness.

■

A simple Corollary of this result is that ever continuous function $f$ from a compact space $K$ to a uniform space $Y$ is automatically uniformly continuous.

## 9.3 The Samuel compactification:

If $(X, \mathcal{D})$ is a uniform space, there is a finest precompact uniformity $\mathcal{D}_{\surd}$ on $X$ which is coarser than the original one. $\mathcal{D}_{\surd}$ is the uniformity generated by all those pseudo-metrics $d$ in $\mathcal{D}$ for which the associated metric space $X_d$ is totally bounded. We remark that if $f$ is a function in $C^b(X)$, then $d_f$ is in $\mathcal{D}_{\surd}$. This implies that $\mathcal{D}_{\surd}$ separates $X$ (since it is finer than the $C^b$-uniformity). (We are regarding $X$ as a topological space with the topology $\tau_{\mathcal{D}}$). The completion $(\hat{X}, \hat{\mathcal{D}}_p)$ of $(X, \mathcal{D}_{\surd})$ is compact. This space is called the **Samuel compactification** of $X$ and denoted by $\sigma X$. It has the following characteristic property: if $f$ is a uniformly continuous mapping from $X$ into a compact space $Y$, then there is a unique continuous extension $\tilde{f}$ of $f$ to a function from the Samuel compactification of $X$ into $Y$.

## 9.4 The Stone-Čech compactification:

We shall now consider the completion of a $T_{3\frac{1}{2}}$-space with respect to the $C^b$-uniformity. It follows immediately from the description of the Samuel compactification that this is also the Samuel compactification of $X$, provided with the $C$-uniformity. The space which is obtained in this manner is called the **Stone-Čech compactification** of $X$ and is denoted by $\underline{X}$. It has the characteristic property that every continuous function $f$ from $X$ into a compact space $K$ has a (unique) continuous extension to a mapping from $\beta X$ into $K$. In particular, every bounded, continuous, real-valued function on $X$ extends to $\beta X$. This means that the spaces $C^b(X)$ and $C(\beta X)$ are essentially the same. In particular, the natural embedding of $X$ into $\beta X$ is a homeomorphism (since both spaces have the initial topology induced by $C^b(X) = C(\beta X)$). Thus we have proved the following characterisation of $T_{3\frac{1}{2}}$-spaces: a topological space is $T_{3\frac{1}{2}}$ if and only if it is homeomorphic to a subspace of a compact space.

The above extension property implies, as in the case of completions, that the Stone-Čech compactification of a space is unique. More precisely, if $X$ is embedded as a dense subspace of a second compact space $K$ in such a way that the corresponding extension property holds, then there is a homeomorphism between $K$ and $\beta X$ which fixes $X$, regarded as a subspace of $K$ and $\beta X$. It follows immediately that if $Y$ is any subset of $\beta X$ which lies between $X$ and $\beta X$, then $\beta X$ is also the Stone-Čech compactification of $Y$. For the extension property of $X$ clearly carries over to $Y$.

We remark that if $A$ and $B$ are disjoint *zero sets* of $X$, then their closures in $\beta X$ are also disjoint. For there is a function $f \in C^b(X)$ which takes on the value 0 on $A$ and 1 on $B$ (easy exercise). Let $\tilde{f}$ be an extension of $f$ to a continuous function on $\beta X$. Then the closure of $A$ lies in the set of zeros of $\tilde{f}$ while that of $B$ lies in the set of zeros of $\tilde{f} - 1$.

A subset $X_0$ of a topological space $X$ is said to be $C^b$**-embedded** in $X$ if each bounded continuous real-valued function on $X_0$ has an extension to a bounded, real-valued continuous function on $X$. Thus we have just seen that the characteristic property of $\beta X$ is that $X$ is $C^b$-embedded there as a dense subspace. The theorem of Tietze can be regarded as stating that each closed subset of a $T_4$- space is $C^b$-embedded. We combine these facts to deduce that every compact subset $K$ of a $T_{3\frac{1}{2}}$-space $X$ is $C^b$-embedded. For we can regard the compact subset as a subset of $\beta X$ where it is still compact and hence closed. By Tietze's theorem, $K$ is $C^b$-embedded in $\beta X$ and so, *a fortiori*, in $X$.

Suppose now that $X_0$ is a subset of a $T_{3\frac{1}{2}}$-space $X$. An obvious candidate for the Stone-Čech compactification of $X_0$ is its closure in $\beta X$. Just when this is the case is the content of the next Proposition:

**Proposition 98** *Let $X_0$ be a subset of the $T_{3\frac{1}{2}}$-space $X$. Then the closure of $X_0$ in $\beta X$ is the Stone-Čech compactification of $X_0$ if and only if $X_0$ is $C^b$-embedded in $X$.*

PROOF. If $X_0$ is $C^b$-embedded in $X$, then it clearly $C^b$- embedded in $\beta X$ and hence *a fortiori* in its closure in $\beta X$. Hence the latter is its compactification by the above remarks. On the other hand, if the closure of $X_0$ in $\beta X$ is the Stone-Čech compactification, then $X_0$ is $C^b$-embedded in its closure which is, in turn, $C^b$-embedded in $\beta X$ by Tietze's theorem. This implies that $X_0$ is $C^b$-embedded in $\beta X$ and hence in $X$.

■

We remark here that the existence of the Stone-Čech compactification can be used to give a short proof of Tychonov's theorem on the compactness of products. Suppose that $(K_\alpha)$ is a family of compact space and let $X$ be its Cartesian product. The latter is, of course, $T_{3\frac{1}{2}}$. Now the natural projections $\pi_\alpha$ from $X$ onto $K_\alpha$ can, by the defining property of the Stone-Čech compactification, be extended to mappings $\tilde{\pi}_\alpha$ from $\beta X$ onto $K_\alpha$. These mappings in turn define one from $\beta X$ onto $X$. Hence $X$, as the continuous image of a compact space, is itself compact.

## 9.5 The real-compactification:

This is, by definition, the completion of a $T_{3\frac{1}{2}}$-space under the $C(X)$- uniformity. It is denoted by the symbol $\upsilon X$, the Greek "upsilon" being used to suggest the adjective "unbounded" because of the role of the space of unbounded (more exactly, not necessarily bounded) continuous functions on $X$. $X$ is **realcompact** if it is complete for the above uniformity i.e. if $X = \upsilon X$. If we recall the construction of the completion as a closed subspace of a product and note that for the above uniformity the associated metric spaces are all subspaces of copies of $\mathbf{R}$, we obtain the following characterisation of real-compact spaces:

**Proposition 99** *A topological space is real compact if and only if it is homeomorphic to a closed subset of a product of copies of the real line (i.e. a space of the form $R_A$ for some indexing set A).*

The reader should compare this to the characterisation of $T_{3\frac{1}{2}}$-spaces as *general* subspaces of such products.

The property of real-compactifications which corresponds to the extension property of $\beta X$ is the following: for every continuous mapping $f$ from $X$ into a real-compact space $Y$ (in particular into $\mathbf{R}$), there is a unique continuous extension $\tilde{f}$ of $f$ to a function from $\upsilon X$ into $Y$.

This implies that $C(X)$ and $C(\upsilon X)$ are naturally isomorphic. If we define a subset $X_0$ of a space $X$ to be **$C$-embedded** whenever each continuous, real-valued function on $X_0$ has a continuous extension to a function on $X$, then the real-compactification can be characterised as a real-compact space which contains $X$ as a dense, $C$-embedded subset.

We list some simple properties of real-compact spaces which can be deduced easily from the above remarks:

- let $X$ be a real-compact space. Then each closed subspace is also real-compact;
- If $X_0$ is $C$-embedded in a real-compact space $X$, then the closure of $X_0$ in $X$ is the latter's real-compactification;
- an arbitrary product of real-compact spaces is real-compact;
- the projective limit of a spectrum of real-compact spaces is real-compact;
- if $(X_\alpha)$ is a family of real-compact subsets of a given space $X$, then their intersection is also real-compact;
- if $f : X \to Y$ is a continuous mapping between $T_{3\frac{1}{2}}$-spaces, whereby $X$ is real-compact and $Y_0$ is a real-compact subset of $Y$, then $f^{-1}(Y_0)$ is also real-compact.

## 9.6 Alternative approaches to uniform spaces:

In our treatment we have regarded uniform spaces as generalised metric spaces since this appears to be the simplest and natural approach. However, many of the standard texts use alternative, but equivalent, approaches which we discuss briefly.

**Uniformities defined by entourages:** If $(X, \mathcal{D})$ is a uniform space, then we define a family $\mathcal{U}$ of subspaces of $X \times X$ as follows:

$U \in \mathcal{U}$ if and only if there is a $d \in \mathcal{D}$ with $\{(\S, \dagger) \in \mathcal{X} \times \mathcal{X} : \lceil(\S, \dagger) < \infty\} \subset \mathcal{U}$.

The sets in this family are called **entourages**. It is easy to see that the family of entourages is a filter and, further, that the following conditions are satisfied;

- each $U \in \mathcal{U}$ contains the diagonal set $\Delta$;
- if $V \in \mathcal{U}$, there is a $U \in \mathcal{U}$ with $U \circ U \subset V$. (Here $U \circ U$ denotes the set

  $\{(x, z) \in X \times X : \text{there exists} \quad y \in U \quad \text{with} \quad (x, y) \in U \quad \text{and} \quad (y, z) \in U\}$).
- if $x \neq y$ in $X$ then there exists an entourage which does not contain $(x, y)$. (In other words, the diagonal set is the intersection of the family of entourages).

Then an alternative definition of a uniform space is to define it as a set provided with a filter $\mathcal{U}$ of subsets of its square which satisfies these conditions. We shall show below that this definition is equivalent to our one.

**Uniformities via uniform coverings:** If $(X, d)$ is a metric space, then we recall that a Lebesgue number for an open covering is a positive $\epsilon$ so that each subset $A$ of $X$ which has diameter at most $\epsilon$ is contained in a set of the covering. This is equivalent to the fact that there is a positive $\epsilon$ so that for each point $x$, $U(x, \epsilon)$ is contained in some set of the covering as the reader can verify. (However, the precise value of the $\epsilon$ may depend on which version of the definition one uses). As we know, every open covering of a compact metric space has a Lebesgue number. We define an open covering $\mathcal{U}$ of a metric space to be a **uniform cover** if it has a Lebesgue number. More generally, if $(X, \mathcal{D})$ is a uniform space then an open covering $\mathcal{U}$ is defined to be **uniform** if there is a $d \in \mathcal{D}$ for which the cover has a Lebesgue number. The set of all uniform covers satisfy the following conditions:

In order to verify the equivalence of the three definitions of uniform spaces, we shall develop a method of constructing pseudo-metrics on topological spaces out of coverings. This has applications in many situations of general topology.

**Definition:** If $\mathcal{U}$ is a cover of a space $X$ and $A$ is a subset, then the **star** of $A$ with respect to $\mathcal{U}$ (written $\mathrm{st}(A;\mathcal{U})$) is defined to be the union of all subsets of $\mathcal{U}$ whose intersection with $A$ is non-empty. If $\mathcal{U}$ and $\mathcal{V}$ are two coverings of $X$, then we say that $\mathcal{V}$ **star refines** $\mathcal{U}$ (or is a star-refinement of $\mathcal{U}$) if for each $U \in \mathcal{U}$ there is a $V$ in $\mathcal{V}$ so that $\mathrm{st}(V;\mathcal{V}) \subset \mathcal{U}$. A **normal sequence** in $X$ is a sequence $(\mathcal{U}_\backslash)$ of open covers so that for each $n$, $\mathcal{U}_{\backslash+\infty}$ star refines $\mathcal{U}_\backslash$. *The* example of such a sequence is the case where $\mathcal{U}_\backslash$ is the family of open balls with radius $2^{-n}$ in a metric (or pseudo-metric) space. We shall now show that this is essentially the *only* case by showing how to use such a sequence to construct a pseudo-metric.

# 10 Compactology

We now discuss briefly a class of spaces whose relationship with the compact spaces is dual to that between the uniform spaces and metric spaces. Although the definition may seem rather artificial at first sight, there are a number of good reasons for regarding these spaces as the natural framework for many applications of point set topology.

**Definition:** A **compactology** on a set $X$ is a family $\mathcal{K}$ of subsets whose union is all of $X$ and which is closed under the formation of finite unions. We also assume the existence of a family $\{\tau_K : K \in \mathcal{K}\}$, where $\tau_K$ is a compact topology on $K$ so that the following condition is satisfied:

if $K$ is a subset of $L$, $(K, L \in \mathcal{K})$, then $\tau_L$ induces $\tau_K$ on $K$.
It is no loss of generality then to assume that if $K$ is in $\mathcal{K}$, so is each closed subspace of $K$. A **compactological space** is a set $X$ together with a compactology $\mathcal{K}$. If $(X, \mathcal{K})$ and $(Y, \mathcal{L})$ are two such spaces, a mapping $f$ from $X$ into $Y$ is **continuous** if for each $K \in \mathcal{K}$, there is an $L \in \mathcal{L}$ so that $f$ maps $K$ into $L$ and its restriction to $K$ is $\tau_K - \tau_L$ continuous. The concept of a (compactological) homeomorphism is defined in the natural way.

Of course, each Hausdorff space $(X, \tau)$ carries a natural compactological structure. One simply takes the family of all compact subsets. On the other hand, if $(X, \mathcal{K})$ is a compactological space, we can define a topology on $X$ as follows: the family $\{K : K \in \mathcal{K}\}$ forms an inductive system of topological spaces (ordered by inclusion) whose set-theoretical inductive limit is just $X$. Hence we can regard the latter as a topological space with the corresponding inductive limit structure. More concretely, we define a subset $U$ of $X$ to be open if $U \cap K$ is $\tau_K$-open for each $K \in \mathcal{K}\mathcal{L}$.

These two constructions show that there is a close connection between the concepts of topological spaces and compactological spaces and that we can go back and forth between the two types of structure with comparative ease. Nevertheless, they are not identical and the following points should be noted:

- if we start with a compactological space, the corresponding topological space need not have good separation properties, in particular, it need not be $T_2$.
- the constructions topological space $\to$ compactological space $\to$ topological space will not necessarily lead back to the original topology.
- if we consider two compactological spaces $X$ and $Y$ with a mapping $f$ between them, then $f$ is continuous (in the sense of the compactologies) if and only if it is continuous for the associated topologies. However, if we start with two topological spaces, then it will be easier for $f$ to be continuous in the compactological sense than in the topological one. (Namely it suffices for the former that the restrictions of $f$ to the compacta of $X$ be continuous).

If we wish to rule out such pathologies we need only have recourse to the following definitions:

**Definition:** A topological space $(X, \tau)$ is called a **Kelley space** if its topology coincides with the topology induced by its natural compactology, in other word, if a set $U$ is open if and only if $U \cap K$ is relatively open in $K$ for each compact subset $K$ of $X$. Such spaces are also called $k$- spaces, or compactly generated spaces.

A compactological space $(X, \mathcal{K})$ is said to be **regular** if the space $C(X)$ of continuous real-valued functions on $X$ separates $X$. It is easy to check that this is equivalent to the fact that there is a completely regular topology $\tau$ on $X$ which induces $\tau_K$ on each $K$. For if such a topology exists, then the functions in $C(X)$ which are continuous for this topology separate $X$ and hence so, *a fortiori*, does $C(X)$. On the other hand if $C(X)$ separates $X$, then the initial topology induced by $C(X)$ has the required property.

# 11 Function spaces and special types of mappings

In this section we shall consider structures on function spaces. At first glance, it is natural to start by looking for *topologies* on spaces of continuous functions between *topological* spaces and this has been the classical approach. However, this soon leads to difficulties and we shall introduce an approach which might seem rather artificial at first sight. We consider spaces of mappings from a uniform space into a compactological one and vice versa. The former will be provided with a compactological structure, the latter with a uniform one. We shall find that we can recover all of the useful classical function spaces from these ones.

We begin with the case of mappings from a compact space $K$ into a metric space $X$. The function space which we consider consists of the continuous or, equivalently, the uniformly continuous functions from $K$ into $X$. This space is denoted by $U(K; X)$ and is provided with the natural metric $D$ where $D(f, g) = \sup_{x \in K} d(f(x), g(x))$ under which it is complete if $X$ is. Convergence for this metric is just uniform convergence.

We now consider the general situation where $X$ is a compactological space and $Y$ is a uniform space which we assume to be complete for the moment. We are interested in the following space of mappings which we shall firstly define by means of the spectral representations of $X$ and $Y$. Indeed, if $X = \lim(K : K \in \mathcal{K})$ and $Y = \lim(Y_d : d \in \mathcal{D})$, then we define $U(X; Y)$ to be the projective limit of the spectrum

$$(U(K; Y_d) : K \in \mathcal{K}, \lceil \in \mathcal{D}).$$

Here the linking mappings are the natural ones i.e. if $K \subset K_1$ and $d_1 \geq d$, then the mapping from $U(K_1; Y_{d_1})$ to $U(K; Y_d)$ is defined by the composition

$$K \subset K_1 \to Y_{d_1} \to Y_d.$$

More formally, the linking mappings are

$$f \mapsto \pi_{d_1,d} \circ f \circ i_{K,K_1}.$$

We can identify $U(X;Y)$ in a less formal manner as follows. Each thread in the projective limit defines in the usual way a mapping from $X$ into $Y$ and these mappings are such that their restrictions to the compacta of $X$ are continuous (and so uniformly continuous). On the other hand, any such mapping from $X$ into $Y$ defines a thread. Hence the space $U(X;Y)$ consists precisely of all such mappings from $X$ into $Y$. Of course, these are just the mappings from $X$ into $Y$ which are continuous for the corresponding topologies induced by the compactology and the uniform structure respectively. We resume some of the simple properties of this space in the following:

The more formal definition of $U(X;Y)$ as a projective limit has the advantage of automatically providing it with structure as the limit of a spectrum of metric spaces. It is then natural (not to say unavoidable) to regard it as a uniform space with the initial structure. It is then automatically complete.

**Remark** If one should ever have to deal with spaces of mappings into non-complete spaces, this can be done easily as follows. The space $U(X;Y)$ is defined to be the subspace of $U(X;\hat{Y})$ consisting of those functions in the latter which take their values in $Y$. This then inherits a uniform structure from $U(X;\hat{Y})$.

Suppose now that $X$ is a metric space and that $K$ is compact. We denote by $U(X;K)$ the space of uniformly continuous functions from $X$ into $K$. We commence by regarding this as a topological space, imposing the topology $\tau_s$ of pointwise convergence on $X$ (sometimes called the **simple topology**). Since we intend to regard $U(X;K)$ as a compactological space we begin by characterising the $\tau_s$- compact subsets.

**Definition:** If $X$ and $K$ are as above, then a family of uniformly continuous mappings from $X$ into $K$ is equi-uniformly continuous if for each continuous function $f$ on $K$, the family $f \circ M$ of mappings in $C(K)$ is equicontinuous i.e. for each positive $\epsilon$ there is a $\delta > 0$ so that if $d(x,y) < \delta$, then $|f \circ g(x) - f \circ g(y)| < \epsilon$ for each $g$ in $M$.

**Proposition 100**

- *$U(X,K)$ is a closed subspace of $U(X,\tilde{K})$;*
- *a family $M \subset U(X;K)$ is $\tau_s$-compact if and only if it is compact in $U(X,\tilde{K})$;*
- *a family $M \subset U(X;K)$ is equi-uniformly continuous if and only if*

**Proposition 101** *(Ascoli) Let $M$ be a subset of $U(X,K)$. Then $M$ is $\tau_s$-compact if and only if it is closed and equi-uniformly continuous.*

PROOF. By the Lemma, we can reduce to the case where $K = I$ which is classical.

∎

We are now in a position to define a compactology on $U(X;K)$ as follows; the compact sets are the closed, equi-uniformly continuous sets, the compact topology is $\tau_s$ (or, equivalently, that of uniform convergence on the compact sets of $X$). Of course, this is just the natural compactology induced by $\tau_s$.

For the general case, we suppose that $(X,\mathcal{D})$ is a complete uniform space and that $(Y,\mathcal{K})$ is a compactological space with $X = \lim_{d\in\mathcal{D}}(\hat{X}_d,\hat{d})$, $Y = \lim_{K\in\mathcal{K}}(K)$. Then we define

$$U(X;Y) = \lim(U(\hat{X}_d;K) : d \in \mathcal{D}, \mathcal{K} \in \mathcal{K})$$

In other words, a mapping $f$ from $X$ into $Y$ belongs to $U(X;Y)$ if there is a $d \in \mathcal{D}$ and a $K \in \mathcal{K}$ so that $f$ factors as follows:

We compare briefly these structures with some classical ones in the following examples:

**Examples:** I. If $X$ is a set and $Y$ is a topological space, then $\mathcal{F}(\mathcal{X};\mathcal{Y})$ denotes the set of *all* mappings form $X$ into $Y$. It is customary to regard this as a topological space with the topology $\tau_s$ of pointwise convergence on $X$ i.e. the simple topology. In terms of nets this means that a net $(f_\alpha)$ of functions converges to a function $f$ if and only if for each $x \in X$, $(f_\alpha(x))$ converges to $f(x)$. If we suppose that $Y$ is uniformisable, then this corresponds to the structure $U(X;Y)$ above, when we regard $X$ as a compactological space by defining its finite subsets (with the discrete topology) to be the compacta. (Compactological spaces of this type are called **discrete**). To be more exact, the above topology is the topology corresponding to this uniform structure on $U(X;Y)$.

In applications, one works with subsets of $\mathcal{F}(\mathcal{X};\mathcal{Y})$. For example, if $X$ is a topological space, then it is natural to consider the set $C(X,Y)$ of continuous functions from $X$ into $Y$. In general, this will not be complete but we shall see shortly that some natural "small" subsets of this space *are*,in fact, complete with respect to the uniformity which corresponds to $\tau_s$.
II. If $K$ is the compact subset $\{\frac{1}{n} : n \in \mathbf{N}\} \cup \{0\}$ of the real line and $X$ is a metric space, then $U(K;X)$ can be identified with the space of convergent sequences in $M$.
III. The space $\mathbf{N}^{\mathbf{N}}$ of irrational numbers which has often been referred to above can be regarded as the function space $U(\mathbf{N},\mathbf{N})$ where the first $\mathbf{N}$ is

treated as a compactology (with the discrete compactology) and the second one as a metric space (with the discrete metric). Other spaces which can be described in an analogous manner are
IV. The space $U(I_d, I)$ of all functions (not necessarily continuous) from the unit interval into itself. (We have used the notation $I_d$ to denote $I$ with the discrete compactology).

Function spaces with the unit interval either as domain or range are important in applications.