S. W. Drury
McGill University

Notes on Metric Spaces

Contents

1 Normed and Metric Spaces **1**
1.1 Some Norms on Euclidean Space . . . 2
1.2 Inner Product Spaces . . . 2
1.3 Geometry of Norms . . . 5
1.4 Metric Spaces . . . 7

2 Topology of Metric Spaces **10**
2.1 Neighbourhoods and Open Sets . . . 10
2.2 Convergent Sequences . . . 11
2.3 Continuity . . . 15
2.4 Compositions of Functions . . . 17
2.5 Product Spaces and Mappings . . . 18
2.6 The Diagonal Mapping and Pointwise Combinations . . . 19
2.7 Interior and Closure . . . 22
2.8 Limits in Metric Spaces . . . 23
2.9 Distance to a Subset . . . 24
2.10 Separability . . . 25
2.11 Relative Topologies . . . 27
2.12 Uniform Continuity . . . 30
2.13 Subsequences . . . 31

3 A Metric Space Miscellany **32**
3.1 The p-norms on $\mathbb{R}^n$ . . . 32
3.2 Minkowski's Inequality and convexity . . . 35
3.3 The sequence spaces ℓ^p . . . 37
3.4 Premetrics . . . 38
3.5 Operator Norms . . . 41
3.6 Continuous Linear Forms . . . 42
3.7 Equivalent Metrics . . . 44
3.8 The Abstract Cantor Set . . . 45
3.9 The Quotient Norm . . . 46

4 Completeness **48**
4.1 Boundedness and Uniform Convergence . . . 49
4.2 Subsets and Products of Complete Spaces . . . 52
4.3 Contraction Mappings . . . 54
4.4 Extension by Uniform Continuity . . . 55
4.5 Completions . . . 57

4.6 Extension of Continuous Functions . . . 59
4.7 Baire's Theorem . . . 60
4.8 Complete Normed Spaces . . . 61

5 Compactness **64**
5.1 Compact Subsets . . . 64
5.2 The Finite Intersection Property . . . 66
5.3 Other Formulations of Compactness . . . 66
5.4 Preservation of Compactness by Continuous Mappings . . . 69
5.5 Compactness and Uniform Continuity . . . 71
5.6 Compactness and Uniform Convergence . . . 73
5.7 Equivalence of Compactness and Sequential Compactness . . . 74
5.8 Compactness and Completeness . . . 76
5.9 Equicontinuous Sets . . . 77
5.10 The Stone–Weierstrass Theorem . . . 78

6 Connectedness **85**
6.1 Connected Subsets . . . 86
6.2 Connectivity of the Real Line . . . 87
6.3 Connected Components . . . 87
6.4 Compactness and Connectedness . . . 90
6.5 Preservation of Connectedness by Continuous Mappings . . . 90
6.6 Path Connectedness . . . 91
6.7 Separation Theorem for Convex Sets . . . 93

1

Normed and Metric Spaces

We start by introducing the concept of a ***norm***. This generalization of the absolute value on $\mathbb{R}$ (or $\mathbb{C}$) to the framework of vector spaces is central to modern analysis.

The zero element of a vector space V (over $\mathbb{R}$ or $\mathbb{C}$) will be denoted 0_V. For an element v of the vector space V the norm of v (denoted $\|v\|$) is to be thought of as the distance from 0_V to v, or as the "size" of v. In the case of the absolute value on the field of scalars, there is really only one possible candidate, but in vector spaces of more than one dimension a wealth of possibilities arises.

DEFINITION *A **norm** on a vector space V over $\mathbb{R}$ or $\mathbb{C}$ is a mapping*

$$v \longrightarrow \|v\|$$

from V to $\mathbb{R}^+$ with the following properties.

- $\|0_V\| = 0$.
- $v \in V, \|v\| = 0 \Rightarrow v = 0_V$.
- $\|tv\| = |t|\|v\| \qquad \forall t$ *a scalar*, $v \in V$.
- $\|v_1 + v_2\| \le \|v_1\| + \|v_2\| \qquad \forall v_1, v_2 \in V$.

The last of these conditions is called the ***subadditivity inequality***. There are really two definitions here, that of a ***real norm*** applicable to real vector spaces and that of a ***complex norm*** applicable to complex vector spaces. However, every complex vector space can also be considered as a real vector space — one simply "forgets" how to multiply vectors by complex scalars that are not real scalars. This process is called ***realification***. In such a situation, the two definitions are different. For instance,

$$\|x + iy\| = \max(|x|, 2|y|) \qquad (x, y \in \mathbb{R})$$

defines a perfectly good real norm on $\mathbb{C}$ considered as a real vector space. On the other hand, the only complex norms on $\mathbb{C}$ have the form

$$\|x + iy\| = t(x^2 + y^2)^{\frac{1}{2}}$$

for some $t > 0$.

The inequality

$$\|t_1v_1 + t_2v_2 + \cdots + t_nv_n\| \le |t_1|\|v_1\| + |t_2|\|v_2\| + \cdots + |t_n|\|v_n\|$$

holds for scalars $t_1, \ldots, t_n$ and elements $v_1, \ldots, v_n$ of V. It is an immediate consequence of the definition.

If $\| \ \|$ is a norm on V and $t > 0$ then

$$|||v||| = t\|v\|$$

defines a new norm $||| \ |||$ on V. We note that in the case of a norm there is often no natural way to normalize it. On the other hand, an absolute value is normalized so that $|1| = 1$, possible since the field of scalars contains a distinguished element 1.

1.1 Some Norms on Euclidean Space

Because of the central role of $\mathbb{R}^n$ as a vector space it is worth looking at some of the norms that are commonly defined on this space.

Example On $\mathbb{R}^n$ we may define a norm by

$$\|(x_1, \ldots, x_n)\|_\infty = \max_{j=1}^{n} |x_j|. \tag{1.1}$$

□

Example Another norm on $\mathbb{R}^n$ is given by

$$\|(x_1, \ldots, x_n)\|_1 = \sum_{j=1}^{n} |x_j|.$$

□

Example The ***Euclidean norm*** on $\mathbb{R}^n$ is given by

$$\|(x_1, \ldots, x_n)\|_2 = \left(\sum_{j=1}^{n} |x_j|^2 \right)^{\frac{1}{2}}.$$

This is the standard norm, representing the standard Euclidean distance to $\mathbf{0}$. The symbol $\mathbf{0}$ will be used to denote the zero vector of $\mathbb{R}^n$ or $\mathbb{C}^n$. □

Later we will generalize these examples by defining in case $1 \le p < \infty$

$$\|(x_1, \ldots, x_n)\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{\frac{1}{p}}.$$

In case that $p = \infty$ we use (1.1) to define $\| \ \|_\infty$. It will be shown (on page 34) that $\| \ \|_p$ is a norm.

1.2 Inner Product Spaces

Inner product spaces play a very central role in analysis. In this section we only scratch the surface of the subject.

Definition *A **real inner product space** is a real vector space V together with an inner product. An **inner product** is a mapping from $V \times V$ to $\mathbb{R}$ denoted by*

$$(v_1, v_2) \longmapsto \langle v_1, v_2 \rangle$$

and satisfying the following properties

- $\langle t_1 v_1 + t_2 v_2, w\rangle = t_1\langle v_1, w\rangle + t_2\langle v_2, w\rangle \qquad \forall w, v_1, v_2 \in V, t_1, t_2 \in \mathbb{R}$.
- $\langle v_1, v_2\rangle = \langle v_2, v_1\rangle \qquad \forall v_1, v_2 \in V$.
- $\langle v, v\rangle \geq 0 \qquad \forall v \in V$.
- *If* $v \in V$ *and* $\langle v, v\rangle = 0$, *then* $v = 0_V$.

The symmetry and the linearity in the first variable implies that the inner product is also linear in the second variable.

$$\langle w, t_1 v_1 + t_2 v_2\rangle = t_1\langle w, v_1\rangle + t_2\langle w, v_2\rangle \qquad \forall w, v_1, v_2 \in V, t_1, t_2 \in \mathbb{R}.$$

EXAMPLE The ***standard inner product*** on $\mathbb{R}^n$ is given by

$$\langle x, y\rangle = \sum_{j=1}^{n} x_j y_j$$

□

The most general inner product on $\mathbb{R}^n$ is given by

$$\langle x, y\rangle = \sum_{j=1}^{n}\sum_{k=1}^{n} p_{j,k} x_j y_k$$

where the $n \times n$ real matrix $P = (p_{j,k})$ is a ***positive definite*** matrix. This means that

- P is a symmetric matrix.
- We have
$$\sum_{j=1}^{n}\sum_{k=1}^{n} p_{j,k} x_j x_k \geq 0$$
for every vector $(x_1, \ldots, x_n)$ of $\mathbb{R}^n$.
- The circumstance
$$\sum_{j=1}^{n}\sum_{k=1}^{n} p_{j,k} x_j x_k = 0$$
only occurs when $x_1 = 0, \ldots, x_n = 0$.

In the complex case, the definition is slightly more complicated.

DEFINITION *A* ***complex inner product space*** *is a complex vector space* V *together with a* ***complex inner product****, that is a mapping from* $V \times V$ *to* $\mathbb{C}$ *denoted*

$$(v_1, v_2) \Leftrightarrow \langle v_1, v_2\rangle$$

and satisfying the following properties

- $\langle t_1 v_1 + t_2 v_2, w\rangle = t_1\langle v_1, w\rangle + t_2\langle v_2, w\rangle \qquad \forall w, v_1, v_2 \in V, t_1, t_2 \in \mathbb{C}$.
- $\langle v_1, v_2\rangle = \overline{\langle v_2, v_1\rangle} \qquad \forall v_1, v_2 \in V$.
- $\langle v, v\rangle \geq 0 \qquad \forall v \in V$.
- *If* $v \in V$ *and* $\langle v, v\rangle = 0$, *then* $v = 0_V$.

It will be noted that a complex inner product is conjugate linear in its second variable.

$$\langle w, t_1 v_1 + t_2 v_2\rangle = \overline{t_1}\langle w, v_1\rangle + \overline{t_2}\langle w, v_2\rangle \qquad \forall w, v_1, v_2 \in V, t_1, t_2 \in \mathbb{C}.$$

EXAMPLE The standard inner product on $\mathbb{C}^n$ is given by

$$\langle x, y \rangle = \sum_{j=1}^{n} x_j \overline{y_j}$$

□

The most general inner product on $\mathbb{C}^n$ is given by

$$\langle x, y \rangle = \sum_{j=1}^{n} \sum_{k=1}^{n} p_{j,k} x_j \overline{y_k}$$

where the $n \times n$ complex matrix $P = (p_{j,k})$ is a ***positive definite*** matrix. This means that

- P is a hermitian matrix, in other words $p_{jk} = \overline{p_{kj}}$.
- We have
$$\sum_{j=1}^{n} \sum_{k=1}^{n} p_{j,k} x_j \overline{x_k} \geq 0$$
for every vector $(x_1, \ldots, x_n)$ of $\mathbb{C}^n$.
- The circumstance
$$\sum_{j=1}^{n} \sum_{k=1}^{n} p_{j,k} x_j x_k = 0$$
only occurs when $x_1 = 0, \ldots, x_n = 0$.

DEFINITION *Let V be an inner product space. Then we define*

$$||v|| = (\langle v, v \rangle)^{\frac{1}{2}} \tag{1.2}$$

the ***associated norm****.*

It is not immediately clear from the definition that the associated norm satisfies the subadditivity condition. Towards this, we establish the abstract Cauchy-Schwarz inequality.

PROPOSITION 1.1 (CAUCHY-SCHWARZ INEQUALITY) *Let $u, v \in V$ an inner product space. Then*

$$|\langle u, v \rangle| \leq ||u|| ||v|| \tag{1.3}$$

holds.

Proof of the Cauchy-Schwarz Inequality. We give the proof in the complex case. The proof in the real case is slightly easier. If $v = 0_V$ then the inequality is evident. We therefore assume that $||v|| > 0$. Similarly, we may assume that $||u|| > 0$.

Let $t \in \mathbb{C}$. Then we have

$$\begin{aligned} 0 \leq ||u + tv||^2 &= \langle u + tv, u + tv \rangle \\ &= \langle u, u \rangle + t\langle v, u \rangle + \overline{t}\langle u, v \rangle + t\overline{t}\langle v, v \rangle \\ &= ||u||^2 + 2\Re\overline{t}\langle u, v \rangle + |t|^2 ||v||^2. \end{aligned} \tag{1.4}$$

Now choose t such that

$$\bar{t}\langle u, v\rangle \text{ is real and } \leq 0 \tag{1.5}$$

and

$$|t| = \frac{||u||}{||v||}. \tag{1.6}$$

Here, (1.6) designates the absolute value of t and (1.5) specifies its argument. Substituting back into (1.4) we obtain

$$2\frac{||u||}{||v||}|\langle u, v\rangle| \leq ||u||^2 + \left(\frac{||u||}{||v||}\right)^2 ||v||^2$$

which simplifies to the desired inequality (1.3). ■

PROPOSITION 1.2 *In an inner product space* (1.2) *defines a norm.*

Proof. We verify the subadditivity of $v \longleftrightarrow ||v||$. The other requirements of a norm are straightforward to establish. We have

$$\begin{aligned} ||u+v||^2 &= \langle u+v, u+v\rangle \\ &= ||u||^2 + 2\Re\langle u, v\rangle + ||v||^2 \\ &\leq ||u||^2 + 2||u||||v|| + ||v||^2 \\ &= (||u|| + ||v||)^2 \end{aligned} \tag{1.7}$$

using the Cauchy-Schwarz Inequality (1.3). Taking square roots yields

$$||u+v|| \leq ||u|| + ||v||$$

as required. ■

1.3 Geometry of Norms

It is possible to understand the concept of norm from the geometrical point of view. Towards this we associate with each norm a geometrical object — its unit ball.

DEFINITION *Let V be a normed vector space. Then the* ***unit ball*** *B of V is defined by*

$$B = \{v; v \in V, ||v|| \leq 1\}.$$

DEFINITION *Let V be a vector space and let $B \subseteq V$. We say that B is* ***convex*** *iff*

$$t_1v_1 + t_2v_2 \in B \qquad \forall v_1, v_2 \in B,\ \forall t_1, t_2 \geq 0 \text{ such that } t_1 + t_2 = 1.$$

In other words, a set B is convex iff whenever we take two points of B, the line segment joining them lies entirely in B.

Definition *Let V be a vector space and let $B \subseteq V$. We say that B satisfies the* ***line condition*** *iff for every $v \in V \setminus \{0_V\}$, there exists a constant $a \in]0, \infty[$ such that*

$$tv \in B \quad \Leftrightarrow \quad |t| \leq a.$$

Thus, the line condition says that the intersection of B with every one-dimensional subspace R of V is the unit ball for some norm on R. The line condition involves a multitude of considerations. It implies that the set B is ***symmetric*** about the zero element. The fact that $a > 0$ is sometimes expressed by saying that B is ***absorbing***. This expresses the fact that every point v of V lies in some (large) multiple of B. Finally the fact that $a < \infty$ is a ***boundedness condition***.

Theorem 1.3 *Let V be a vector space and let $B \subseteq V$. Then the following two statements are equivalent.*

- *There is a norm on V for which B is the unit ball.*
- *B is convex and satisfies the line condition.*

Proof. We assume first that the first statement holds and establish the second. Let $v_1, v_2 \in B$ and let $t_1, t_2 > 0$ be such that $t_1 + t_2 = 1$. Then

$$\begin{aligned} \|t_1 v_1 + t_2 v_2\| &\leq \|t_1 v_1\| + \|t_2 v_2\| \\ &\leq |t_1|\|v_1\| + |t_2|\|v_2\| \\ &\leq t_1 + t_2 = 1. \end{aligned}$$

It follows that B is convex. Now let $v \in V$ and suppose that $v \neq 0_V$. Then it is straightforward to show that the line condition holds with $a = \|v\|^{-1}$.

The real meat of the Theorem is contained in the converse to which we now turn. Let B be a convex subset of V satisfying the line condition. We define for $v \in V \setminus \{0_V\}$

$$\|v\| = a^{-1}$$

where a is the constant of the line condition. We also define $\|0_V\| = 0$. We aim to show that $\| \ \|$ is a norm and that B is its unit ball. Let $v \neq 0_V$ and $s \neq 0$. Then, applying the line condition to V and sv we have constants a and b with $\|v\| = a^{-1}$ and $\|sv\| = b^{-1}$ such that

$$tv \in B \quad \Leftrightarrow \quad |t| \leq a$$

and

$$r(sv) \in B \quad \Leftrightarrow \quad |r| \leq b.$$

Substituting $t = rs$ we find that

$$|rs| \leq a \quad \Leftrightarrow \quad |r| \leq b$$

so that $a = b|s|$. It now follows that

$$\|sv\| = b^{-1} = |s|a^{-1} = |s|\|v\|. \tag{1.8}$$

On the other hand if $s = 0$ or if $v = 0_V$, then (1.8) also holds.

We turn next to the subadditivity of the norm. Let v_1 and v_2 be non-zero vectors in V. Let $t_1 = \|v_1\|$ and $t_2 = \|v_2\|$. Then $t_1^{-1} v_1 \in B$ and $t_2^{-1} v_2 \in B$. Hence, we find that

$$\begin{aligned} v_1 + v_2 &= t_1 t_1^{-1} v_1 + t_2 t_2^{-1} v_2 \\ &= (t_1 + t_2) \left(\frac{t_1}{t_1 + t_2} t_1^{-1} v_1 + \frac{t_2}{t_1 + t_2} t_2^{-1} v_2 \right) \\ &= (t_1 + t_2) v \end{aligned}$$

where $v \in B$ by the convexity of B. If $v_1 + v_2 \neq 0_V$ we have the desired conclusion

$$\|v_1 + v_2\| \leq t_1 + t_2 = \|v_1\| + \|v_2\| \tag{1.9}$$

by the definition of the $\|\ \|$. If $v_1 + v_2 = 0_V$, then (1.9) follows trivially. We also observe that (1.9) follows if either v_1 or v_2 vanishes. The remaining properties of the norm follow directly from the definition.

It is routine to check that for $v \in V \setminus \{0_V\}$

$$\|v\| \leq 1 \quad \Leftrightarrow \quad v \in B.$$

and both sides are true if $v = 0_V$. It follows that B is precisely the unit ball of $\|\ \|$. ■

EXAMPLE Let us define the a subset B of $\mathbb{R}^2$ by

$$(x, y) \in B \text{ if } \begin{cases} x^2 + y^2 \leq 1 & \text{in case } x \geq 0 \text{ and } y \geq 0, \\ \max(\Leftrightarrow x, y) \leq 1 & \text{in case } x \leq 0 \text{ and } y \geq 0, \\ x^2 + y^2 \leq 1 & \text{in case } x \leq 0 \text{ and } y \leq 0, \\ \max(x, \Leftrightarrow y) \leq 1 & \text{in case } x \geq 0 \text{ and } y \leq 0. \end{cases}$$

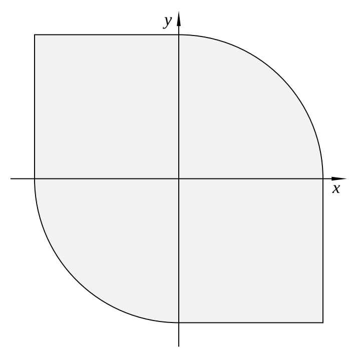

FIGURE 1: *The unit ball for a norm on* $\mathbb{R}^2$.

It is geometrically obvious that B is a convex subset of $\mathbb{R}^2$ and satisfies the line condition — see Figure 1. Therefore it defines a norm. Clearly this norm is given by

$$\|(x, y)\| = \begin{cases} (x^2 + y^2)^{\frac{1}{2}} & \text{if } x \geq 0 \text{ and } y \geq 0, \\ \max(|x|, |y|) & \text{if } x \leq 0 \text{ and } y \geq 0, \\ (x^2 + y^2)^{\frac{1}{2}} & \text{if } x \leq 0 \text{ and } y \leq 0, \\ \max(|x|, |y|) & \text{if } x \geq 0 \text{ and } y \leq 0. \end{cases}$$

□

1.4 Metric Spaces

In the previous section we discussed the concept of the norm of a vector. In a normed vector space, the expression $\|u \Leftrightarrow v\|$ represents the size of the difference $u \Leftrightarrow v$ of two vectors u and v. It can be thought of as the distance between u and v. Just as a vector space may have many possible norms, there can be many possible concepts of distance.

In this section we introduce the concept of a ***metric space***. A metric space is simply a set together with a distance function which measures the distance between any two points of the space. While normed spaces give interesting examples of metric spaces, there are many interesting examples of metric spaces that do not come from norms.

DEFINITION *A **metric space** (X, d) is a set X together with a **distance function** or **metric** $d : X \times X \longrightarrow \mathbb{R}^+$ satisfying the following properties.*

- $d(x, x) = 0 \qquad \forall x \in X.$
- $x, y \in X, d(x, y) = 0 \quad \Rightarrow \quad x = y.$
- $d(x, y) = d(y, x) \qquad \forall x, y \in X.$
- $d(x, z) \leq d(x, y) + d(y, z) \qquad \forall x, y, z \in X.$

The fourth axiom for a distance function is called the ***triangle inequality***. It is easy to derive the ***extended triangle inequality***

$$d(x_1, x_n) \leq d(x_1, x_2) + d(x_2, x_3,) + \cdots + d(x_{n-1}, x_n) \quad \forall x_1, \ldots, x_n \in X \tag{1.10}$$

directly from the axioms.

Sometimes we will abuse notation and say that X is a metric space when the intended distance function is understood.

Let X be a metric space and let $Y \subseteq X$. Then the restriction of the distance function of X to the subset $Y \times Y$ of $X \times X$ is a distance function on Y. Sometimes this is called the ***restriction metric*** or the ***relative metric***. If the four axioms listed above hold for all points of X then *a fortiori* they hold for all points of Y. Thus every subset of a metric space is again a metric space in its own right. This idea will be used very frequently in the sequel.

EXAMPLE Let V be a normed vector space with norm $\| \ \|$. Then V is a metric space with the distance function

$$d(u, v) = \|u - v\|.$$

The reader should check that the triangle inequality is a consequence of the subadditivity of the norm. □

EXAMPLE As an example of an infinite dimensional normed vector space we consider the space ℓ^∞. Its elements are the bounded real sequences (x_n) and the norm is defined by

$$\|(x_n)\|_\infty = \sup_{n \in \mathbb{N}} |x_n|.$$

□

EXAMPLE Another example of an infinite dimensional normed vector space is the space ℓ^1. Its elements are the absolutely summable real sequences (x_n) and the norm is defined by

$$\|(x_n)\|_1 = \sum_{n=1}^{\infty} |x_n|.$$

□

EXAMPLE It follows that every subset X of a normed vector space is a metric space in the distance function induced from the norm. □

Example Let $\langle\ ,\ \rangle$ and $\|\ \|$ denote the standard inner product and Euclidean norm on $\mathbb{R}^n$. Let S^{n-1} denote the unit sphere

$$S^{n-1} = \{x; x \in \mathbb{R}^n, \|x\| = 1\}$$

then we can define the ***geodesic distance*** between two points x and y of S^{n-1} by

$$d(x, y) = \arccos(\langle x, y\rangle). \tag{1.11}$$

We will show that d is a metric on S^{n-1}. This metric is of course different from the Euclidean distance $\|x \Leftrightarrow y\|$.

To verify that (1.11) is in fact a metric, at least the symmetry of the metric is evident. Suppose that $x, y \in S^{n-1}$ and that $d(x, y) = 0$. Then $\langle x, y\rangle = 1$ and

$$\|x \Leftrightarrow y\|^2 = \|x\|^2 \Leftrightarrow 2\langle x, y\rangle + \|y\|^2 = 1 \Leftrightarrow 2 + 1 = 0.$$

It follows that $x = y$.

To establish the triangle inequality, let $x, y, z \in S^{n-1}$, $\theta = \arccos(\langle x, y\rangle)$ and $\varphi = \arccos(\langle y, z\rangle)$. Then we can write $x = \cos\theta\, y + \sin\theta\, u$ and $z = \cos\varphi\, y + \sin\varphi\, v$ where u and v are unit vectors orthogonal to y. An easy calculation now gives

$$\langle x, z\rangle = \cos\theta\ \cos\varphi + \langle u, v\rangle \sin\theta\ \sin\varphi.$$

Now, since $0 \le \theta, \varphi \le \pi$, we have $\sin\theta\ \sin\varphi \ge 0$. By the Cauchy-Schwarz Inequality (1.3), we find that $\langle u, v\rangle \ge \Leftrightarrow 1$. Hence

$$\langle x, z\rangle \ge \cos\theta\ \cos\varphi \Leftrightarrow \sin\theta\ \sin\varphi = \cos(\theta + \varphi).$$

Since arccos is decreasing on $[\Leftrightarrow 1, 1]$ this immediately yields

$$d(x, z) \le \theta + \varphi = d(x, y) + d(y, z).$$

$\square$

2

Topology of Metric Spaces

2.1 Neighbourhoods and Open Sets

It is customary to refer to the elements of a metric space as ***points***. In this chapter we will develop the ***point-set topology*** of metric spaces. This is done through concepts such as ***neighbourhoods***, ***open sets***, ***closed sets*** and ***sequences***. Any of these concepts can be used to define more advanced concepts such as the continuity of mappings from one metric space to another. They are, as it were, languages for the further development of the subject. We study them all and most particularly the relationships between them.

DEFINITION *Let (X, d) be a metric space. For $t > 0$ and $x \in X$, we define*

$$U(x,t) = \{y; y \in X, d(x,y) < t\}$$

and

$$B(x,t) = \{y; y \in X, d(x,y) \le t\}.$$

*the **open ball** $U(x,t)$ centred at x of radius t and the corresponding **closed ball** $B(x,t)$.*

DEFINITION *Let V be a subset of a metric space X and let $x \in V$. Then we say that V is a **neighbourhood** of x or x is an **interior point** of V iff there exists $t > 0$ such that $U(x,t) \subseteq V$.*

Thus V is a neighbourhood of x iff all points sufficiently close to x lie in V.

PROPOSITION 2.1

- *If V is a neighbourbood of x and $V \subseteq W \subseteq X$. Then W is a neighbourhood of x.*
- *If $V_1, V_2, \ldots, V_n$ are finitely many neighbourhoods of x, then $\cap_{j=1}^n V_j$ is also a neighbourhood of x.*

Proof. The first statement is left as an exercise for the reader. For the second, applying the definition, we may find $t_1, t_2, \ldots, t_n > 0$ such that $U(x, t_j) \subseteq V_j$. It follows that

$$\bigcap_{j=1}^{n} U(x,t_j) \subseteq \bigcap_{j=1}^{n} V_j. \tag{2.1}$$

But the left-hand side of (2.1) is just $U(x,t)$ where $t = \min t_j > 0$. It now follows that $\cap_{j=1}^n V_j$ is a neighbourhood of x. ∎

Neighbourhoods are a local concept. We now introduce the corresponding global concept.

Definition *Let (X,d) be a metric space and let $V \subseteq X$. Then V is an* ***open subset*** *of X iff V is a neighbourhood of every point x that lies in V.*

Example For all $t > 0$, the open ball $U(x,t)$ is an open set. To see this, let $y \in U(x,t)$, that is $d(x,y) < t$. We must show that $U(x,t)$ is a neighbourhood of y. Let $s = t - d(x,y) > 0$. We claim that $U(y,s) \subseteq U(x,t)$. To prove the claim, let $z \in U(y,s)$. Then $d(y,z) < s$. We now find that

$$d(x,z) \leq d(x,y) + d(y,z) < d(x,y) + s = t,$$

so that $z \in U(x,t)$ as required. □

Example In $\mathbb{R}$ every interval of the form $]a,b[$ is an open set. Here, a and b are real and satisfy $a < b$. We also allow the possibilities $a = -\infty$ and $b = \infty$. □

Theorem 2.2 *In a metric space (X,d) we have*

- *X is an open subset of X.*
- *$\emptyset$ is an open subset of X.*
- *If V_α is open for every α in some index set I, then $\cup_{\alpha \in I} V_\alpha$ is again open.*
- *If V_j is open for $j = 1, \ldots, n$, then the finite intersection $\cap_{j=1}^n V_j$ is again open.*

Proof. For every $x \in X$ and any $t > 0$, we have $U(x,t) \subseteq X$, so X is open. On the other hand, $\emptyset$ is open because it does not have any points. Thus the condition to be checked is vacuous.

To check the third statement, let $x \in \cup_{\alpha \in I} V_\alpha$. Then there exists $\alpha \in I$ such that $x \in V_\alpha$. Since V_α is open, V_α is a neighbourhood of x. The result now follows from the first part of Proposition 2.1.

Finally let $x \in \cap_{j=1}^n V_j$. Then since V_j is open, it is a neighbourhood of x for $j = 1, \ldots, n$. Now apply the second part of Proposition 2.1. ■

Definition *Let X be a set. Let $\mathcal{V}$ be a "family of open sets" satisfying the four conditions of Theorem 2.2. Then $\mathcal{V}$ is a* ***topology*** *on X and $(X, \mathcal{V})$ is a* ***topological space****.*

Not every topology arises from a metric. In these notes we are *not* concerned with topological spaces in their own right. For some applications topological spaces are needed to capture key ideas (like the weak$\star$ topology). On the other hand, some theorems true for general metric spaces are false for topological spaces (separation theorems for example). Finally some metric space concepts (such as uniform continuity) cannot be defined on topological spaces.

It is worth recording here that there is a complete description of the open subsets of $\mathbb{R}$. A subset V of $\mathbb{R}$ is open iff it is a disjoint union of open intervals (possibly of infinite length). Furthermore, such a union is necessarily countable.

2.2 Convergent Sequences

A sequence $x_1, x_2, x_3, \ldots$ of points of a set X is really a mapping from $\mathbb{N}$ to X. Normally, we denote such a sequence by (x_n). For $x \in X$ the sequence given by $x_n = x$ is called the ***constant sequence*** with value x.

Definition *Let X be a metric space. Let (x_n) be a sequence in X. Then (x_n)* ***converges to*** *$x \in X$ iff for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \epsilon$ for all $n > N$. In this case, we write $x_n \longrightarrow x$ or*

$$x_n \xrightarrow{n \to \infty} x.$$

Sometimes, we say that x is the ***limit*** *of (x_n). Proposition 2.3 below justifies the use of the indefinite article. To say that (x_n) is a* **convergent sequence** *is to say that there exists some $x \in X$ such that (x_n) converges to x.*

EXAMPLE Perhaps the most familiar example of a convergent sequence is the sequence

$$x_n = \frac{1}{n}$$

in $\mathbb{R}$. This sequence converges to 0. To see this, let $\epsilon > 0$ be given. Then choose a natural number N so large that $N > \epsilon^{-1}$. It is easy to see that

$$n > N \qquad \Rightarrow \qquad \left|\frac{1}{n}\right| < \epsilon$$

Hence $x_n \Leftrightarrow\!\rightarrow 0$. □

PROPOSITION 2.3 *Let (x_n) be a convergent sequence in X. Then the limit is unique.*

Proof. Suppose that x and y are both limits of the sequence (x_n). We will show that $x = y$. If not, then $d(x,y) > 0$. Let us choose $\epsilon = \frac{1}{2}d(x,y)$. Then there exist natural numbers N_x and N_y such that

$$\begin{aligned} n > N_x \qquad &\Rightarrow \qquad d(x_n, x) < \epsilon, \\ n > N_y \qquad &\Rightarrow \qquad d(x_n, y) < \epsilon. \end{aligned}$$

Choose now $n = \max(N_x, N_y) + 1$ so that both $n > N_x$ and $n > N_y$. It now follows that

$$2\epsilon = d(x,y) \le d(x,x_n) + d(x_n,y) < \epsilon + \epsilon$$

a contradiction. ■

PROPOSITION 2.4 *Let X be a metric space and let (x_n) be a sequence in X. Let $x \in X$. The following conditions are equivalent to the convergence of (x_n) to x.*

- *For every neighbourhood V of x in X, there exists $N \in \mathbb{N}$ such that*

$$n > N \qquad \Rightarrow \qquad x_n \in V. \tag{2.2}$$

- *The sequence $(d(x_n, x))$ converges to 0 in $\mathbb{R}$.*

We leave the details of the proof to the reader. The first item here is significant because it leads to the concept of the ***tail*** of a sequence. The sequence (t_n) defined by $t_k = x_{N+k}$ is called the Nth tail sequence of (x_n). The set of points $T_N = \{x_n; n > N\}$ is the Nth tail set. The condition (2.2) can be rewritten as $T_N \subseteq V$.

Sequences provide one of the key tools for understanding metric spaces. They lead naturally to the concept of ***closed subsets*** of a metric space.

DEFINITION *Let X be a metric space. Then a subset $A \subseteq X$ is said to be* ***closed*** *iff whenever (x_n) is a sequence in A (that is $x_n \in A \quad \forall n \in \mathbb{N}$) converging to a limit x in X, then $x \in A$.*

The link between closed subsets and open subsets is contained in the following result.

THEOREM 2.5 *In a metric space X, a subset A is closed if and only if $X \setminus A$ is open.*

It follows from this Theorem that U is open in X iff $X \setminus U$ is closed.

Proof. First suppose that A is closed. We must show that $X \setminus A$ is open. Towards this, let $x \in X \setminus A$. We claim that there exists $\epsilon > 0$ such that $U(x, \epsilon) \subseteq X \setminus A$. Suppose not. Then taking for each $n \in \mathbb{N}$, $\epsilon_n = \frac{1}{n}$ we find that there exists $x_n \in A \cap U(x, \frac{1}{n})$. But now (x_n) is a sequence of elements of A converging to x. Since A is closed $x \in A$. But this is a contradiction.

For the converse assertion, suppose that $X \setminus A$ is open. We will show that A is closed. Let (x_n) be a sequence in A converging to some $x \in X$. If $x \in X \setminus A$ then since $X \setminus A$ is open, there exists $\epsilon > 0$ such that

$$U(x, \epsilon) \subseteq X \setminus A. \tag{2.3}$$

But since (x_n) converges to x, there exists $N \in \mathbb{N}$ such that $x_n \in U(x, \epsilon)$ for $n > N$. Choose $n = N + 1$. Then we find that $x_n \in A \cap U(x, \epsilon)$ which contradicts (2.3). ■

Combining now Theorems 2.2 and 2.5 we have the following corollary.

COROLLARY 2.6 *In a metric space (X, d) we have*

- *X is an closed subset of X.*
- *$\emptyset$ is an closed subset of X.*
- *If A_α is closed for every α in some index set I, then $\cap_{\alpha \in I} A_\alpha$ is again closed.*
- *If A_j is closed for $j = 1, \ldots, n$, then the finite union $\cup_{j=1}^{n} A_j$ is again closed.*

EXAMPLE In a metric space every singleton is closed. To see this we remark that a sequence in a singleton is necessarily a constant sequence and hence convergent to its constant value. □

EXAMPLE Combining the previous example with the last assertion of Corollary 2.6, we see that in a metric space, every finite subset is closed. □

EXAMPLE Let (x_n) be a sequence converging to x. Then the set

$$\{x_n; n \in \mathbb{N}\} \cup \{x\}$$

is a closed subset. □

EXAMPLE In $\mathbb{R}$, the intervals $[a, b]$, $[a, \infty[$ and $]-\infty, b]$ are closed subsets. □

EXAMPLE A more complicated example of a closed subset of $\mathbb{R}$ is the ***Cantor set***. There are several ways of describing the Cantor set. Let $E_0 = [0, 1]$. To obtain E_1 from E_0 we remove the middle third of E_0. Thus $E_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. To obtain E_2 from E_1 we remove the middle thirds from both the constituent intervals of E_1. Thus

$$E_2 = [0, \tfrac{1}{9}] \cup [\tfrac{2}{9}, \tfrac{1}{3}] \cup [\tfrac{2}{3}, \tfrac{7}{9}] \cup [\tfrac{8}{9}, 1].$$

Continuing in this way, we find that E_k is a union of 2^k closed intervals of length 3^{-k}. The Cantor set E is now defined as

$$E = \bigcap_{k=0}^{\infty} E_k.$$

By Corollary 2.6 it is clear that E is a closed subset of $\mathbb{R}$.

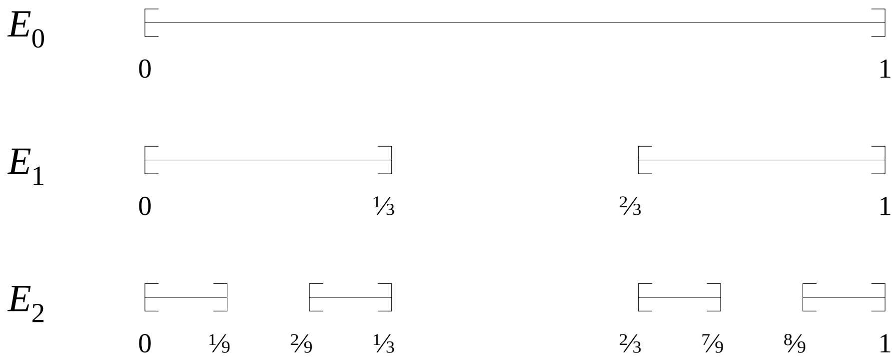

FIGURE 2: *The sets E_0, E_1 and E_2.*

The sculptor Rodin once said that to make a sculpture one starts with a block of marble and removes everything that is unimportant. This is the approach that we have just taken in building the Cantor set. The second way of constructing the Cantor set works by building the set from the inside out.

Let us define

$$K = \{\sum_{k=1}^{\infty} \omega_k 3^{-k}; \omega_k \in \{0, 2\}, k = 1, 2, \ldots\}.$$

A moment's thought shows us that the points $\sum_{k=1}^{n} \omega_k 3^{-k}$ given by the 2^n choices of ω_k for $k = 1, 2, \ldots, n$ are precisely the left hand endpoints of the 2^n constituent subintervals of E_n. Also a straightforward estimate on the tail sum

$$0 \leq \sum_{k=n+1}^{\infty} \omega_k 3^{-k} \leq \sum_{k=n+1}^{\infty} 2 \cdot 3^{-k} \leq 3^{-n},$$

shows that each sum $\sum_{k=1}^{\infty} \omega_k 3^{-k}$ lies in E_n for each $n \in \mathbb{N}$. It follows that $K \subseteq E$.

For the reverse inclusion, suppose that $x \in E$. Then for every $n \in \mathbb{N}$, let x_n be the left hand endpoint of the subinterval of E_n to which x belongs. Then

$$|x \Leftrightarrow x_n| \leq 3^{-n}. \tag{2.4}$$

We write

$$x_n = \sum_{k=1}^{n} \omega_k 3^{-k} \tag{2.5}$$

where ω_k takes one or other of the values 0 and 2. It is not difficult to see that the values of ω_k do not depend on the value n under consideration. Indeed, suppose that (2.5) holds for a specific value of n. Then $x \in [x_n, x_n + 3^{-n}]$. At the next step, we look to see whether x lies in the left hand third or the right hand third of this interval. This determines x_{n+1} by

$$x_{n+1} = x_n + \omega_{n+1} 3^{-(n+1)}$$

where $\omega_{n+1} = 0$ if it is the left hand interval and $\omega_{n+1} = 2$ if it is the right hand interval. The values of ω_k for $k = 1, 2, \ldots, n$ are not affected by this choice. It now follows from (2.5) and (2.4) that

$$x = \sum_{k=1}^{\infty} \omega_k 3^{-k}$$

so that $x \in K$ as required. □

2.3 Continuity

The primary purpose of the preceding sections is to define the concept of ***continuity*** of mappings. This concept is the mainspring of mathematical analysis.

Definition *Let X and Y be metric spaces. Let $f : X \to Y$. Let $x \in X$. Then f is* ***continuous at*** *x iff for all $\epsilon > 0$, there exists $\delta > 0$ such that*

$$z \in U(x, \delta) \qquad \Rightarrow \qquad f(z) \in U(f(x), \epsilon). \tag{2.6}$$

The $\forall \ldots \exists \ldots$ combination suggests the role of the "devil's advocate" type of argument. Let us illustrate this with an example.

Example The mapping $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is continuous at $x = 1$. To prove this, we suppose that the devil's advocate provides us with a number $\epsilon > 0$ chosen cunningly small. We have to "reply" with a number $\delta > 0$ (depending on ϵ) such that (2.6) holds. In the present context, we choose

$$\delta = \min(\tfrac{1}{4}\epsilon, 1)$$

so that for $|x - 1| < \delta$ we have

$$|x^2 - 1| \le |x - 1||x + 1| < (\tfrac{1}{4}\epsilon)(3) < \epsilon$$

since $|x - 1| < \delta$ and $|x + 1| = |(x - 1) + 2| \le |x - 1| + 2 < 3$. □

Example Continuity at a point — a single point that is, does not have much strength. Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q}, \\ x & \text{if } x \in \mathbb{Q}. \end{cases}$$

This function is continuous at 0 but at no other point of $\mathbb{R}$. □

Example An interesting contrast is provided by the function $g : \mathbb{R} \to \mathbb{R}$ given by

$$g(x) = \begin{cases} 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \text{ or if } x = 0, \\ \frac{1}{q} & \text{if } x = \frac{p}{q} \text{ where } p \in \mathbb{Z} \setminus \{0\},\ q \in \mathbb{N} \text{ are coprime.} \end{cases}$$

The function g is continuous at x iff x is zero or irrational. To see this, we first observe that if $x \in \mathbb{Q} \setminus \{0\}$, then $g(x) \neq 0$ but there are irrational numbers z as close as we like to x which satisfy $g(z) = 0$. Thus g is not continuous at the points of $\mathbb{Q} \setminus \{0\}$. On the other hand, if $x \in \mathbb{R} \setminus \mathbb{Q}$ or $x = 0$, we can establish continuity of g at x by an epsilon delta argument. We agree that whatever $\epsilon > 0$ we will always choose $\delta < 1$. Then the number of points z in the interval $]x - \delta, x + \delta[$ where $|g(z)| \ge \epsilon$ is finite because such a z is necessarily a rational number that can be expressed in the form $\frac{p}{q}$ where $1 \le q < \epsilon^{-1}$. With only finitely many points to avoid, it is now easy to find $\delta > 0$ such that

$$|z - x| < \delta \qquad \Longrightarrow \qquad |g(z) - g(x)| = |g(z)| < \epsilon.$$

□

There are various other ways of formulating continuity at a point.

Theorem 2.7 *Let X and Y be metric spaces. Let $f : X \to Y$. Let $x \in X$. Then the following statements are equivalent.*

- *f is continuous at x.*
- *For every neighbourhood V of $f(x)$ in Y, $f^{-1}(V)$ is a neighbourhood of x in X.*
- *For every sequence (x_n) in X converging to x, the sequence $(f(x_n))$ converges to $f(x)$ in Y.*

Proof. We show that the first statement implies the second. Let f be continuous at x and suppose that V is a neighbourhood of $f(x)$ in Y. Then there exists $\epsilon > 0$ such that $U(f(x), \epsilon) \subseteq V$ in Y. By definition of continuity at a point, there exists $\delta > 0$ such that

$$\begin{aligned} z \in U(x, \delta) \quad &\Rightarrow \quad f(z) \in U(f(x), \epsilon) \\ &\Rightarrow \quad f(z) \in V \\ &\Rightarrow \quad z \in f^{-1}(V). \end{aligned}$$

Hence $f^{-1}(V)$ is a neighbourhood of x in X.

Next, we assume the second statement and establish the third. Let (x_n) be a sequence in X converging to x. Let $\epsilon > 0$. Then $U(f(x), \epsilon)$ is a neighbourhood of $f(x)$ in Y. By hypothesis, $f^{-1}(U(f(x), \epsilon))$ is a neighbourhood of x in X. By the first part of Proposition 2.4 there exists $N \in \mathbb{N}$ such that

$$n > N \quad \Rightarrow \quad x_n \in f^{-1}(U(f(x), \epsilon)).$$

But this is equivalent to

$$n > N \quad \Rightarrow \quad f(x_n) \in U(f(x), \epsilon).$$

Thus $(f(x_n))$ converges to $f(x)$ in Y.

Finally we show that the third statement implies the first. We argue by contradiction. Suppose that f is not continuous at x. Then there exists $\epsilon > 0$ such that for all $\delta > 0$, there exists $z \in X$ with $d(x, z) < \delta$, but $d(f(x), f(z)) \geq \epsilon$. We take choice $\delta = \frac{1}{n}$ for $n = 1, 2, \ldots$ in sequence. We find that there exist x_n in X with $d(x, x_n) < \frac{1}{n}$, but $d(f(x), f(x_n)) \geq \epsilon$. But now, the sequence (x_n) converges to x in X while the sequence $(f(x_n))$ does not converge to $f(x)$ in Y. ■

We next build the global version of continuity from the concept of continuity at a point.

DEFINITION *Let X and Y be metric spaces and let $f : X \longrightarrow Y$. Then the mapping f is* ***continuous*** *iff f is continuous at every point x of X.*

There are also many possible reformulations of global continuity.

THEOREM 2.8 *Let X and Y be metric spaces. Let $f : X \longrightarrow Y$. Then the following statements are equivalent to the continuity of f.*

- *For every open set U in Y, $f^{-1}(U)$ is open in X.*
- *For every closed set A in Y, $f^{-1}(A)$ is closed in X.*
- *For every convergent sequence (x_n) in X with limit x, the sequence $(f(x_n))$ converges to $f(x)$ in Y.*

Proof. Let f be continuous. We check that the first statement holds. Let $x \in f^{-1}(U)$. Then $f(x) \in U$. Since U is open in Y, U is a neighbourhood of $f(x)$. Hence, by Theorem 2.7 $f^{-1}(U)$ is a neighbourhood of x. We have just shown that $f^{-1}(U)$ is a neighbourhood of each of its points. Hence $f^{-1}(U)$ is open in X. For the converse, we assume that the first statement holds. Let x be an arbitrary point of X. We must show that f is continuous at x. Again we plan to use Theorem 2.7. Let V be a neighbourhood of $f(x)$ in Y. Then, there exists $t > 0$ such that $U(f(x), t) \subseteq V$. It is shown on page 11 that $U(f(x), t)$ is an open subset of Y. Hence using the hypothesis, $f^{-1}(U(f(x),t))$ is open in X. Since $x \in f^{-1}(U(f(x),t))$, this set is a neighbourhood of x, and it follows that so is the larger subset $f^{-1}(V)$.

The second statement is clearly equivalent to the first. For instance if A is closed in Y, then $Y \setminus A$ is an open subset. Then

$$X \setminus f^{-1}(A) = f^{-1}(Y \setminus A)$$

is open in X and it follows that $f^{-1}(A)$ is closed in X. The converse entirely similar.

The third statement is equivalent directly from the definition. ■

One very useful condition that implies continuity is the Lipschitz condition.

DEFINITION *Let X and Y be metric spaces. Let $f : X \longrightarrow Y$. Then f is a* ***Lipschitz map*** *iff there is a constant C with $0 < C < \infty$ such that*

$$d_Y(f(x_1), f(x_2)) \leq C d_X(x_1, x_2) \qquad \forall x_1, x_2 \in X.$$

In the special case that $C = 1$ we say that f is a ***nonexpansive mapping****. In the even more restricted case that*

$$d_Y(f(x_1), f(x_2)) = d_X(x_1, x_2) \qquad \forall x_1, x_2 \in X,$$

we say that f is an ***isometry****.*

PROPOSITION 2.9 *Every Lipschitz map is continuous.*

Proof. We work directly. Let $\epsilon > 0$. The set $\delta = C^{-1}\epsilon$. Then $d_X(z,x) < \delta$ implies that

$$d_Y(f(z), f(x)) \leq C d_X(z,x) \leq C\delta = \epsilon.$$

as required. ■

2.4 Compositions of Functions

DEFINITION *Let X, Y and Z be sets. Let $f : X \longrightarrow Y$ and $g : Y \longrightarrow Z$ be mappings. Then we can make a new mapping $h : X \longrightarrow Z$ by $h(x) = g(f(x))$. In other words, to map by h we first map by f from X to Y and then by g from Y to Z. The mapping h is called the* ***composition*** *or* ***composed mapping*** *of f and g. It is usually denoted by $h = g \circ f$.*

Composition occurs in very many situations in mathematics. It is the primary tool for building new mappings out of old.

THEOREM 2.10 *Let X, Y and Z be metric spaces. Let $f : X \longrightarrow Y$ and $g : Y \longrightarrow Z$ be continuous mappings. Then the composition $g \circ f$ is a continuous mapping from X to Z.*

THEOREM 2.11 *Let X, Y and Z be metric spaces. Let $f : X \longrightarrow Y$ and $g : Y \longrightarrow Z$ be mappings. Suppose that $x \in X$, that f is continuous at x and that g is continuous at $f(x)$. Then the composition $g \circ f$ is a continuous at x.*

Proof of Theorems 2.10 and 2.11. There are many possible ways of proving these results using the tools from Theorem 2.8 and 2.7. It is even relatively easy to work directly from the definition.

Let us use sequences. In the local case, we take x as a fixed point of X whereas in the global case we take x to be a generic point of X.

Let (x_n) be a sequence in X convergent to x. Then since f is continuous at x, $(f(x_n))$ converges to $f(x)$. But, then using the fact that g is continuous at $f(x)$, we find that $(g(f(x_n)))$ converges to $g(f(x))$. This says that $(g \circ f(x_n))$ converges to $g \circ f(x)$. Since this holds for every sequence (x_n) convergent to x, it follows that $g \circ f$ is continuous (respectively continuous at x). ■

2.5 Product Spaces and Mappings

In order to discuss combinations of functions we need some additional machinery.

DEFINITION *Let (X, d_X) and (Y, d_Y) be metric spaces. Then we define a* ***product metric*** *d on the product set $X \times Y$ which allows us to consider $X \times Y$ as a* ***product metric space****. We do this as follows*

$$d((x_1, y_1), (x_2, y_2)) = \max(d_X(x_1, x_2), d_Y(y_1, y_2)) \tag{2.7}$$

PROPOSITION 2.12 *Equation (2.7) defines a bona fide metric on $X \times Y$.*

Proof. The first three conditions in the definition of a metric (on page 8) are obvious. It remains to check the triangle inequality. Let $(x_1, y_1), (x_2, y_2)$ and (x_3, y_3) be three generic points of $X \times Y$. Then $d((x_1, y_1), (x_3, y_3))$ is the maximum of $d_X(x_1, x_3)$ and $d_Y(y_1, y_3)$. Let us suppose without loss of generality that $d_X(x_1, x_3)$ is the larger of the two quantities. Then, by the triangle inequality on X, we have

$$d_X(x_1, x_3) \le d_X(x_1, x_2) + d_X(x_2, x_3). \tag{2.8}$$

But the right hand side of (2.8) is in turn less than $d((x_1, y_1), (x_2, y_2)) + d((x_2, y_2), (x_3, y_3))$ providing the required result. ■

With the definition out of the way, the next step is to see how it relates to other topological constructs.

LEMMA 2.13 *Let X and Y be metric spaces. Let $x \in X$ and let (x_n) be a sequence in X. Let $y \in Y$ and let (y_n) be a sequence in Y. Then the sequence $((x_n, y_n))$ converges to (x, y) in $X \times Y$ if and only if the sequence (x_n) converges to x in X and the sequence (y_n) converges to y in Y.*

Proof. First, suppose that $((x_n, y_n))$ converges to (x, y) in $X \times Y$. We must show that (x_n) converges to x in X. (It will follow similarly that (y_n) converges to y in Y.) This amounts then to showing that the projection $\pi : X \times Y \longrightarrow X$ onto the first coordinate, given by

$$\pi((x, y)) = x$$

is continuous. But the definition of the product metric ensures that π is nonexpansive (see page 17) and hence is continuous. The key inequality is

$$d_X(x_1, x_2) \le d_{X \times Y}((x_1, y_1), (x_2, y_2)).$$

For the converse, we have to get our hands dirtier. Let $\epsilon > 0$. Then there exists N such that $d_X(x_n, x) < \epsilon$ for $n > N$. Also, there exists M such that $d_Y(y_n, y) < \epsilon$ for $n > M$. It follows that for $n > \max(N, M)$ both of the above inequalities hold, so that

$$\max(d_X(x_n, x), d_Y(y_n, y)) < \epsilon.$$

But this is exactly equivalent to

$$d_{X \times Y}((x_n, y_n), (x, y)) < \epsilon$$

as required for the convergence of $((x_n, y_n))$ to (x, y). ■

There is a simple way to understand neighbourhoods and hence open sets in product spaces.

PROPOSITION 2.14 *Let X and Y be metric spaces. Let $x \in X$ and $y \in Y$. Let $U \subseteq X \times Y$. Then the following two statements are equivalent*

- *U is a neighbourhood of (x, y).*
- *There exist V a neighbourhood of x and W a neighbourhood of y such that $V \times W \subseteq U$.*

Proof. Suppose that the first statement holds. Then there exists $t > 0$ such that $U_{X\times Y}((x, y), t) \subseteq U$. But it is easy to check that

$$U_{X\times Y}((x, y), t) = U_X(x, t) \times U_Y(y, t).$$

Of course, $U_X(x, t)$ is a neighbourhood of x in X and $U_Y(y, t)$ is a neighbourhood of y in Y.

Conversely, let V and W be neighbourhoods of x and y in X and Y respectively. Then there exist $t, s > 0$ such that $U_X(x, t) \subseteq V$ and $U_Y(y, s) \subseteq W$. It is then easy to verify that

$$U_{X\times Y}((x, y), \min(t, s)) \subseteq U_X(x, t) \times U_Y(y, s) \subseteq V \times W \subseteq U,$$

so that U is a neighbourhood of (x, y) as required. ■

Next, we introduce ***product mappings***.

DEFINITION *Let X, Y, P and Q be sets. Let $f : X \longrightarrow P$ and $g : Y \longrightarrow Q$. Then we define the* ***product mapping*** *$f \times g : X \times Y \longrightarrow P \times Q$ by*

$$(f \times g)(x, y) = (f(x), g(y)).$$

PROPOSITION 2.15 *Let X, Y, P and Q be metric spaces. Let $f : X \longrightarrow P$ and $g : Y \longrightarrow Q$ be continuous mappings. Then the product mapping $f \times g$ is also continuous.*

Proof. We argue using sequences. We could equally well use neighbourhoods or epsilons and deltas. Let $((x_n, y_n))$ be an arbitrary sequence in $X \times Y$ converging to (x, y). Then (x_n) converges to x in X by Lemma 2.13. By Theorem 2.8 we find that $(f(x_n))$ converges to $f(x)$. Similar reasoning shows that $(g(y_n))$ converges to $g(y)$. Now we use Lemma 2.13 again to show that $((f(x_n), g(y_n)))$ converges to $(f(x), g(y))$. Finally since $((x_n, y_n))$ is an arbitrary sequence in $X \times Y$ converging to (x, y), it follows again by Theorem 2.13 that $f \times g$ is continuous. ■

There is also a local version of Proposition 2.15. We leave both the statement and the proof to the reader.

2.6 The Diagonal Mapping and Pointwise Combinations

DEFINITION *Let X be a metric space. Then the* ***diagonal mapping*** *on X is the mapping $\Delta_X : X \longrightarrow X \times X$ given by*

$$\Delta_X(x) = (x, x) \qquad \forall x \in X.$$

If X is a metric space it is easy to check that Δ_X is an isometry (for the definition, see page 17). In particular, Δ_X is a continuous mapping. This gives us the missing link to discuss the continuity of pointwise combinations.

THEOREM 2.16 *Let P, Q and R be metric spaces. Let $\mu : P \times Q \longrightarrow R$ be a continuous mapping. Let $f : X \longrightarrow P$ and $g : X \longrightarrow Q$ also be continuous mappings. Then the combination $h : X \longrightarrow R$ given by*

$$h(x) = \mu(f(x), g(x)) \qquad \forall x \in X$$

is also continuous.

Proof. It suffices to write $h = \mu \circ (f \times g) \circ \Delta_X$ and to apply Theorem 2.10 and Proposition 2.15 together with the continuity of Δ_X. ■

There are numerous examples of Theorem 2.16. In effect, the examples that follow are examples of continuous binary operations.

EXAMPLE Let $P = Q = R = \mathbb{R}$. Let $\mu(x, y) = x + y$, addition on $\mathbb{R}$. Then if $f, g : X \longrightarrow \mathbb{R}$ are continuous so is the sum function $f + g$ defined by

$$(f + g)(x) = f(x) + g(x) \qquad \forall x \in X.$$

It remains to check the continuity of μ. We have

$$\begin{aligned} |\mu(x_1, y_1) - \mu(x_2, y_2)| &= |(x_1 - x_2) + (y_1 - y_2)| \\ &\leq |x_1 - x_2| + |y_1 - y_2| \\ &\leq d_{\mathbb{R}\times\mathbb{R}}((x_1, y_1), (x_2, y_2)) + d_{\mathbb{R}\times\mathbb{R}}((x_1, y_1), (x_2, y_2)) \\ &= 2d_{\mathbb{R}\times\mathbb{R}}((x_1, y_1), (x_2, y_2)), \end{aligned}$$

so that μ is Lipschitz with constant $C = 2$ and hence continuous. □

EXAMPLE Let $P = Q = R = \mathbb{R}$. Let $\mu(x, y) = xy$, multiplication on $\mathbb{R}$. Then if $f, g : X \longrightarrow \mathbb{R}$ are continuous so is the pointwise product function fg defined by

$$(fg)(x) = f(x)g(x) \qquad \forall x \in X.$$

We check that μ is continuous at (x_1, y_1). Observe that

$$xy - x_1y_1 = x_1(y - y_1) + (x - x_1)y_1 + (x - x_1)(y - y_1)$$

so that

$$|xy - x_1y_1| \leq |x_1||y - y_1| + |x - x_1||y_1| + |x - x_1||y - y_1|$$

Now let $\epsilon > 0$ be given. We choose $\delta = \min(1, (|x_1| + |y_1| + 1)^{-1}\epsilon)$. Then

$$d_{\mathbb{R}\times\mathbb{R}}((x, y), (x_1, y_1)) < \delta$$

implies that

$$\begin{aligned} |xy - x_1y_1| &< |x_1|\delta + \delta|y_1| + \delta^2 \\ &\leq (|x_1| + |y_1| + 1)\delta \\ &\leq \epsilon. \end{aligned}$$

This estimate establishes that μ is continuous at (x_1, y_1). □

We leave the reader to check that addition and multiplication are continuous operations in $\mathbb{C}$. Two other operations on $\mathbb{R}$ that are continuous are max and min. We leave the reader to show that these are distance decreasing.

Example One very important binary operation on a metric space is the distance function itself. Let X be a metric space, $P = Q = X$ and $R = \mathbb{R}^+$. Let $\mu(x, y) = d(x, y)$. We check that μ is continuous. By the extended triangle inequality (page 8) we have

$$\begin{aligned} d(x_2, y_2) &\leq d(x_2, x_1) + d(x_1, y_1) + d(y_1, y_2) \\ &\leq d(x_1, y_1) + 2d_{X \times X}((x_1, y_1), (x_2, y_2)), \end{aligned}$$

and similarly

$$d(x_1, y_1) \leq d(x_2, y_2) + 2d_{X \times X}((x_1, y_1), (x_2, y_2)).$$

We may combine these two inequalities into one as

$$|d(x_1, y_1) - d(x_2, y_2)| \leq 2d_{X \times X}((x_1, y_1), (x_2, y_2)).$$

This shows that the distance function is Lipschitz with constant $C = 2$, and hence is continuous. □

Other examples of continuous binary operations are found in the context of normed spaces. Let us recall that in a normed space $(V, \| \ \|)$, the metric d_V is given by

$$d_V(v_1, v_2) = \|v_1 - v_2\| \qquad \forall v_1, v_2 \in V.$$

We will treat only the case of real normed spaces. The complex case is similar.

Example In a normed space $(V, \| \ \|)$, the vector addition operator is continuous. Let $\mu(v, w) = v + w$. We have

$$\begin{aligned} \|\mu(v_1, w_1) - \mu(v_2, w_2)\| &= \|(v_1 - v_2) + (w_1 - w_2)\| \\ &\leq \|v_1 - v_2\| + \|w_1 - w_2\| \\ &\leq d_{V \times V}((v_1, w_1), (v_2, w_2)) + d_{V \times V}((v_1, w_1), (v_2, w_2)) \\ &= 2d_{V \times V}((v_1, w_1), (v_2, w_2)), \end{aligned}$$

so that μ is Lipschitz with constant $C = 2$. □

While the previous example parallelled addition in $\mathbb{R}$, the next is similar to multiplication in $\mathbb{R}$.

Example In a normed space $(V, \| \ \|)$, the scalar multiplication operator is continuous. Thus $P = \mathbb{R}$, $Q = R = V$, and $\mu : \mathbb{R} \times V \longrightarrow V$ is the map $\mu(t, v) = tv$. We leave the details to the reader. □

Example Now let V be a real inner product space. Then the inner product is continuous. Thus $P = Q = V$, $R = \mathbb{R}$ and $\mu(v, w) = \langle v, w \rangle$.

We check that μ is continuous at (v_1, w_1). Observe that

$$\langle v, w \rangle - \langle v_1, w_1 \rangle = \langle v_1, w - w_1 \rangle + \langle v - v_1, w_1 \rangle + \langle v - v_1, w - w_1 \rangle$$

so that by the Cauchy-Schwarz inequality (page 4) we have

$$|\langle v, w \rangle - \langle v_1, w_1 \rangle| \leq \|v_1\| \|w - w_1\| + \|v - v_1\| \|w_1\| + \|v - v_1\| \|w - w_1\|.$$

Now let $\epsilon > 0$ be given. We choose $\delta = \min(1, (\|v_1\| + \|w_1\| + 1)^{-1}\epsilon)$. Then

$$d_{V \times V}((v, w), (v_1, w_1)) < \delta$$

implies that

$$\begin{aligned} |\langle v, w \rangle - \langle v_1, w_1 \rangle| &< \|v_1\|\delta + \delta\|w_1\| + \delta^2 \\ &\leq (\|v_1\| + \|w_1\| + 1)\delta \\ &\leq \epsilon. \end{aligned}$$

This estimate establishes that μ is continuous at (v_1, w_1). □

2.7 Interior and Closure

We return to discuss subsets and sequences in metric spaces in greater detail. Let X be a metric space and let A be an arbitrary subset of X. Then $\emptyset$ is an open subset of X contained in A, so we can define the ***interior*** $\operatorname{int}(A)$ of A by

$$\operatorname{int}(A) = \bigcup_{U \text{ open } \subseteq A} U. \tag{2.9}$$

By Theorem 2.2 (page 11), we see that $\operatorname{int}(A)$ is itself an open subset of X contained in A. Thus $\operatorname{int}(A)$ is the unique open subset of X contained in A which in turn contains all open subsets of X contained in A. There is a simple characterization of $\operatorname{int}(A)$ in terms of interior points (page 10).

PROPOSITION 2.17 *Let X be a metric space and let $A \subseteq X$. Then*

$$\operatorname{int}(A) = \{x; x \text{ is an interior point of } A\}.$$

Proof. Let $x \in \operatorname{int}(A)$. Then since $\operatorname{int}(A)$ is open, it is a neighbourhood of x. But then the (possibly) larger set A is also a neighbourhood of x. This just says that x is an interior point of A.

For the converse, let x be an interior point of A. Then by definition, there exists $t > 0$ such that $U(x,t) \subseteq A$. But it is shown on page 11, that $U(x,t)$ is open. Thus $U = U(x,t)$ figures in the union in (2.9), and since $x \in U(x,t)$ it follows that $x \in \operatorname{int}(A)$. ■

EXAMPLE The interior of the closed interval $[a,b]$ of $\mathbb{R}$ is just $]a,b[$. □

EXAMPLE The Cantor set E has empty interior in $\mathbb{R}$. Suppose not. Let x be an interior point of E. Then there exist $\epsilon > 0$ such that $U(x,\epsilon) \subseteq E$. Choose now n so large that $3^{-n} < \epsilon$. Then we also have $U(x,\epsilon) \subseteq E_n$. For the notation see page 13. This says that E_n contains an open interval of length $2(3^{-n})$ which is clearly not the case. □

By passing to the complement and using Theorem 2.5 (page 13) we see that there is a unique closed subset of X containing A which is contained in every closed subset of X which contains A. The formal definition is

$$\operatorname{cl}(A) = \bigcap_{E \text{ closed } \supseteq A} E. \tag{2.10}$$

The set $\operatorname{cl}(A)$ is called the ***closure*** of A. We would like to have a simple characterization of the closure.

PROPOSITION 2.18 *Let X be a metric space and let $A \subseteq X$. Let $x \in X$. Then $x \in \operatorname{cl}(A)$ is equivalent to the existence of a sequence of points (x_n) in A converging to x.*

Proof. Let $x \in \operatorname{cl}(A)$. Then x is not in $\operatorname{int}(X \setminus A)$. Then by Proposition 2.17, x is not an interior point of $X \setminus A$. Then, for each $n \in \mathbb{N}$, there must be a point $x_n \in A \cap U(x, \frac{1}{n})$. But now, $x_n \in A$ and (x_n) converges to x.

For the converse, let (x_n) be a sequence of points of A converging to x. Then $x_n \in \operatorname{cl}(A)$ and since $\operatorname{cl}(A)$ is closed, it follows from the definition of a closed set that $x \in \operatorname{cl}(A)$. ■

While Proposition 2.18 is perfectly satisfactory for many purposes, there is a subtle variant that is sometimes necessary.

DEFINITION *Let X be a metric space and let $A \subseteq X$. Let $x \in X$. Then x is an* ***accumulation point*** *or a* ***limit point*** *of A iff $x \in \text{cl}(A \setminus \{x\})$.*

PROPOSITION 2.19 *Let X be a metric space and let $A \subseteq X$. Let $x \in X$. Then the following statements are equivalent.*

- *$x \in \text{cl}(A)$.*
- *$x \in A$ or x is an accumulation point of A.*

Proof. That the second statement implies the first follows easily from Proposition 2.18. We establish the converse. Let $x \in \text{cl}(A)$. We may suppose that $x \notin A$, for else we are done. Now apply the argument of Proposition 2.18 again. For each $n \in \mathbb{N}$, there is a point $x_n \in A \cap U(x, \frac{1}{n})$. Since $x \notin A$, we have $A = A \setminus \{x\}$. Thus we have found $x_n \in A \setminus \{x\}$ with (x_n) converging to x. ■

DEFINITION *Let X be a metric space and let $A \subseteq X$. Let $x \in A$. Then x is an* ***isolated point*** *of A iff there exists $t > 0$ such that $A \cap U(x,t) = \{x\}$.*

We leave the reader to check that a point of A is an isolated point of A if and only if it is not an accumulation point of A.

A very important concept related to closure is the concept of density.

DEFINITION *Let X be a metric space and let $A \subseteq X$. Then A is said to be* ***dense*** *in X if $\text{cl}(A) = X$.*

If A is dense in X, then by definition, for every $x \in X$ there exists a sequence (x_n) in A converging to x.

PROPOSITION 2.20 *Let f and g be continuous mappings from X to Y. Suppose that A is a dense subset of X and that $f(x) = g(x)$ for all $x \in A$. Then $f(x) = g(x)$ for all $x \in X$.*

Proof. Let $x \in X$ and let (x_n) be a sequence in A converging to x. Then $f(x_n) = g(x_n)$ for all $n \in \mathbb{N}$. So the sequences $(f(x_n))$ and $(g(x_n))$ which converge to $f(x)$ and $g(x)$ respectively, are in fact identical. By the uniqueness of the limit, Proposition 2.3 (page 12), it follows that $f(x) = g(x)$. This holds for all $x \in X$ so that $f = g$. ■

We leave the proof of the following Proposition to the reader.

PROPOSITION 2.21 *Let A be a dense subset of a metric space X and let B be a dense subset of a metric space Y. Then $A \times B$ is dense in $X \times Y$.*

2.8 Limits in Metric Spaces

DEFINITION *Let X be a metric space and let $t > 0$. Then for $x \in X$ the* ***deleted open ball*** *$U'(x,t)$ is defined by*

$$U'(x,t) = \{z; z \in X, 0 < d(x,z) < t\} = U(x,t) \setminus \{x\}.$$

Let A be a subset of X then it is routine to check that x is an accumulation point of A if and only if for all $t > 0$, $U'(x,t) \cap A \neq \emptyset$. Deleted open balls are also used to define the concept of a ***limit***.

Definition *Let X and Y be metric spaces. Let x be an accumulation point of X. Let $f : X \setminus \{x\} \Leftrightarrow\!\rightarrow Y$. Then $f(z)$ has* **limit** y **as** z **tends to** x **in** X*, in symbols*

$$\lim_{z \to x} f(z) = y \tag{2.11}$$

if and only if for all $\epsilon > 0$ there exists $\delta > 0$ such that

$$z \in U'(x, \delta) \Longrightarrow f(z) \in U(y, \epsilon).$$

In the same way one also defines $f(z)$ has **a limit as** z **tends to** x **in** X*, which simply means that (2.11) holds for some $y \in Y$.*

Note that in the above definition, the quantity $f(x)$ is undefined. The purpose of taking the limit is to "attach a value" to $f(x)$. The following Lemma connects this idea with the concept of continuity at a point. We leave the proof to the reader.

Lemma 2.22 *Let X and Y be metric spaces. Let x be an accumulation point of X. Let $f : X \setminus \{x\} \Leftrightarrow\!\rightarrow Y$. Suppose that (2.11) holds for some $y \in Y$. Now define $\tilde{f} : X \Leftrightarrow\!\rightarrow Y$ by*

$$\tilde{f}(z) = \begin{cases} f(z) & \text{if } z \in X \setminus \{x\}, \\ y & \text{if } z = x. \end{cases}$$

Then $\tilde{f}$ is continuous at x.

One of the most standard uses of limits is in the definition of the derivative.

Definition *Let $g :]a, b[\Leftrightarrow\!\rightarrow V$ where V may as well be a general normed vector space. Let $t \in]a, b[$. Then the quotient*

$$f(s) = (s \Leftrightarrow t)^{-1}(g(s) \Leftrightarrow g(t)) \in V$$

is defined for s in $]a, b[\setminus \{t\}$. It is not defined at $s = t$. If

$$\lim_{s \to t} f(s)$$

exists, then we say that g is **differentiable** *at t and the value of the limit is denoted $g'(t)$ and called the* **derivative** *of g at t. It is an element of V.*

2.9 Distance to a Subset

Definition *Let X be a metric space and let A be a non-empty subset of X. Then we may define for every element $x \in X$, the real number $\mathrm{dist}_A(x) \geq 0$ by*

$$\mathrm{dist}_A(x) = \inf_{a \in A} d(x, a).$$

This is the distance from x to the subset A. We view dist_A as a mapping $\mathrm{dist}_A : X \Leftrightarrow\!\rightarrow \mathbb{R}^+$.

Proposition 2.23 *Let X be a metric space and let $A \subseteq X$. Then*

- $\mathrm{dist}_A : X \Leftrightarrow\!\rightarrow \mathbb{R}^+$ *is continuous.*
- $\mathrm{dist}_A(x) = 0 \Leftrightarrow x \in \mathrm{cl}(A)$.
- $\mathrm{dist}_A(x) = \mathrm{dist}_{\mathrm{cl}(A)}(x) \qquad \forall x \in X$.

Proof. Let $x_1, x_2 \in X$ and $a \in A$. Then by the triangle inequality

$$d(x_1, a) \le d(x_1, x_2) + d(x_2, a).$$

Take infimums of both sides as a runs over the elements of A to obtain

$$\mathrm{dist}_A(x_1) \le d(x_1, x_2) + \mathrm{dist}_A(x_2). \tag{2.12}$$

An exactly similar argument yields

$$\mathrm{dist}_A(x_2) \le d(x_1, x_2) + \mathrm{dist}_A(x_1). \tag{2.13}$$

Now we combine (2.12) and (2.13) to find that

$$|\mathrm{dist}_A(x_1) \Leftrightarrow \mathrm{dist}_A(x_2)| \le d(x_1, x_2), \tag{2.14}$$

which asserts that dist_A is nonexpansive. The first assertion follows.

The second assertion follows directly from the definition of $\mathrm{cl}(A)$.

For the third assertion, it is clear that $\mathrm{dist}_{\mathrm{cl}(A)}(x) \le \mathrm{dist}_A(x)$ since $\mathrm{cl}(A)$ is a (possibly) larger set than A. It therefore remains to show that $\mathrm{dist}_{\mathrm{cl}(A)}(x) \ge \mathrm{dist}_A(x)$. By the definition of $\mathrm{dist}_{\mathrm{cl}(A)}(x)$, it suffices to take a an arbitrary point of $\mathrm{cl}(A)$ and show that

$$\mathrm{dist}_A(x) \le d(a, x). \tag{2.15}$$

Using the fact that $a \in \mathrm{cl}(A)$, we see that there is a sequence (a_n) of points of A converging to a. By definition of $\mathrm{dist}_A(x)$ we have

$$\mathrm{dist}_A(x) \le d(a_n, x) \tag{2.16}$$

But since d is a continuous function on $X \times X$, it now follows that $d(a_n, x) \Leftrightarrow\rightarrow d(a, x)$ as $n \Leftrightarrow\rightarrow \infty$. Combining this with (2.16) yields (2.15) as required. ■

2.10 Separability

In this text, we use the term ***countable*** to mean finite or countably infinite. Thus a set A is countable iff it can be put in one to one correspondence with some subset of $\mathbb{N}$.

DEFINITION *A metric space X is said to be **separable** iff it has a countable dense subset.*

EXAMPLE The real line $\mathbb{R}$ is a separable metric space with the standard metric because the set $\mathbb{Q}$ of rational numbers is dense in $\mathbb{R}$. □

The nomenclature is somewhat misleading. Separability has nothing to do with separation. In fact separability is a measure of the smallness of a metric space. Unfortunately this fact is not obvious. The following Theorem clarifies the situation.

THEOREM 2.24 *Let X be a separable metric space. Let Y be a subset of X. Then Y is separable when considered as a metric space with the restriction metric.*

Proof. Let A be a countable dense subset of X. Then it is certainly possible that $A \cap Y = \emptyset$. We need therefore to build a subset of Y in a more complicated way. Let (a_n) be an enumeration of A. By the definition of $\mathrm{dist}_Y(a_n)$ we can deduce the existence of an element $b_{n,k}$ of Y such that

$$d(a_n, b_{n,k}) < \mathrm{dist}_Y(a_n) + \tfrac{1}{k}. \tag{2.17}$$

We will show that the set $\{b_{n,k}; n, k \in \mathbb{N}\}$ is dense in Y. Towards this, let $y \in Y$. We will show that for every $\epsilon > 0$ there exist n and k such that $b_{n,k} \in U(y, \epsilon)$. We choose n such that $d(a_n, y) < \frac{1}{3}\epsilon$, possible because A is dense in X. We choose k so large that $\frac{1}{k} < \frac{1}{3}\epsilon$. It follows that

$$\begin{aligned} d(b_{n,k}, y) &\leq d(b_{n,k}, a_n) + d(a_n, y) \\ &\leq \operatorname{dist}_Y(a_n) + \tfrac{1}{k} + d(a_n, y) \\ &\leq d(a_n, y) + \tfrac{1}{k} + d(a_n, y) \\ &< \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon = \epsilon. \end{aligned}$$

as required. ■

Much easier is the following Theorem the proof of which we leave as an exercise.

THEOREM 2.25 *Let X and Y be separable metric spaces. Then $X \times Y$ is again a separable metric space (with the product metric).*

The following result is needed in applications to measure theory.

THEOREM 2.26 *In a separable metric space X, every open subset U can be written as a countable union of open balls.*

Proof. We leave the reader to prove the Theorem in case that $U = X$ and assume henceforth that $U \neq X$. By Theorem 2.24, the set U itself possesses a countable dense subset. Let us enumerate this subset as (x_n). We claim that

$$U = \bigcup_n U(x_n, \tfrac{1}{2}\operatorname{dist}_{X \setminus U}(x_n)). \tag{2.18}$$

Obviously, the right hand side of (2.18) is contained in the left hand side. To establish the claim, we let $x \in U$ and show that x is in the right hand member of (2.18). Let $t = \operatorname{dist}_{X \setminus U}(x) > 0$ because of Proposition 2.23 and since $X \setminus U$ is a closed set. Using the density of (x_n), we may find $n \in \mathbb{N}$ such that

$$d(x, x_n) < \tfrac{1}{3}t. \tag{2.19}$$

By (2.14), we have that

$$|\operatorname{dist}_{X \setminus U}(x) \Leftrightarrow \operatorname{dist}_{X \setminus U}(x_n)| \leq d(x, x_n), \tag{2.20}$$

and it follows from (2.19), (2.20) and the definition of t that

$$\operatorname{dist}_{X \setminus U}(x_n) > \tfrac{2}{3}t.$$

Then we have

$$d(x, x_n) < \tfrac{1}{3}t = \tfrac{1}{2}(\tfrac{2}{3}t) < \tfrac{1}{2}\operatorname{dist}_{X \setminus U}(x_n).$$

It follows that $x \in U(x_n, \frac{1}{2}\operatorname{dist}_{X \setminus U}(x_n))$ as required. ■

EXAMPLE Every subset of $\mathbb{R}^d$ is separable. □

EXAMPLE The normed vector space ℓ^∞ (page 8) is not separable. To see this, suppose that S is a dense subset of ℓ^∞. Let $\omega = (\omega_n)$ be a sequence taking the values ± 1. There are uncountably many such sequences ω. For each such sequence, there is a sequence $s = (s_n)$ in S such that $\|s \Leftrightarrow \omega\| < \frac{1}{3}$. It is easy to see that two distinct values of ω necessarily lead to distinct elements of S. It follows that S is also uncountable. □

Example On the other hand, the space ℓ^1 (page 8) is separable. Let S be the set of sequences with rational entries eventually zero. Then S is a countable set. Given a sequence $x = (x_n)$ in ℓ^1 and a strictly positive real number ϵ, we first choose N so large that

$$\sum_{n>N} |x_n| < \frac{\epsilon}{2}.$$

Let $y = (y_n)$ be the truncated sequence given by

$$y_n = \begin{cases} x_n & \text{if } n \le N, \\ 0 & \text{if } n > N. \end{cases}$$

Then $\|x \Leftrightarrow y\|_1 < \frac{1}{2}\epsilon$. It now remains to find a slightly perturbed sequence $s = (s_n) \in S$ such that $\|y \Leftrightarrow s\|_1 < \frac{1}{2}\epsilon$. We leave this as an exercise. For more on this example see Proposition 2.29 □

2.11 Relative Topologies

We remarked on page 8 that if X is a metric space and Y is a subset of X then Y can be considered as a metric space in its own right. From the point of view of convergent sequences, this causes no problems. The sequences in Y that converge in Y to an element of Y are simply the sequences in Y that converge in X to an element of Y. Of course, it is possible to have a sequence of elements of Y which converges in X to an element of $X \setminus Y$. Such a sequence will not converge in Y.

The situation with regard to open and closed sets is more complicated, and certainly more difficult to understand. A subset A of Y can be said to be open in Y or said to be open in X. These concepts are different in general. To distinguish the difference, we sometimes say that A is ***relatively open*** when it is an open subset of the subset Y. In general the adverb ***relatively*** is reserved for properties considered with respect to the subset (in this case Y) rather than the whole space (in this case X). Thus when we say that A is ***relatively closed***, we mean that it is closed in Y. If A is ***relatively dense***, then it is dense in Y.

Let us consider an example to illustrate the difference.

Example Let $X = \mathbb{R}$ with the usual metric and $Y = [0, 1]$ with the relative metric. Then the subset $A = [0, \frac{1}{2}[$ of Y is not open in X because $0 \in A$ and every neighbourhood of 0 in X contains small negative numbers that are not in A. However 0 *is* an interior point of A with respect to Y. This is because $U_Y(0, \epsilon) = [0, \epsilon[\subseteq A$ provided $0 < \epsilon < \frac{1}{2}$. Those small negative numbers are not in Y and do not cause a problem when we are considering openness in Y. The reader should ponder this point until he understands it, because it is fundamental to so much that follows. In fact the subset A is open relative to Y. □

Example Let $X = \mathbb{R}$ with the usual metric and $Y = [0, 1[$ with the relative metric. Then the subset $A = [\frac{1}{2}, 1[$ is not closed in X, but it *is* closed in Y. The skeptic will immediately consider the sequence $(x_n = \frac{n}{n+1})$ which lies in A and "converges to 1". This is certainly true in X, but it is not true that $x_n \Leftrightarrow\!\!\rightarrow 1$ in Y for the simple reason that $1 \notin Y$. □

What is required is a way of understanding the open subsets of Y in terms of those of X. The following result fills that role.

Theorem 2.27 *Let X be a metric space and let $Y \subseteq X$.*

- *A subset U of Y is open in Y iff there exists an open subset V of X such that $U = V \cap Y$.*
- *A subset F of Y is closed in Y iff there exists a closed subset E of X such that $F = E \cap Y$.*

Proof. We work on the first statement. Let U be a subset of Y open in Y. By definition, for every $y \in U$ there exists $t_y > 0$ such that $U_Y(y, t_y) \subseteq U$. Now define

$$V = \bigcup_{y \in U} U_X(y, t_y).$$

Then V is an open subset of X by Theorem 2.2 (page 11). We have

$$\begin{aligned} V \cap Y &= \bigcup_{y \in U} (U_X(y, t_y) \cap Y) \\ &= \bigcup_{y \in U} U_Y(y, t_y) \\ &= U, \end{aligned}$$

since, for every $y \in U$, we have $y \in U_X(y, t_y)$.

Conversely, if V is open in X and $y \in V \cap Y$, then there exists $t > 0$ such that $U_X(y, t) \subseteq V$. Then obviously

$$U_Y(y, t) = U_X(y, t) \cap Y \subseteq V \cap Y.$$

Thus $V \cap Y$ is a neighbourhood of each of its points in Y. In other words $V \cap Y$ is open in Y. This completes the proof of the first assertion.

The second assertion follows immediately from the first and Theorem 2.5 (page 13). ■

EXAMPLE Consider $Y = \mathbb{R}$ embedded as the real axis in $X = \mathbb{R}^2$. The interval $]\Leftrightarrow 1, 1[$ is a relatively open subset of the real axis Y. It is clearly not an open subset of $\mathbb{R}^2$. However, the disc

$$\{(x, y); x^2 + y^2 < 1\}$$

is open in the plane X and meets the real axis Y in precisely $]\Leftrightarrow 1, 1[$. □

COROLLARY 2.28 *We maintain the notations of the Theorem. Thus X is a metric space, Y is a subset of X which we are considering as a metric space in its own right. Further U and F are subsets of Y*

- *If U is open in X, then it is open in Y.*
- *If Y is open in X and U is open in Y, then U is open in X.*
- *If F is closed in X, then it is closed in Y.*
- *If Y is closed in X and F is closed in Y, then F is closed in X.*

We can use relative topologies to elucidate the proof of the fact that the sequence space ℓ^1 is separable on page 27. Here there are three spaces $X = \ell^1$, Y the set of all real sequences that are eventually zero, and S the set of all rational sequences that are eventually zero. We have $S \subset Y \subset X$. We show that Y is dense in X and that S is relatively dense in Y. The density of S in X then follows from the following general principle which might be called the ***transitivity of density***.

PROPOSITION 2.29 *Let X be a metric space and let $S \subseteq Y \subseteq X$. Suppose that Y is dense in X and that S is relatively dense in Y. Then S is dense in X.*

Proof. Let $\epsilon > 0$ and suppose that $x \in X$. Then, since Y is dense in X there exists $y \in Y$ such that $d(x, y) < \frac{1}{2}\epsilon$. Now, since S is dense in Y, there exists $s \in S$ such that $d(y, s) < \frac{1}{2}\epsilon$. The triangle inequality now yields $d(x, s) < \epsilon$ as required. ■

Theorem 2.30 (Glueing Theorem) *Let X and Y be metric spaces. Let X_1 and X_2 be subsets of X such that $X = X_1 \cup X_2$. Let $f_j : X_j \Leftrightarrow\!\!\rightarrow Y$ be continuous maps for $j = 1, 2$. Suppose that f_1 and f_2 agree on their overlap — explicitly*

$$f_1(x) = f_2(x) \qquad \forall x \in X_1 \cap X_2,$$

so that the ***glued mapping*** *$f : X \Leftrightarrow\!\!\rightarrow Y$ given by*

$$f(x) = \begin{cases} f_1(x) & \text{if } x \in X_1, \\ f_2(x) & \text{if } x \in X_2, \end{cases}$$

is well defined. Suppose that one or other of the two following conditions holds.

- *Both X_1 and X_2 are open in X.*
- *Both X_1 and X_2 are closed in X.*

Then f is continuous.

Proof. Suppose that both X_1 and X_2 are open in X. We work with sequences. Let $x \in X$ and suppose that (x_n) is a sequence in X converging to x. Without loss of generality we may suppose that $x \in X_1$. Then since X_1 is open in X, the sequence (x_n) is eventually in X_1. Explicitly, there exists $N \in \mathbb{N}$ such that $x_n \in X_1$ for $n > N$. Since this tail of the sequence converges to x in X_1 and since f_1 is continuous as a mapping from X_1 to Y, the image sequence of the tail converges to $f_1(x)$. But this just says that $(f(x_n))$ converges to $f(x)$.

Let us go back and fill in the details in glorious technicolour. We define a new sequence (the tail) by $z_k = x_{N+k}$. We claim that z_k converges to x. Towards this, let $\epsilon > 0$. Then since (x_n) converges to x, there exists $M \in \mathbb{N}$ such that $d(x, x_n) < \epsilon$ for $n > M$. Then, certainly $d(x, z_k) < \epsilon$ for $k > M$. This proves the claim. Since for all k, $z_k \in X_1$ and since f_1 is continuous on X_1 we now find that $(f(z_k))$ converges to $f(x)$ in X. Now we claim that $(f(x_n))$ converges to $f(x)$. Let $\epsilon > 0$. Then there exists $K \in \mathbb{N}$, such that $k > K$ implies $d(f(z_k), f(x)) < \epsilon$. Then, for $n > N + K$ we have $d(f(x_n), f(x)) = d(f(z_k), f(x)) < \epsilon$ where $k = n \Leftrightarrow N > K$ as needed.

In case that X_1 and X_2 are both closed in X we use a completely different strategy, namely the characterization of continuity by closed subsets in Theorem 2.8 (page 16). Let A be a closed subset of Y. We must show that $f^{-1}(A)$ is closed in X. We write $f^{-1}(A) = (f^{-1}(A) \cap X_1) \cup (f^{-1}(A) \cap X_2)$ possible since $X = X_1 \cup X_2$. It is enough to show that the two sets $f^{-1}(A) \cap X_1$ and $f^{-1}(A) \cap X_2$ are closed in X. Without loss of generality we need only handle the first of these. Now $f^{-1}(A) \cap X_1 = f_1^{-1}(A)$, so that, by the continuity of f_1 this set is closed in X_1. Therefore, according to the last assertion of Corollary 2.28, it is also closed in X since X_1 is itself closed in X. ■

Example The Glueing Theorem is used in homotopy theory. Let f and g be continuous maps from a metric space X to a metric space Y. Then we say that f and g are ***homotopic*** iff there exist a continuous map

$$F : [0, 1] \times X \Leftrightarrow\!\!\rightarrow Y$$

such that

$$F(0, x) = f(x) \qquad \forall x \in X$$

and

$$F(1, x) = g(x) \qquad \forall x \in X.$$

It turns out that being homotopic is an equivalence relation. We leave the reflexivity and symmetry conditions to be verified by the reader. We now sketch the transitivity.

Let g and h also be homotopic. Then there is (with slight change in notation) a continuous mapping

$$G : [1, 2] \times X \Leftrightarrow\!\!\rightarrow Y$$

such that

$$G(1, x) = g(x) \qquad \forall x \in X$$

and

$$G(2, x) = h(x) \qquad \forall x \in X.$$

Since the subsets $[0, 1] \times X$ and $[1, 2] \times X$ are closed in $[0, 2] \times X$, the mappings F and G can be glued together to make a continuous mapping

$$H : [0, 2] \times X \Leftrightarrow\!\!\rightarrow Y$$

such that

$$H(0, x) = f(x) \qquad \forall x \in X$$

and

$$H(2, x) = h(x) \qquad \forall x \in X.$$

It follows that f and h are homotopic. □

2.12 Uniform Continuity

For many purposes, continuity of mappings is not enough. The following strong form of continuity is often needed.

DEFINITION *Let X and Y be metric spaces and let $f : X \Leftrightarrow\!\!\rightarrow Y$. Then we say that f is* ***uniformly continuous*** *iff for all $\epsilon > 0$ there exists $\delta > 0$ such that*

$$x_1, x_2 \in X, d_X(x_1, x_2) < \delta \qquad \Rightarrow \qquad d_Y(f(x_1), f(x_2)) < \epsilon. \tag{2.21}$$

In the definition of continuity, the number δ is allowed to depend on the point x_1 as well as ϵ.

EXAMPLE The function $f(x) = x^2$ is continuous, but not uniformly continuous as a mapping $f : \mathbb{R} \Leftrightarrow\!\!\rightarrow \mathbb{R}$. Certainly the identity mapping $x \Leftrightarrow\!\!\rightarrow x$ is continuous because it is an isometry. So f, which is the pointwise product of the identity mapping with itself is also continuous. We now show that f is not uniformly continuous. Let us take $\epsilon = 1$. Then, we must show that for all $\delta > 0$ there exist points x_1 and x_2 with $|x_1 \Leftrightarrow x_2| < \delta$, but $|x_1^2 \Leftrightarrow x_2^2| \geq 1$. Let us take $x_2 = x \Leftrightarrow \frac{1}{4}\delta$ and $x_1 = x + \frac{1}{4}\delta$. Then

$$x_1^2 \Leftrightarrow x_2^2 = (x_1 \Leftrightarrow x_2)(x_1 + x_2) = x\delta.$$

It remains to choose $x = \delta^{-1}$ to complete the argument. □

EXAMPLE Any function satisfying a Lipschitz condition (page 17) is uniformly continuous. Let X and Y be metric spaces. Let $f : X \Leftrightarrow\!\!\rightarrow Y$ with constant C. Then

$$d_Y(f(x_1), f(x_2)) \leq C d_X(x_1, x_2) \qquad \forall x_1, x_2 \in X.$$

Given $\epsilon > 0$ it suffices to choose $\delta = C^{-1}\epsilon > 0$ in order for $d_X(x_1, x_2) < \delta$ to imply $d_Y(f(x_1), f(x_2)) < \epsilon$. □

It should be noted that one cannot determine (in general) if a mapping is uniformly continuous from a knowledge only of the open subsets of X and Y. Thus, uniform continuity is not a topological property. It depends upon other aspects of the metrics involved.

In order to clarify the concept of uniform continuity and for other purposes, one introduces the ***modulus of continuity*** ω_f of a function f. Suppose that $f : X \longrightarrow Y$. Then $\omega_f(t)$ is defined for $t \geq 0$ by

$$\omega_f(t) = \sup\{d_Y(f(x_1), f(x_2)); x_1, x_2 \in X, d_X(x_1, x_2) \leq t\}. \tag{2.22}$$

It is easy to see that the uniform continuity of f is equivalent to

$$\forall \epsilon > 0,\ \exists \delta > 0 \text{ such that } 0 < t < \delta \quad \Rightarrow \quad \omega_f(t) < \epsilon.$$

We observe that $\omega_f(0) = 0$ and regard $\omega_f : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$. Then the uniform continuity of f is also equivalent to the continuity of ω_f at 0.

2.13 Subsequences

Subsequences are used extensively in analysis. Some advanced metric space concepts such as compactness can be handled quite nicely using subsequences. We start by defining a subsequence of the sequence of natural numbers.

DEFINITION *A sequence (n_k) of natural numbers is called a **natural subsequence** if $n_k < n_{k+1}$ for all $k \in \mathbb{N}$.*

Since $n_1 \geq 1$, a straightforward induction argument yields that $n_k \geq k$ for all $k \in \mathbb{N}$.

DEFINITION *Let (x_n) be a sequence of elements of a set X. A **subsequence** of (x_n) is a sequence (y_k) of elements of X given by*

$$y_k = x_{n_k}$$

where (n_k) is a natural subsequence.

The key result about subsequences is very easy and is left as an exercise for the reader.

LEMMA 2.31 *Let (x_n) be a sequence in a metric space X converging to an element $x \in X$. Then any subsequence (x_{n_k}) also converges to x.*

One way of showing that a sequence fails to converge is to find two convergent subsequences with different limits. Indeed, this idea can also be turned around. One way of showing that two sequences converge to the same limit is to build a new sequence that possesses both of the given sequences as subsequences. It is then enough to establish the convergence of the new sequence. This idea will be used in our discussion of completeness.

3

A Metric Space Miscellany

In this chapter we introduce some topics from metric spaces that are slightly out of the mainstream and which can be tackled with the rather meagre knowledge of the subject that we have amassed up to this point. This chapter is primarily intended to enrich the material presented thus far.

3.1 The p-norms on $\mathbb{R}^n$

Let $1 \leq p < \infty$. We define

$$\|(x_1, \ldots, x_n)\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{\frac{1}{p}}.$$

Our aim is to show that $\| \ \|_p$ is a norm. It is easy to verify all the conditions defining a norm except the last one — the subadditivity condition.

In case that $p = \infty$ we use (1.1) to define $\| \ \|_\infty$. This fits into the scheme in that

$$\max_{k=1}^{n} |x_k| = \lim_{p \longrightarrow \infty} \|(x_1, \ldots, x_n)\|_p.$$

Let $1 \leq p \leq \infty$. Then we define $p' = \frac{p}{p-1}$ the ***conjugate index*** of p. In case $p = 1$ we take $p' = \infty$, and in case $p = \infty$ we take $p' = 1$. We have

$$\frac{1}{p} + \frac{1}{p'} = 1,$$

so that the relationship between index and conjugate index is symmetric.

Proposition 3.1 (Hölder's Inequality) *For $x, y \in \mathbb{C}^n$ we have*

$$|\sum_{j=1}^{n} x_j y_j| \leq \|x\|_p \, \|y\|_{p'} \tag{3.1}$$

If $p = 1$ or $p = \infty$, Hölder's Inequality is easy to verify. In the general case we use the following lemma.

Lemma 3.2 *Let $x \geq 0$ and $y \geq 0$. Let $1 < p < \infty$ and let p' be the conjugate index of p, so that $1 < p' < \infty$. Then*

$$xy \leq \frac{1}{p} x^p + \frac{1}{p'} y^{p'}. \tag{3.2}$$

Proof. First of all, if $x = 0$ or $y = 0$ the inequality is obvious. We therefore assume that $x > 0$ and $y > 0$.

Next, observe that if $t > 0$ and we replace x by $t^{\frac{1}{p}}x$ and y by $t^{\frac{1}{p'}}y$ in (3.2) then since $\frac{1}{p} + \frac{1}{p'} = 1$, (3.2) is multiplied by t and its content is unchanged. Choosing t appropriately (in fact with $t = y^{-p'}$), we can assume without loss of generality that $y = 1$. The problem is now reduced to one-variable calculus.

Let us define a function f on $]0, \infty[$ by

$$f(x) = \frac{1}{p}x^p \Leftrightarrow x + \frac{1}{p'}.$$

Taking the derivative of f we obtain

$$f'(x) = x^{p-1} \Leftrightarrow 1.$$

Since $p > 1$ this leads to

$$f'(x) \geq 0 \qquad \text{if } x \geq 1, \tag{3.3}$$
$$f'(x) \leq 0 \qquad \text{if } x \leq 1. \tag{3.4}$$

It follows from (3.3), (3.4) and the Mean-Value Theorem that

$$f(x) \geq f(1) \qquad \text{if } x \geq 1, \tag{3.5}$$
$$f(x) \geq f(1) \qquad \text{if } x \leq 1. \tag{3.6}$$

Since $f(1) = 0$, (3.5) and (3.6) lead to

$$f(x) \geq 0 \qquad \forall x > 0. \tag{3.7}$$

But (3.7) is equivalent to (3.2) in the case $y = 1$, completing the proof of Lemma 3.2. ■

Proof of Hölder's Inequality. We first suppose that $||x||_p = 1$ and $||y||_{p'} = 1$. Then, by multiple applications of Lemma 3.2 we have

$$\begin{aligned} |\sum_{j=1}^n x_j y_j| &\leq \sum_{j=1}^n |x_j||y_j| \\ &\leq \sum_{j=1}^n \frac{1}{p}|x_j|^p + \frac{1}{p'}|y_j|^{p'} \\ &= \frac{1}{p}||x||_p^p + \frac{1}{p'}||y||_{p'}^{p'} \\ &= \frac{1}{p} + \frac{1}{p'} = 1. \end{aligned} \tag{3.8}$$

For the general case, we first observe that if $||x||_p = 0$, then $x = 0$ and the result is straightforward. We may assume that $||x||_p > 0$ and similarly that $||y||_{p'} > 0$. Then, applying (3.8) with x replaced by $||x||_p^{-1}x$ and y replaced by $||y||_{p'}^{-1}y$, we obtain

$$|\sum_{j=1}^n \frac{x_j}{||x||_p}\frac{y_j}{||y||_{p'}}| \leq 1.$$

Finally, multiplying by $||x||_p||y||_{p'}$ yields Hölder's inequality. ■

Theorem 3.3 (Minkowski's Inequality) *Let $1 \le p \le \infty$ and $x, y \in \mathbb{R}^n$. Then*

$$\|x+y\|_p \le \|x\|_p + \|y\|_p \tag{3.9}$$

holds.

Proof. The result is easy if $p = 1$ or if $p = \infty$. We therefore suppose that $1 < p < \infty$. We have

$$\begin{aligned} \|x+y\|_p^p &= \sum_{j=1}^{n} |x_j + y_j|^p \\ &= \sum_{j=1}^{n} |x_j + y_j||x_j + y_j|^{p-1} \\ &\le \sum_{j=1}^{n} (|x_j| + |y_j|)|x_j + y_j|^{p-1} \\ &= \sum_{j=1}^{n} |x_j||x_j + y_j|^{p-1} + \sum_{j=1}^{n} |y_j||x_j + y_j|^{p-1} \end{aligned} \tag{3.10}$$

The key is to apply Hölder's inequality to each of the two sums in (3.10). We have

$$\begin{aligned} \sum_{j=1}^{n} |x_j||x_j + y_j|^{p-1} &\le \left(\sum_{j=1}^{n} |x_j|^p \right)^{\frac{1}{p}} \left(\sum_{j=1}^{n} |x_j + y_j|^{p'(p-1)} \right)^{\frac{1}{p'}} \\ &= \|x\|_p \|x+y\|_p^{p-1}. \end{aligned} \tag{3.11}$$

since $p'(p \Leftrightarrow 1) = p$ and $\frac{1}{p'} = (p \Leftrightarrow 1)\frac{1}{p}$. Similarly

$$\sum_{j=1}^{n} |y_j||x_j + y_j|^{p-1} \le \|y\|_p \|x+y\|_p^{p-1}. \tag{3.12}$$

Combining now (3.10), (3.11) and (3.12), we obtain

$$\|x+y\|_p^p \le (\|x\|_p + \|y\|_p)\|x+y\|_p^{p-1}. \tag{3.13}$$

Now if $\|x+y\|_p = 0$ we have the conclusion (3.9). If not, then it is legitimate to divide (3.13) by $\|x+y\|_p^{p-1}$ and again the conclusion follows. ■

The p-norms are used most frequently in the cases $p = 1$, $p = 2$ and $p = \infty$. The case $p = 2$ is special in that the 2-norm is the Euclidean norm which arises from an inner product. In particular the standard Cauchy-Schwarz inequality

$$\left|\sum_{k=1}^{n} x_k y_k\right| \le \left(\sum_{j=1}^{n} |x_j|^2 \right)^{\frac{1}{2}} \left(\sum_{k=1}^{n} |y_k|^2 \right)^{\frac{1}{2}}$$

is just the case $p = 2$ of Hölder's Inequality (3.1).

3.2 Minkowski's Inequality and convexity

The proof of Theorem 3.9 is really very slick. However, it is not easy to understand the motivating forces behind the proof. We have seen in Theorem 1.3 (which relates to the line condition) that the subadditivity of a norm is related to convexity. If there is justice, it should be possible to understand Minkowski's Inequality as a convexity inequality. This is the purpose of this section.

Definition *Let $a < b$ and suppose that $f :]a, b[\longrightarrow \mathbb{R}$. Then we say that f is* ***convex****, or more precisely a* ***convex function*** *iff it satisfies the inequality*

$$f((1 - t)x_1 + tx_2) \leq (1 - t)f(x_1) + tf(x_2)$$

for all $x_1, x_2 \in]a, b[$ and for all t satisfying $0 \leq t \leq 1$.

The rationale for this definition and one way in which it relates to convex sets is that f is a convex function iff the region

$$\{(x, y); a < x < b, y > f(x)\}$$

lying above the graph of f is a convex subset of the plane $\mathbb{R}^2$.

The connection with norms is also clear. If $\|\ \|$ is a norm on a real vector space V then the function

$$f(t) = \|v_1 + tv_2\|$$

is convex on $\mathbb{R}$ for every fixed v_1 and v_2 in V. This result has a converse.

Lemma 3.4 *Let $\|\ \|$ be a quantity defined on a real vector space V and such that the first three conditions of the definition of a norm hold. Suppose that*

$$f(t) = \|v_1 + tv_2\| \tag{3.14}$$

is convex on $\mathbb{R}$ for every fixed v_1 and v_2 in V. Then $\|\ \|$ satisfies the fourth condition and in consequence is a norm on V.

Proof. Let u_1 and u_2 be elements of V. We must show that

$$\|u_1 + u_2\| \leq \|u_1\| + \|u_2\|.$$

Towards this we write $v_1 = \frac{1}{2}(u_1 + u_2)$ and $v_2 = \frac{1}{2}(u_1 - u_2)$. Then using the fact that the function f in (3.14) is convex, we have

$$\|\tfrac{1}{2}(u_1 + u_2)\| = f(0) \leq \tfrac{1}{2}(f(-1) + f(1)) = \tfrac{1}{2}(\|u_1\| + \|u_2\|).$$

The desired result now follows from the homogeneity of $\|\ \|$. ■

The next step is to understand convex functions using differential calculus. The precise statement of the result depends on the degree of smoothness of the function. In fact, convex functions necessarily have a certain degree of regularity, but this issue is beyond the scope of this discussion.

Theorem 3.5 *Let $a < b$ and suppose that $f :]a, b[\longrightarrow \mathbb{R}$. Then we have*

- *If f is differentiable, then f is convex on $]a, b[$ iff f' is increasing (in the wide sense) on $]a, b[$.*
- *If f is twice differentiable, then f is convex on $]a, b[$ iff f'' is nonnegative on $]a, b[$.*

Proof. It is enough to prove the first statement. First, we assume that f' is increasing in the wide sense on $]a,b[$. Then, using the Mean Value Theorem we have

$$\begin{aligned}
& (1-t)f(x_1)+tf(x_2)-f((1-t)x_1+tx_2) \\
= \; & (1-t)(f(x_1)-f((1-t)x_1+tx_2))+t(f(x_2)-f((1-t)x_1+tx_2)) \\
= \; & (1-t)t(x_1-x_2)f'(\xi_1)-t(1-t)(x_1-x_2)f'(\xi_2) \qquad (3.15) \\
= \; & (1-t)t(x_1-x_2)(f'(\xi_1)-f'(\xi_2)) \\
\geq \; & 0 \qquad (3.16)
\end{aligned}$$

where in (3.15) ξ_1 is between $(1-t)x_1+tx_2$ and x_1, and ξ_2 is between x_2 and $(1-t)x_1+tx_2$. The crucial point is that the two quantities x_1-x_2 and $\xi_1-\xi_2$ have the same sign. Together with the fact that f' is increasing in the wide sense, this justifies (3.16). This result is often called Jensen's Inequality.

Conversely, suppose that f is convex on $]a,b[$. Let

$$a < x_1 < x_2 < b \qquad (3.17)$$

and suppose that $0 < t < (x_2-x_1)$. Now let

$$s = \frac{x_2-x_1-t}{x_2-x_1}$$

a number that satisfies $0 \leq s \leq 1$ and is defined so that

$$\begin{aligned}
x_1+t &= (1-s)x_1+sx_2 \qquad (3.18) \\
x_2-t &= sx_1+(1-s)x_2 \qquad (3.19)
\end{aligned}$$

Applying the convexity of f to (3.18) and (3.19) yields

$$\begin{aligned}
f(x_1+t) &\leq (1-s)f(x_1)+sf(x_2) \\
f(x_2-t) &\leq sf(x_1)+(1-s)f(x_2)
\end{aligned}$$

which combine to give

$$f(x_1+t)+f(x_2-t) \leq f(x_1)+f(x_2). \qquad (3.20)$$

But (3.20) can be rewritten in the form

$$\frac{f(x_1+t)-f(x_1)}{t} \leq \frac{f(x_2)-f(x_2-t)}{t}.$$

Passing to the limit as $t \to 0$ we find that $f'(x_1) \leq f'(x_2)$. Since x_1 and x_2 are arbitrary points satisfying (3.17) we see that f' is increasing in the wide sense. ■

In theory, our plan should now be to use Lemma 3.4 and Theorem 3.5 to establish the Minkowski Inequality. However, in order to succeed, we will need to get at the second derivative and unfortunately for $1 < p < 2$ the function

$$t \mapsto \{\sum_{j=1}^{n} |x_j+ty_j|^p\}^{\frac{1}{p}}$$

is not twice differentiable. To avoid this problem, we realise that it is enough to establish

$$\left\{\sum_{j=1}^{n}(x_j+y_j)^p\right\}^{\frac{1}{p}} \leq \left\{\sum_{j=1}^{n} x_j^p\right\}^{\frac{1}{p}} + \left\{\sum_{j=1}^{n} y_j^p\right\}^{\frac{1}{p}} \qquad (3.21)$$

in the case that $x_j, y_j \geq 0$. Indeed, by continuity, it will be enough to prove (3.21) in the case $x_j, y_j > 0$. For this, it suffices to establish that $f''(0) \geq 0$ where

$$f(t) = \left\{ \sum_{j=1}^{n} (x_j + t y_j)^p \right\}^{\frac{1}{p}}$$

supposing that $x_j > 0$ and $y_j \in \mathbb{R}$. The proof of this fact follows that of Lemma 3.4. We leave the details to the reader.

This is much better because f is twice differentiable in a neighbourhood of 0. To aid calculations, let us set

$$f(t) = \{g(t)\}^{\frac{1}{p}}, \quad g(t) = \sum_{j=1}^{n} (x_j + t y_j)^p.$$

Then

$$\begin{aligned} f'(t) &= \frac{1}{p} \{g(t)\}^{\frac{1}{p}-1} g'(t), \\ f''(t) &= \frac{1}{p} \left(\frac{1}{p} \Leftrightarrow 1 \right) \{g(t)\}^{\frac{1}{p}-2} (g'(t))^2 + \frac{1}{p} \{g(t)\}^{\frac{1}{p}-1} g''(t). \end{aligned}$$

When the derivatives of g are calculated, the condition $f''(0) \geq 0$ finally boils down to

$$\left\{ \sum_{j=1}^{n} x_j^p \right\} \left\{ \sum_{j=1}^{n} x_j^{p-2} y_j^2 \right\} \geq \left\{ \sum_{j=1}^{n} x_j^{p-1} y_j \right\}^2$$

which is true in light of

$$\sum_{j=1}^{n} \sum_{k=1}^{n} (x_j y_k \Leftrightarrow x_k y_j)^2 x_j^{p-2} x_k^{p-2} \geq 0.$$

Again, we leave the details to the reader.

3.3 The sequence spaces ℓ^p

DEFINITION *Let $1 \leq p < \infty$. The the space ℓ^p is the vector space of all real sequences (x_k) for which the expression*

$$\|(x_k)\|_p = \left(\sum_{k \in \mathbb{N}} |x_k|^p \right)^{\frac{1}{p}} \tag{3.22}$$

is finite. The vector space operations on ℓ^p are defined coordinatewise. Thus

$$(tx + sy)_k = t x_k + s y_k \qquad k \in \mathbb{N}$$

where $x = (x_k)$ and $y = (y_k)$ are given elements of ℓ^p and t and s are reals.

Unfortunately, it is not immediately obvious that ℓ^p is a vector space, nor is it clear that (3.22) defines a genuine norm.

To see that ℓ^p is a vector space, we first invoke Minkowski's Inequality (3.9) in the form

$$\left(\sum_{k=1}^{n} |t x_k + s y_k|^p \right)^{\frac{1}{p}} \leq |t| \left(\sum_{k=1}^{n} |x_k|^p \right)^{\frac{1}{p}} + |s| \left(\sum_{k=1}^{n} |y_k|^p \right)^{\frac{1}{p}} \tag{3.23}$$

Assuming that x, y in ℓ^p and bounding the right hand side of (3.23) we obtain

$$\left(\sum_{k=1}^{n} |tx_k + sy_k|^p\right)^{\frac{1}{p}} \leq |t|||x||_p + |s|||y||_p \tag{3.24}$$

for all $n \in \mathbb{N}$. Letting now n tend to ∞ on the left in (3.24) we find

$$||tx + sy||_p \leq |t|||x||_p + |s|||y||_p.$$

since the left hand side of (3.24) increases with n. This shows simultaneously that ℓ^p is a vector space and that (3.22) defines a norm on ℓ^p.

We would next like to establish the sequence space version of Hölder's Inequality. First, use (3.1) in the form

$$\sum_{j=1}^{n} |x_j||y_j| \leq \left(\sum_{j=1}^{n} |x_j|^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} |y_k|^{p'}\right)^{\frac{1}{p'}}.$$

Again, let n tend to ∞ on the right to obtain

$$\sum_{j=1}^{n} |x_j y_j| \leq ||x||_p ||y||_{p'}$$

Letting n tend to infinity on the left, we again obtain an increasing limit

$$\sum_{j=1}^{\infty} |x_j y_j| \leq ||x||_p ||y||_{p'}.$$

Finally this gives

$$|\sum_{j=1}^{\infty} x_j y_j| \leq ||x||_p ||y||_{p'}. \tag{3.25}$$

the sequence space version of Hölder's Inequality.

We leave the reader to check two points. Firstly, ℓ^2 is an inner product space under

$$\langle (x_j), (y_j) \rangle = \sum_{j=1}^{\infty} x_j y_j.$$

The norm associated to this inner product is just the ℓ^2 norm. Secondly, the space ℓ^p is separable for $1 \leq p < \infty$.

3.4 Premetrics

Some examples of Metric Spaces stem naturally from the concept of a ***premetric***.

DEFINITION *A **premetric function** on a set X is a function $\rho : X \times X \longrightarrow [0, \infty]$ such that*

- $\rho(x, x) = 0 \qquad \forall x \in X.$
- $\rho(x, y) > 0 \qquad \forall x, y \in X$ *such that* $x \neq y$.
- $\rho(x, y) = \rho(y, x) \qquad \forall x, y \in X.$

We think of $\rho(x, y)$ as the cost of moving from x to y in a single step. If it is impossible to move from x to y in a single step, this cost is infinite. A path from x to y is a finite chain $x = x_1, x_2, \ldots, x_{n-1}, x_n = y$ such that each link in the chain can be achieved in a single step, that is

$$\rho(x_j, x_{j+1}) < \infty \qquad \forall j = 1, 2, \ldots, n \Leftrightarrow 1.$$

The metric is then defined by

$$d(x, y) = \inf \sum_{j=1}^{n-1} \rho(x_j, x_{j+1}) \tag{3.26}$$

where the infimum is taken over all paths from x to y. The metric function d then automatically satisfies the triangle inequality. However two remaining conditions have to be checked.

- For all x and y in X there must be some path from x to y in which each link has finite cost.
- It remains to be checked that $d(x, y) = 0$ implies that $x = y$. This may not be easy. It may be possible for a path from x to y with many links of very small cost to have arbitrarily small total cost.

Example Let X be the set of finite character strings on a finite alphabet, say the lower case letters "a" through "z". We say that two strings t and s are adjacent iff one can be transformed into the other by one of the following operations (which simulate typing errors).

- Deletion of a single character anywhere in the string.
- Insertion of a single character anywhere in the string.
- Replacement of one character in the string by some other character.
- Transposition of two adjacent characters in the string.

We define $\rho(s, t) = 0$ if $s = t$, $\rho(s, t) = 1$ if s and t are adjacent and $\rho(s, t) = \infty$ in all other cases. Then (3.26) defines an integer valued metric on X. □

Example A more general example relates to an undirected graph with (possibly infinite) vertex set V and edge set E. We assume that each edge $e \in E$ has a "weight" $w_e > 0$ attached to it. The we can define a premetric function

$$\rho_V(u, v) = \begin{cases} w_{\{u,v\}} & \text{if } \{u, v\} \in E, \\ \infty & \text{otherwise.} \end{cases}$$

Let us assume that the graph is connected in the sense that any two vertices can be linked by a *finite* path (chain of edges). Then it is clear that the function d_V defined by (3.26) is everywhere finite.

The function d_V may fail to be a metric on V however. Consider a graph with vertices x, y and z_{kj} for $j = 1, \ldots, k \Leftrightarrow 1$ and $k \in \mathbb{N}$. The edges are $\{x, y\}$ corresponding to $k = 1$, $\{x, z_{21}\}$ and $\{z_{21}, y\}$ corresponding to $k = 2$, $\{x, z_{31}\}$, $\{z_{31}, z_{32}\}$ and $\{z_{32}, y\}$ corresponding to $k = 3$ and so forth. Let the edges corresponding to a given value of k have weight k^{-2}. The it is clear that $d_V(x, y) = 0$ in spite of the fact that $x \neq y$ since for each $k \in \mathbb{N}$ there is a path from x to y having k links each with weight k^{-2} for a total cost of k^{-1}.

Nevertheless, there will also be many cases in which d_V is a genuine metric. □

Example A more interesting example relates to this last example in case that d_V does define a metric. We construct a set X by "joining with a line segment" any two vertices that are linked by an edge. Thus each edge of the graph is "replaced" by a line segment and these line segments are "glued together" at the vertices. The descriptive notation

$$t\langle u\rangle + (1 \Leftrightarrow t)\langle v\rangle \qquad (\{u, v\} \in E, t \in [0, 1]) \tag{3.27}$$

specifies a typical point of X. The "scalar multiplications" and $+$ in this expression are purely symbolic and in no way represent algebraic operations. It is also understood that the expression $(1 \Leftrightarrow t)\langle v\rangle + t\langle u\rangle$ represents exactly the same point as in (3.27). The point $1\langle v\rangle + 0\langle u\rangle$ represents the vertex v and is independent of u.

A premetric function will now be defined using the same weights $w_e > 0$ of the previous example. Two points can be joined in a single step if they lie on a common segment. Supposing that this segment corresponds to the edge $e = \{u, v\}$, we set

$$\rho(t\langle u\rangle + (1 \Leftrightarrow t)\langle v\rangle, s\langle u\rangle + (1 \Leftrightarrow s)\langle v\rangle) = w_e|t \Leftrightarrow s|$$

for such points. In all other cases we set $\rho(x, y) = \infty$. We assume as before that the graph is connected in the sense that any two vertices can be linked by a *finite* path. Then it is clear that the metric d_X defined by (3.26) is everywhere finite. It remains to show that

$$d_X(x, y) = 0 \qquad \Longrightarrow \qquad x = y.$$

Consider a path from x to y

$$x = x_1, x_2, \ldots, x_{n-1}, x_n = y \tag{3.28}$$

such that each link in the chain can be achieved in a single step. We claim that we can find a chain $x = \xi_1, \xi_2, \ldots, \xi_k = y$ at least as efficient as (3.28) and such that $\xi_2, \ldots, \xi_{k-1}$ are vertex points and all the points in the path are distinct. Typically the path (3.28) will pass through several vertex points. Let us suppose that x_ℓ and x_m are vertex points and that none of the intervening points $x_{\ell+1}, \ldots, x_{m-1}$ are vertex points. Then these intervening points necessarily lie on the same segment and it follows from the extended triangle inequality for $[0, 1]$ that it is at least as efficient to remove them. In this way we obtain a path $x = \xi_1, \xi_2, \ldots, \xi_k = y$ in which $\xi_2, \ldots, \xi_{k-1}$ are vertex points. If a vertex appears twice in this path, for instance if $\xi_p = \xi_q$ with $p < q$, then we can obtain a more efficient path by omitting the points $\xi_{p+1}, \ldots, \xi_q$. We repeat this procedure until all the vertex points in the path are distinct.

At this point, it is clear that if x and y are vertex points then we have $d_X(x, y) = d_V(x, y)$. Furthermore we can calculate d_X completely from d_V. Let $x = t\langle u\rangle + (1 \Leftrightarrow t)\langle v\rangle$ and $y = s\langle z\rangle + (1 \Leftrightarrow s)\langle w\rangle$ corresponding to distinct edges e and f respectively. Then $d_X(x, y)$ is the minimum of the four quantities

$$\begin{aligned}
&(1 \Leftrightarrow t)w_e + (1 \Leftrightarrow s)w_f + d_V(u, z),\\
&tw_e + (1 \Leftrightarrow s)w_f + d_V(v, z),\\
&(1 \Leftrightarrow t)w_e + sw_f + d_V(u, w),\\
&tw_e + sw_f + d_V(v, w).
\end{aligned}$$

If $x = t\langle u\rangle + (1 \Leftrightarrow t)\langle v\rangle$ and $y = s\langle u\rangle + (1 \Leftrightarrow s)\langle v\rangle$ lie on the same segment corresponding to the edge e and we assume without loss of generality that $t < s$, then $d_X(x, y)$ is the minimum of the two quantities

$$\begin{aligned}
&(1 + t \Leftrightarrow s)w_e + d_V(u, v)\\
&(s \Leftrightarrow t)w_e
\end{aligned}$$

In particular one can verify that $d_X(x, y) = 0$ implies $x = y$. □

EXAMPLE One particular example based on the previous one will be needed for a counterexample later in these notes. Let S be any set. Form a graph with vertex set $V = S \cup \{c\}$ where c is a special vertex called the centre. The edges of the graph all have the form $\{c, s\}$ where $s \in S$ and they all have unit weight. The corresponding space X is the ***star space*** based on S. In this case d_X is a metric.

The metric d_X can be described colloquially as follows. If two points lie on the same segment, the distance between them is the standard linear distance along the segment. If two points lie on different segments, then the distance between them is the linear distance of the first to the centre plus the linear distance of the second to the centre. □

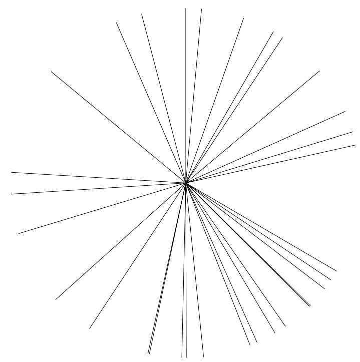

FIGURE 3: *A typical star space.*

3.5 Operator Norms

In this section we study continuous linear mappings between two normed vector spaces.

THEOREM 3.6 *Let U and V be normed vector spaces. Let T be a linear mapping from U to V. Then the following are equivalent.*

- *T is continuous from U to V.*
- *T is continuous at 0_U.*
- *T is uniformly continuous from U to V.*
- *There exists a constant C such that $\|T(u)\|_V \le C\|u\|_U$ for all $u \in U$.*

Proof. We show that the fourth condition implies the third. In the fourth condition we replace u by $u_1 \Leftrightarrow u_2$ where u_1 and u_2 are arbitrary elements of U. Then, using the linearity of T in the form $T(u_1 \Leftrightarrow u_2) = T(u_1) \Leftrightarrow T(u_2)$ we see that the Lipschitz condition (page 17)

$$\|T(u_1) \Leftrightarrow T(u_2)\|_V \le C\|u_1 \Leftrightarrow u_2\|_U$$

holds. Thus T is uniformly continuous.

It is easy to see that the third condition implies the first, and the first condition implies the second.

It remains only to show that the second condition implies the fourth. For this we take $\epsilon = 1$ in the definition of the continuity of T at 0_U. There exists $\delta > 0$ such that

$$\|w\|_U < \delta \Rightarrow \|T(w)\|_V < 1. \tag{3.29}$$

We take $C = 2\delta^{-1}$. Then for $u \in U$ let $w = tu$ where $t = \frac{2}{3}\delta\|u\|^{-1}$. Then $\|w\| = \frac{2}{3}\delta < \delta$ and it follows from (3.29) that $\|T(w)\|_V < 1$ or equivalently $\|T(u)\|_V < \frac{3}{2}\delta^{-1}\|u\| \le C\|u\|$. ■

Sometimes a continuous linear mapping is described as ***bounded linear***. This is a different use of the word "bounded" from the one we have already met — bounded linear maps are not bounded in the metric space sense. Care is needed to make the correct interpretation of the word. We shall make use of the term ***continuous linear*** instead.

The space of all continuous linear maps from U to V is denoted $\mathcal{CL}(U, V)$. For $T \in \mathcal{CL}(U, V)$ we define

$$\|T\|_{\mathcal{CL}(U,V)} = \sup_{\|u\| \le 1} \|T(u)\|. \tag{3.30}$$

It is an exercise to show that (3.30) defines a norm on $\mathcal{CL}(U, V)$ called the ***operator norm***. A most particular case arises when U is finite dimensional. It then turns out that *all* linear mappings from U to V are continuous linear and we can view (3.30) as defining a norm on the space $\mathcal{L}(U, V)$ of all such linear maps. This fact is by no means obvious and can be obtained as a consequence of Corollary 5.17 on page 71.

If U, V and W are all normed spaces, $T \in \mathcal{CL}(U, V)$ and $S \in \mathcal{CL}(V, W)$ then it is easy to see that

$$||S \circ T|| \leq ||S|| ||T||$$

where all the norms are the appropriate operator norms. A particular case that is often used arises when all the space are equal. If $T \in \mathcal{CL}(U, U)$ then

$$||T^n|| \leq ||T||^n \quad \forall n \in \mathbb{Z}^+. \tag{3.31}$$

In particular the operator norm of the identity map I is unity. This is the case $n = 0$ in (3.31).

Operator norms are in general difficult to compute explicitly. When the underlying norms are Euclidean, the following result from linear algebra helps.

THEOREM 3.7 (SINGULAR VALUE DECOMPOSITION THEOREM) *Let A be a real $m \times n$ matrix. Then there exist orthogonal matrices U and V of shapes $m \times m$ and $n \times n$ respectively such that $A = UBV$ and B satisfies $b_{jk} = 0$ if $j \neq k$. Further the diagonal values b_{jj} for $j = 1, 2, \ldots, \min(m, n)$ may be taken nonnegative. They are called the **singular values** of A. Any two such decompositions yield the same singular values up to rearrangement.*

If $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is given by

$$(Tx)_i = \sum_{j=1}^{n} a_{ij} x_j$$

and the norms taken on $\mathbb{R}^n$ and $\mathbb{R}^m$ are the Euclidean norms, then the operator norm of T is seen to be the largest singular value of the $m \times n$ matrix A. We leave the proof to the reader.

3.6 Continuous Linear Forms

If V is a normed real vector space, then it has a ***dual space***, $V^\star$, the linear space of all linear mappings from V to $\mathbb{R}$. If V is finite dimensional, then *all* linear forms on V are continuous (this fact is not obvious and can be obtained from Corollary 5.17). If V is infinite dimensional, then one may have linear forms that are not continuous.

EXAMPLE Let F be the linear subspace of ℓ^1 of finitely supported sequences (x_n). A sequence is finitely supported if it satisfies the criterion

$$\exists N \in \mathbb{N} \text{ such that } (x_n = 0 \qquad \forall n \geq N).$$

Then F is a normed vector space with the norm restricted from ℓ^1. The form φ given by

$$(x_n) \quad \stackrel{\varphi}{\longrightarrow} \quad \sum_{n=1}^{\infty} n x_n \tag{3.32}$$

is a perfectly good linear form on F. Note that the sum in (3.32) is in fact a finite sum, even though it is written as an infinite one. This ensures that φ is everywhere defined in F. Clearly $\varphi(e_n) = n$ while $||e_n||_{\ell^1} = 1$ so that φ is not continuous on F. □

EXAMPLE Finding a discontinuous linear form on ℓ^1 itself is considerably more difficult. Let e_n denote the sequence in ℓ^1 that has a 1 in the n-th place and 0 everywhere else. Consider the set $\{e_1, e_2, \ldots\}$. This set is linearly independent in ℓ^1. It is not however a spanning set. In fact its linear span is just the set F of the previous example. According to the basis extension theorem for linear spaces, we may extend it to a basis of ℓ^1 with $\{f_\alpha; \alpha \in I\}$, where I is some index set. It needs to be said that linear bases in infinite dimensional spaces are strange things — one needs to be able to write *every* vector in the space as a *finite* linear combination of basis vectors. So strange in fact that the ***Axiom of Choice*** is normally used in order to show the existence of $\{f_\alpha; \alpha \in I\}$. Indeed, nobody has ever written down an *explicit* example of such a set! Let f be one of these f_α's. Then we define the form θ as the mapping that takes an element of ℓ^1 to the coefficient of f in its corresponding basis representation. We will show that θ fails to be continuous. Indeed, θ vanishes on F, because the basis representation of an element (x_n) of F is just the *finite* sum

$$(x_n) = \sum_{n=1}^{\infty} x_n e_n.$$

The term in f is not required and the corresponding coefficient is zero. Since F is a dense subset of ℓ^1, it follows from Proposition 2.20, that if θ were continuous then θ would be identically zero. But $\theta(f) = 1$ so this is not possible. □

The space V' of continuous linear forms on a normed vector space V has a natural norm, namely the operator norm (using $|\ |$ as a norm on $\mathbb{R}$).

$$\|\varphi\|_{V'} = \sup_{\|v\|_V \le 1} |\varphi(v)|. \tag{3.33}$$

A most interesting situation develops in the case that V is finite dimensional and $V' = V^\star$. Here, since $V^\star$ is itself again finite dimensional, the second dual $V'' = V^{\star\star}$ and it is well known that $V^{\star\star}$ and V are naturally isomorphic. This allows us to construct a second dual norm on V. The following result asserts that this second dual norm is identical to the original norm on V.

PROPOSITION 3.8 *Let V be a finite dimensional real normed vector space. Let $v \in V$. Then*

$$\|v\| = \sup_{\|\varphi\|_{V'} \le 1} |\varphi(v)|. \tag{3.34}$$

This Proposition will be obtained as a consequence of the following Theorem, the proof of which will be given later (on page 94).

THEOREM 3.9 (SEPARATION THEOREM FOR CONVEX SETS) *Let C be an open convex subset of V a finite dimensional real normed vector space. Let $v \in V \setminus C$. Then there exists a linear form φ on V such that $\varphi(c) < \varphi(v)$ for all $c \in C$.*

Geometrically, the conclusion of Theorem 3.9 asserts that the convex set C lies entirely in the open halfspace

$$\{u; u \in V, \varphi(u) < \varphi(v)\}$$

whose boundary is the affine hyperplane

$$M = \{u; u \in V, \varphi(u) = \varphi(v)\}.$$

Proof of Proposition 3.8. Clearly we have

$$\sup_{\|\varphi\|_{V'} \le 1} |\varphi(v)| \le \|v\|. \tag{3.35}$$

If equality does not hold in (3.35), then after renormalizing we can suppose that there exists $v \in V$ such that $\|v\| = 1$ while $|\varphi(v)| < \|\varphi\|_{V'}$ for all $\varphi \in V'$. Now let $C = \{c; c \in V, \|c\| < 1\}$ the open unit ball of V. Clearly C is an open convex subset of V. Also $v \notin C$. Hence, by the Separation Theorem, there exists $\theta \in V'$ such that $\theta(c) < \theta(v)$ for all $c \in C$. But since $c \in C$ implies that $\Leftrightarrow c \in C$ we can also write

$$|\theta(c)| < \theta(v) \qquad \forall c \in C.$$

It now follows that $\|\theta\|_{V'} \leq \theta(v)$. Finally this yields

$$\|\theta\|_{V'} \leq \theta(v) \leq |\theta(v)| < \|\theta\|_{V'},$$

the required contradiction. ■

3.7 Equivalent Metrics

There are various notions of ***equivalence*** for metrics on a set X. The standard definition follows.

Definition *Let X be a set and suppose that d_1 and d_2 are two metrics on X. Then d_1 and d_2 are said to be **metrically equivalent** if there exist constants C_1 and C_2 with $0 < C_j < \infty$ for $j = 1, 2$ such that*

$$C_1 d_1(x_1, x_2) \leq d_2(x_1, x_2) \leq C_2 d_1(x_1, x_2)$$

for all $x_1, x_2 \in X$.

We leave the reader to show that metric equivalence is an equivalence relation. This is the strongest form of equivalence. Some authors call this ***uniform equivalence***, but in these notes we reserve this terminology for a concept to be defined shortly. Metric equivalence is not very useful, except in the case of normed spaces where there is really only one form of equivalence and we drop the adverb *metrically*.

Definition *Let V be a vector space over $\mathbb{R}$ or $\mathbb{C}$. Then two norms $\|\ \|_1$ and $\|\ \|_2$ are **equivalent** iff there exist strictly positive constants C_1 and C_2 such that*

$$C_1 \|v\|_1 \leq \|v\|_2 \leq C_2 \|v\|_1$$

for all $v \in V$.

In the metric space context there are much more interesting forms of equivalence that preserve underlying properties.

Definition *Let X be a set and suppose that d_1 and d_2 are two metrics on X. Then d_1 and d_2 are said to be **topologically equivalent** iff the metric spaces (X, d_1) and (X, d_2) have the same open sets.*

It is clear from the definition that topological equivalence is an equivalence relation. There is a more subtle way of rephrasing the definition. Two metrics d_1 and d_2 are topologically equivalent iff the identity mapping I_X on X is continuous as a mapping from (X, d_1) to (X, d_2) and also from (X, d_2) to (X, d_1). This makes it clear that one could also say that two metrics are topologically equivalent iff they have the same convergent sequences. There are many other equivalent formulations.

This idea also suggests the final form of equivalence.

Definition *Let X be a set and suppose that d_1 and d_2 are two metrics on X. Then d_1 and d_2 are said to be **uniformly equivalent** iff the identity mapping I_X on X is uniformly continuous as a mapping from (X, d_1) to (X, d_2) and also as a map from (X, d_2) to (X, d_1).*

Metric equivalence implies uniform equivalence and uniform equivalence implies topological equivalence.

For normed spaces, all three forms of equivalence are the same. This follows immediately from Theorem 3.6.

EXAMPLE On $\mathbb{R}$ consider

- $d_1(x,y) = |x \Leftrightarrow y|$.
- $d_2(x,y) = 2|x \Leftrightarrow y|$.
- $d_3(x,y) = \arctan(|x \Leftrightarrow y|)$.
- $d_4(x,y) = |\arctan(x) \Leftrightarrow \arctan(y)|$.

It is not immediately obvious that d_3 is a metric. To see this, one needs to establish

$$\arctan(x+y) \leq \arctan x + \arctan y \tag{3.36}$$

for $x, y \geq 0$. Let $t = \arctan x$ and $s = \arctan y$. Then in case that $t + s \geq \frac{\pi}{2}$, (3.36) is obvious. Thus, we may assume that $t, s \geq 0$ and that $t + s < \frac{\pi}{2}$. We need to show that

$$\tan t + \tan s \leq \tan(t+s). \tag{3.37}$$

But (3.37) follows from the trig identity

$$\tan(t+s) = \frac{\tan t + \tan s}{1 \Leftrightarrow \tan t \tan s},$$

and the observation that $1 \geq 1 \Leftrightarrow \tan t \tan s > 0$ since $t + s < \frac{\pi}{2}$.

It is immediately obvious that d_1, d_2 and d_4 are metrics on $\mathbb{R}$. It is straightforward to show that d_1 and d_2 are metrically equivalent, that d_1 and d_3 are uniformly equivalent, but not metrically equivalent and finally that d_1 and d_4 are topologically equivalent but not uniformly equivalent. □

3.8 The Abstract Cantor Set

Consider X_j to be a copy of the two point space $\{0,1\}$ for $j = 1, 2, 3, \ldots$. To define the abstract Cantor set X we simply consider

$$X = \prod_{j=1}^{\infty} X_j,$$

the infinite product of the X_j. In effect, a point x of X is a sequence $x = (x_j)$ with $x_j \in \{0,1\}$ for $j = 1, 2, 3, \ldots$.

Next, we define a metric on X. We want entries far up the sequence to have less weight than entries near the beginning of the sequence, so we define

$$d((x_j),(y_j)) = \sum_{j=1}^{\infty} 2^{-j} |x_j \Leftrightarrow y_j|.$$

Observe that since $|x_j \Leftrightarrow y_j| \leq 2$, the series always converges. The use of the weights 2^{-j} is somewhat arbitrary here. It is routine to verify that d defines a metric on X. It will be observed that convergence in (X,d) is ***coordinatewise*** or ***pointwise convergence***. The case is somewhat special here because two coordinates either agree or differ by 1.

We see that if (x_j) and (y_j) agree in the first k coordinates, then $d((x_j),(y_j)) \leq 2^{-k}$. Conversely, if $d((x_j),(y_j)) \leq 2^{-k}$ then (x_j) and (y_j) agree in the first $k \Leftrightarrow 1$ coordinates.

The mapping $\alpha : X \Leftrightarrow\!\rightarrow \mathbb{R}$ given by

$$\alpha((x_j)) = 2 \sum_{j=1}^{\infty} 3^{-j} x_j$$

maps X onto the standard Cantor set in $\mathbb{R}$. The metric

$$d_1((x_j),(y_j)) = |\alpha(x) \Leftrightarrow \alpha(y)| = 2\sum_{j=1}^{\infty} |3^{-j}(x_j \Leftrightarrow y_j)|.$$

on X reflects the standard metric on $\mathbb{R}$ through the mapping α. For k an integer with say $k \geq 3$ it is easy to show that

$$d(x,y) \leq 2^{-(k+1)} \Longrightarrow d_1(x,y) \leq 3^{-k}$$

and

$$d_1(x,y) \leq 3^{-(k+1)} \Longrightarrow d(x,y) \leq 2^{-k}.$$

It follows that d and d_1 are uniformly equivalent metrics on X.

3.9 The Quotient Norm

Let V be a normed vector space and let N be a closed linear subspace. Then we can consider the quotient space $Q = V/N$. This is a new linear space with a complicated definition highly unpopular with students.

One starts by defining a relation $\sim$ on V by

$$v_1 \sim v_2 \qquad \Leftrightarrow \qquad v_1 \Leftrightarrow v_2 \in N.$$

Here v_1, v_2 denote elements of V. One verifies that $\sim$ is an equivalence relation on V. There is then a quotient space Q and a canonical projection π,

$$\pi : V \Leftrightarrow\!\rightarrow Q.$$

It is now possible to show that Q can be given the structure of a linear space in such a way that π is a linear mapping. In addition one has that $\ker(\pi) = N$.

We now define a norm on Q known as the quotient norm by

$$||q||_Q = \inf_{\pi(v)=q} ||v||_V, \tag{3.38}$$

for $q \in Q$. The infimum is taken over all elements $v \in V$ such that $\pi(v) = q$. It is more or less obvious that $|| \; ||_Q$ is homogenous.

To show the subadditivity of the norm, we argue by contradiction. Suppose that there exists $\epsilon > 0$, $q_1, q_2 \in Q$ such that

$$||q_1 + q_2|| \geq ||q_1|| + ||q_2|| + 3\epsilon. \tag{3.39}$$

Then using the definition (3.38), we can find $v_1, v_2 \in V$ such that $\pi(v_j) = q_j$ and

$$||v_j||_V \leq ||q_j|| + \epsilon,$$

for $j = 1, 2$. Obviously, $\pi(v_1 + v_2) = q_1 + q_2$ so that

$$||q_1 + q_2|| \leq ||v_1|| + ||v_2|| \leq ||q_1|| + ||q_2|| + 2\epsilon.$$

This contradiction with (3.39) establishes the subadditivity.

There is one final detail that requires a little proof. Suppose that $q \in Q$ and that $||q||_Q = 0$. Then, using (3.38) we can find a sequence (v_j) of elements of V with $\pi(v_j) = q$ for $j = 1, 2, \ldots$ and $||v_j||$ tending to zero. Clearly $v_j \Leftrightarrow\!\rightarrow 0_V$ and hence $(v_1 \Leftrightarrow v_j) \Leftrightarrow\!\rightarrow v_1$. Since $(v_1 \Leftrightarrow v_j) \in N$ and since N is supposed to be closed in V, we conclude that $v_1 \in N$ and consequently that $q = 0_Q$.

Some nice points lie outside our present reach since they depend on compactness. If V is finite dimensional then any linear subspace N of V is necessarily closed. Furthermore, in this case, the infimum of (3.38) is necessarily attained. A consequence is that the unit ball of Q is just the direct image by π of the unit ball of V. In the finite dimensional case, this gives a geometric way of understanding the quotient norm.

It should be pointed out that one can try to define general quotient metrics in much that same way, but the issues are much more problematic. If

$$\pi : X \twoheadrightarrow Q.$$

is a quotienting of a metric space X we can define

$$d_Q(q_1, q_2) = \inf_{\substack{\pi(x_1)=q_1 \\ \pi(x_2)=q_2}} d_X(x_1, x_2)$$

but only under very stringent additional assumptions will this define a metric on Q.

4

Completeness

In this chapter we will assume that the reader is familiar with the completeness of $\mathbb{R}$. Usually $\mathbb{R}$ is defined as the unique order-complete totally ordered field. The order completeness postulate is that every subset B of $\mathbb{R}$ which is bounded above possesses a least upper bound (or supremum). From this the metric completeness of $\mathbb{R}$ is deduced. Metric completeness is formulated in terms of the convergence of Cauchy sequences. It is true that in making the link between the two for $\mathbb{R}$, one uses the Bolzano–Weierstrass Theorem which is a form of compactness. Nevertheless, we believe that for metric spaces, completeness is a more fundamental concept than compactness and should be treated first.

DEFINITION *Let X be a metric space. Let (x_n) be a sequence in X. Then (x_n) is a* ***Cauchy sequence*** *iff for every number $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that*

$$p, q > N \qquad \Rightarrow \qquad d(x_p, x_q) < \epsilon.$$

LEMMA 4.1 *Every convergent sequence is Cauchy.*

Proof. Let X be a metric space. Let (x_n) be a sequence in X converging to $x \in X$. Then given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \frac{1}{2}\epsilon$ for $n > N$. Thus for $p, q > N$ the triangle inequality gives

$$d(x_p, x_q) \leq d(x_p, x) + d(x, x_q) < \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon = \epsilon.$$

Hence (x_n) is Cauchy. ■

The Cauchy condition on a sequence says that the diameters of the successive tails of the sequence converge to zero. One feels that this is almost equivalent to convergence except that no limit is explicitly mentioned. Sometimes, Cauchy sequences fail to converge because the "would be limit" is not in the space. It is the existence of such "gaps" in the space that prevent it from being complete.

DEFINITION *Let X be a metric space. Then X is* ***complete*** *iff every Cauchy sequence in X converges in X.*

EXAMPLE The real line $\mathbb{R}$ is complete. □

EXAMPLE The set $\mathbb{Q}$ of rational numbers is not complete. Consider the sequence defined inductively by

$$x_1 = 2 \quad \text{and} \quad x_{n+1} = \frac{1}{2}\left(x_n + \frac{2}{x_n}\right), \quad n = 1, 2, \ldots. \tag{4.1}$$

Then one can show that (x_n) converges to $\sqrt{2}$ in $\mathbb{R}$. It follows that (x_n) is a Cauchy sequence in $\mathbb{Q}$ which does not converge in $\mathbb{Q}$. Hence $\mathbb{Q}$ is not complete.

To fill in the details, observe first that (4.1) can also be written in both of the alternative forms

$$\begin{aligned} 2x_n(x_{n+1} \Leftrightarrow \sqrt{2}) &= (x_n \Leftrightarrow \sqrt{2})^2, \\ x_{n+1} \Leftrightarrow x_n &= \Leftrightarrow \left(\frac{x_n^2 \Leftrightarrow 2}{2x_n} \right). \end{aligned}$$

We now observe the following in succession.

- $x_n > 0$ for all $n \in \mathbb{N}$.
- $x_n > \sqrt{2}$ for all $n \in \mathbb{N}$.
- x_n is decreasing with n.
- $x_n \leq 2$ for all $n \in \mathbb{N}$.
- $|x_{n+1} \Leftrightarrow \sqrt{2}| \leq \dfrac{|x_n \Leftrightarrow \sqrt{2}|^2}{2\sqrt{2}}$ for all $n \in \mathbb{N}$.
- $|x_{n+1} \Leftrightarrow \sqrt{2}| \leq \dfrac{2 \Leftrightarrow \sqrt{2}}{2\sqrt{2}} |x_n \Leftrightarrow \sqrt{2}|$ for all $n \in \mathbb{N}$.

The convergence of (x_n) to $\sqrt{2}$ follows easily. □

4.1 Boundedness and Uniform Convergence

DEFINITION *Let A be a subset of a metric space X. Then the* ***diameter*** *of A is defined by*

$$\operatorname{diam}(X) = \sup_{x_1, x_2 \in A} d(x_1, x_2). \tag{4.2}$$

We say that A is a ***bounded subset*** *iff $\operatorname{diam}(A) < \infty$. We note that we regard the subset $A = \emptyset$ as being bounded, even though formally the supremum in (4.2) is illegal. We say that the metric space X is bounded iff it is a bounded subset of itself. We say that a sequence (x_n) is bounded iff its underlying set is bounded. We say that a function $f : Y \Leftrightarrow\rightarrow X$ is bounded iff $f(Y)$ is a bounded subset of X.*

These definitions are coloured by the fact that they relate to metric spaces. In the case of a normed vector space V, there is another alternative, provided by the distinguished element 0_V.

DEFINITION *A subset A of a normed vector space V is bounded iff*

$$\sup_{v \in A} \|v\| < \infty.$$

A moment's thought shows that the two concepts of boundedness are equivalent.

The following lemma follows directly from the definition of a Cauchy sequence and Lemma 4.1.

LEMMA 4.2

- *Every Cauchy sequence is bounded.*
- *Every convergent sequence is bounded.*

Our main immediate motivation for introducing boundedness at this point is the construction of additional examples of complete metric spaces. Let X be a non-empty set and Y a metric space. Then we denote by $B(X, Y)$ the set of all bounded mappings from X to Y. We can turn this into a metric space by defining the metric

$$d(f, g) = \sup_{x \in X} d_Y(f(x), g(x)). \tag{4.3}$$

This is the ***supremum metric*** or ***uniform metric***. In case that Y is a normed vector space, it is easy to check that

$$\|f\|_{B(X,Y)} = \sup_{x \in X} \|f(x)\|_Y \tag{4.4}$$

is a norm on $B(X, Y)$ and that the metric that it induces is precisely the one given by (4.3). The norm defined by (4.4) is called the ***supremum norm*** or ***uniform norm***.

The convergence of $B(X, Y)$ is called ***uniform convergence***. On the other hand, we say that a sequence of functions (f_n) converges pointwise iff $(f_n(x))$ converges for every x in X. If a sequence converges uniformly, then it converges pointwise. But the converse is not true.

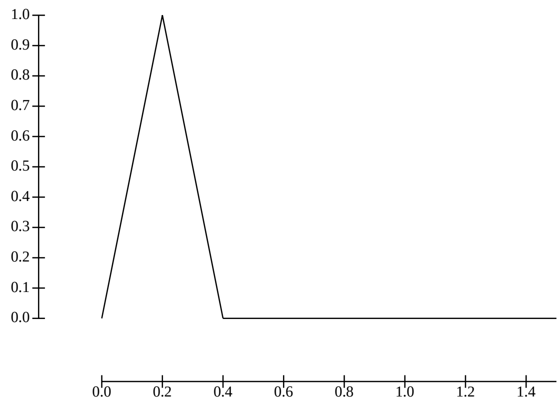

FIGURE 4: *The function f_5.*

EXAMPLE Consider the case $X = \mathbb{R}^+$, $Y = \mathbb{R}$ and suppose that the sequence (f_n) is given by

$$f_n(x) = \begin{cases} nx & \text{if } 0 \le x \le n^{-1}, \\ 2 - nx & \text{if } n^{-1} \le x \le 2n^{-1}, \\ 0 & \text{if } x \ge 2n^{-1}. \end{cases} \tag{4.5}$$

Now, if $x = 0$ then $f_n(0) = 0$ for all n so that $f_n(0) \to 0$ as $n \to \infty$. On the other hand, if $x > 0$, then for large values of n, it is the third line of (4.5) that applies. Once again, we obtain $f_n(x) \to 0$. Thus the sequence (f_n) converges pointwise to 0.

This convergence is not uniform. To see this, we simply put $x = n^{-1}$. Then $f_n(\frac{1}{n}) = 1$ and it follows that $d_{B(\mathbb{R}^+,\mathbb{R})}(f_n, 0) = 1$ for all n. □

While uniform convergence is just convergence in the metric of $B(X, Y)$, there is *no metric* which gives rise to pointwise convergence. From the point of view of Metric Spaces, pointwise convergence of sequences of functions is some kind of *rogue* convergence that does not fit the theory. In these notes we just have to live with this unfortunate circumstance. However there is a topology on $B(X,Y)$ (and on other spaces of functions) with the property that convergence in this topology is exactly pointwise convergence. The need to have a unified theory of convergence therefore forces one into the realm of topological spaces with all of its associated pathology.

PROPOSITION 4.3 *If Y is complete, so is $B(X,Y)$.*

Proof. The pattern of most completeness proofs is the same. Take a Cauchy sequence. Use some existing completeness information to deduce that the sequence converges in some weak sense. Use the Cauchy condition again to establish that the sequence converges in the metric sense.

Let (f_n) be a Cauchy sequence in $B(X,Y)$. Then, for each $x \in X$, it is straightforward to check that $(f_n(x))$ is a Cauchy sequence in Y and hence converges to some element of Y. This can be viewed as a rule for assigning an element of Y to every element of X — in other words, a function f from X to Y. We have just shown that (f_n) converges to f pointwise.

Now let $\epsilon > 0$. Then for each $x \in X$ there exists $N_x \in \mathbb{N}$ such that

$$q > N_x \qquad \Rightarrow \qquad d(f_q(x), f(x)) < \tfrac{1}{3}\epsilon. \tag{4.6}$$

Now we reuse the Cauchy condition — there exists $N \in \mathbb{N}$ such that

$$p, q > N \qquad \Rightarrow \qquad \sup_{x \in X} d(f_p(x), f_q(x)) < \tfrac{1}{3}\epsilon. \tag{4.7}$$

Now, combining (4.6) and (4.7) with the triangle inequality and choosing q explicitly as $q = \max(N, N_x) + 1$, we find that

$$p > N \qquad \Rightarrow \qquad d(f_p(x), f(x)) < \tfrac{2}{3}\epsilon \quad \forall x \in X. \tag{4.8}$$

We emphasize the crucial point that N depends only on ϵ. It does not depend on x. Thus we may deduce

$$p > N \qquad \Rightarrow \qquad \sup_{x \in X} d(f_p(x), f(x)) < \epsilon. \tag{4.9}$$

from (4.8).

This would be the end of the proof, if it were not for the fact that we still do not know that $f \in B(X,Y)$. For this, choose an explicit value of ϵ, say $\epsilon = 1$. Then, using the corresponding specialization of (4.9), we see that there exists $r \in \mathbb{N}$ such that

$$\sup_{x \in X} d(f_r(x), f(x)) < 1. \tag{4.10}$$

Now, substitute (4.10) into the extended triangle inequality

$$d(f(x_1), f(x_2)) \le d(f(x_1), f_r(x_1)) + d(f_r(x_1), f_r(x_2)) + d(f_r(x_2), f(x_2))$$

to obtain

$$d(f(x_1), f(x_2)) \le 1 + d(f_r(x_1), f_r(x_2)) + 1.$$

It now follows that since f_r is bounded, so is f. Finally, with the knowledge that $f \in B(X,Y)$ we see that (f_n) converges to f in $B(X,Y)$ by (4.9). ■

There is an alternative way of deducing (4.9) from (4.7) which worth mentioning. Conceptually it is simpler than the argument presented above, but perhaps less rigorous. We write (4.7) in the form

$$p, q > N \qquad \Rightarrow \qquad d(f_p(x), f_q(x)) < \tfrac{1}{3}\epsilon. \tag{4.11}$$

where x is a general point of X. The vital key is that N depends only on ϵ and not on x. Now, letting $q \Leftrightarrow\!\!\rightarrow \infty$ in (4.11) we find

$$p > N \qquad \Rightarrow \qquad d(f_p(x), f(x)) \le \tfrac{1}{3}\epsilon. \tag{4.12}$$

because $f_q(x)$ converges pointwise to $f(x)$. Here we are using the fact that $[0, \frac{1}{3}\epsilon]$ is a closed subset of $\mathbb{R}$. Since N depends only on ϵ we can then deduce (4.9) from (4.12).

Example An immediate Corollary of the above is that the space ℓ^∞ is complete. The same is true of ℓ^p for $1 \le p < \infty$. We sketch the details. Let (x_n) be a Cauchy sequence of elements of ℓ^p. Then each such element x_n is actually a sequence x_{nk} of real numbers. It is easy to see that for each fixed k, $(x_{nk})_{n=1}^\infty$ is a Cauchy sequence in $\mathbb{R}$. Using the completeness of $\mathbb{R}$ we infer the existence of $\xi_k \in \mathbb{R}$ such that

$$x_{nk} \longrightarrow \xi_k$$

as $n \to \infty$. We now use again the fact that (x_n) is a Cauchy sequence. Let $\epsilon > 0$. Then there exist $N \in \mathbb{N}$ such that

$$m, n > N \Longrightarrow \|x_m - x_n\|_p < \epsilon.$$

Then, for all $m, n > N$ and for all $K \in \mathbb{N}$ we have

$$\sum_{k=1}^{K} |x_{mk} - x_{nk}|^p \le \epsilon^p.$$

Letting $m \to \infty$, this leads to

$$\sum_{k=1}^{K} |\xi_k - x_{nk}|^p \le \epsilon^p,$$

because only finitely many values of k are involved. Finally letting $K \to \infty$ we get $\|\xi - x_n\|_p \le \epsilon$ for all $n > N$. This gives the desired convergence to an element ξ of ℓ^p. As above, a little extra work is necessary to show that $\xi \in \ell^p$. □

4.2 Subsets and Products of Complete Spaces

We seek other ways of building new complete spaces from old.

Proposition 4.4

- *Let X be a complete metric space. Let Y be a closed subset of X. Then Y is complete (as a metric space in its own right).*
- *Let X be a metric space. Let Y be a subset of X which is complete (as a metric space in its own right). Then Y is closed in X.*

Proof. We establish the first statement. Let (x_n) be a Cauchy sequence in Y. Then (x_n) is a Cauchy sequence in X. Since X is complete, there exists $x \in X$ such that (x_n) converges to x. But since the sequence (x_n) lies in Y and Y is closed, $x \in Y$.

For the second statement, let (x_n) be a sequence in Y converging to some element x of X. We aim to show that $x \in Y$. By Lemma 4.1, (x_n) is a Cauchy sequence in X. Hence (x_n) is also a Cauchy sequence in Y. But Y is complete. It follows that there exists an element $y \in Y$ such that (x_n) converges to y. Then by Proposition 2.3, $x = y$. Hence $x \in Y$. ■

This Proposition gives the correct impression that completeness is a kind of global closedness property. We have the following useful corollary.

Corollary 4.5 *Let X and Y be metric spaces. Let $f : X \longrightarrow Y$ be an isometry. Suppose that X is complete and that $f(X)$ is dense in Y. Then f is onto.*

Proof. Since X is complete and f is an isometry, $f(X)$ is complete and hence closed in Y. But since $f(X)$ is also dense in Y, $f(X) = Y$. ■

PROPOSITION 4.6 *Let X and Y be complete spaces. Then so is $X \times Y$.*

Proof. Let (x_n, y_n) be a Cauchy sequence in $X \times Y$. Then it is easy to see that the component sequences (x_n) and (y_n) are Cauchy in X and Y respectively. Since X and Y are both complete, it follows that there exist limits x and y respectively. The result now follows directly from Lemma 2.13. ■

PROPOSITION 4.7 *Let X and Y be metric spaces. Let (f_n) be a sequence of continuous functions $f_n : X \Leftrightarrow\!\rightarrow Y$, converging uniformly to a function f. Then f is also continuous.*

Proof. The proof we give is by epsilons and deltas. Let $x_0 \in X$ — we will show that f is continuous at x_0. Suppose that $\epsilon > 0$. Then by the uniform convergence, there exists $N \in \mathbb{N}$ such that

$$n > N \qquad \Rightarrow \qquad \sup_{x \in X} d_Y(f(x), f_n(x)) < \tfrac{1}{3}\epsilon. \tag{4.13}$$

Let us fix $n = N + 1$. Now we use the fact that this one function f_n of the sequence is continuous at x_0. There exists $\delta > 0$ such that

$$d_X(x, x_0) < \delta \qquad \Rightarrow \qquad d_Y(f_n(x), f_n(x_0)) < \tfrac{1}{3}\epsilon. \tag{4.14}$$

Combining (4.13) and (4.14) we now obtain for $x \in X$ satisfying $d_X(x, x_0) < \delta$

$$\begin{aligned} d_Y(f(x), f(x_0)) &\leq d_Y(f(x), f_n(x)) + d_Y(f_n(x), f_n(x_0)) + d_Y(f_n(x_0), f(x_0)) \\ &< \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon = \epsilon. \end{aligned}$$

This shows that f is continuous. ■

EXAMPLE We give next an important example of Proposition 4.4. Let X and Y be metric spaces. Then we denote by $C(X, Y)$ the subset of $B(X, Y)$ consisting of bounded continuous functions from X to Y. We claim that $C(X, Y)$ is a closed subset of $B(X, Y)$. This is an immediate consequence of Proposition 4.7. Applying the first assertion of Proposition 4.4 shows that $C(X, Y)$ is a complete space if Y is. □

EXAMPLE The star space X based on a set S is always complete. The definition is found on page 40. Let (x_n) be any sequence in X. Then we can denote $x_n = (1 \Leftrightarrow t_n)\langle c\rangle + t_n\langle s_n\rangle$ where c denotes the centre of X, (t_n) is a sequence in $[0, 1]$ and (s_n) is a sequence in S.

Suppose now that (x_n) is a Cauchy sequence in X. If $t_n \Leftrightarrow\!\rightarrow 0$ as $n \Leftrightarrow\!\rightarrow \infty$ then $x_n \Leftrightarrow\!\rightarrow c$ in X and we have convergence. Thus, to establish convergence we may assume that there exists a strictly positive number ϵ such that for all $N \in \mathbb{N}$ there exists $p > N$ such that $t_p > \epsilon$. Apply now the Cauchy condition with this ϵ. There exists $N \in \mathbb{N}$ such that

$$p, q > N \qquad \Rightarrow \qquad d_X(x_p, x_q) < \epsilon.$$

Choose p as described above. If $s_p \neq s_q$ so that x_p and x_q lie on different rays then $d_X(x_p, x_q) = t_p + t_q > \epsilon$ a contradiction. Hence $s_q = s_p$ for all $q > N$. Thus a tail of the sequence lies on a single ray where the metric is essentially just that of $[0, 1]$. In other words, (t_n) is a Cauchy sequence in $[0, 1]$. Since $[0, 1]$ is complete (it is a closed subset of $\mathbb{R}$), it follows that there exists $t \in [0, 1]$ such that $t_n \Leftrightarrow\!\rightarrow t$. The sequence (s_n) is eventually constant, so that (x_n) converges to $(1 \Leftrightarrow t)\langle c\rangle + t\langle s_p\rangle$. □

A major application of completeness in normed spaces is the existence of ***absolutely convergent sums***.

PROPOSITION 4.8 *Let V be a complete normed space and let v_j be elements of V for $j \in \mathbb{N}$. Suppose that*

$$\sum_{j=1}^{\infty} \|v_j\|_V < \infty.$$

Then the sequence of partial sums (s_n) *given by*

$$s_n = \sum_{j=1}^{n} v_j$$

converges to an element $s \in V$. *Furthermore we have the norm estimate*

$$\|s\|_V \le \sum_{j=1}^{\infty} \|v_j\|_V . \tag{4.15}$$

In this situation, it is natural to denote

$$s = \sum_{j=1}^{\infty} v_j .$$

Proof. We show that (s_j) is a Cauchy sequence. Let $1 \le q \le p$. Then applying the extended triangle inequality (1.10) to

$$s_p - s_q = \sum_{j=q+1}^{p} v_j ,$$

we obtain

$$\|s_p - s_q\|_V \le \sum_{j=q+1}^{p} \|v_j\|_V \le \sum_{j=q+1}^{\infty} \|v_j\|_V . \tag{4.16}$$

But since the right hand term of (4.16) tends to 0 as $q \to \infty$, it follows that (s_j) is a Cauchy sequence. Since V is complete we deduce that (s_j) converges to some element s of V. Putting $q = 0$ in (4.16) shows that

$$\|s_p\|_V \le \sum_{j=1}^{\infty} \|v_j\|_V . \tag{4.17}$$

Since the norm is continuous on V, we see that (4.15) follows from (4.17) as $p \to \infty$. ■

4.3 Contraction Mappings

Definition *Let* X *be a metric space. Let* $f : X \to X$. *Then* f *is a* ***contraction mapping*** *iff there exists a constant* α *with* $0 \le \alpha < 1$ *such that*

$$d_X(f(x_1), f(x_2)) \le \alpha d_X(x_1, x_2).$$

The following Theorem will be used extensively in the calculus section of this book.

Theorem 4.9 (Contraction Mapping Theorem) *Let* f *be a contraction mapping on a complete non-empty metric space* X. *Then there is a unique point* $x \in X$ *such that* $f(x) = x$.

Proof. Let $x_1 \in X$. Define $x_n \in X$ inductively by $x_{n+1} = f(x_n)$ $(n \in \mathbb{N})$. An easy induction argument establishes that

$$d(x_n, x_{n+1}) \le \alpha^{n-1} d(x_1, x_2) \qquad \forall n \in \mathbb{N}.$$

We then apply the extended triangle inequality to obtain for $p \leq q$

$$\begin{aligned} d(x_p, x_q) &\leq \sum_{n=p}^{q-1} d(x_n, x_{n+1}), \\ &\leq \sum_{n=p}^{q-1} \alpha^{n-1} d(x_1, x_2), \\ &\leq \alpha^{p-1}(1-\alpha)^{-1} d(x_1, x_2), \end{aligned}$$

since $0 \leq \alpha < 1$. It follows that (x_n) is a Cauchy sequence in X. Since X is complete, (x_n) converges to some element $x \in X$. Now

$$\begin{aligned} d(x, f(x)) &\leq d(x, x_n) + d(x_n, f(x_n)) + d(f(x_n), f(x)) \\ &\leq d(x, x_n) + \alpha^{n-1} d(x_1, x_2) + \alpha d(x_n, x) \end{aligned} \tag{4.18}$$

Since $0 \leq \alpha < 1$ and since $d(x, x_n) \to 0$ as $n \to \infty$, it follows that we can make the right hand side of (4.18) as small as we like, by taking n sufficiently large. It follows that $d(x, f(x)) = 0$ which can only occur if $f(x) = x$. This completes the existence part of the proof.

There is another way of seeing this last step of the proof that is worth mentioning. We know that $x_n \to x$ as $n \to \infty$. Since f is a Lipschitz mapping it is continuous, so $f(x_n) \to f(x)$. But $f(x_n) = x_{n+1}$ and $x_{n+1} \to x$. It follows from the uniqueness of the limit that $f(x) = x$.

It remains to check that the fixed point x is unique. Suppose that y also satisfies $f(y) = y$. Then we have

$$d(x, y) = d(f(x), f(y)) \leq \alpha d(x, y).$$

Since $0 \leq \alpha < 1$ the only way out is that $d(x, y) = 0$ which gives $x = y$. ■

Example Here we present an example of a mapping $f : X \to X$ of a complete space X such that

$$d_X(f(x_1), f(x_2)) < d_X(x_1, x_2) \qquad \text{for } x_1 \neq x_2 \tag{4.19}$$

but which does not have a fixed point. Let $X = \mathbb{R}$ and let

$$f(x) = x + \tfrac{1}{2}(1 - \tanh(x)).$$

Then, applying the Mean Value Theorem, we have

$$f(x_1) - f(x_2) = f'(\xi)(x_1 - x_2)$$

for ξ between x_1 and x_2. In any case,

$$\tfrac{1}{2} \leq f'(\xi) = 1 - \tfrac{1}{2}\operatorname{sech}^2 x < 1$$

so that (4.19) holds. Since $\tanh(x) < 1$ we see that f has no fixed point. □

4.4 Extension by Uniform Continuity

In this section we tackle extension by continuity as it is usually called. Actually as we shall see, this is a misnomer.

Lemma 4.10 *Let X and Y be metric spaces. Let $f : X \longrightarrow Y$ be a uniformly continuous mapping. Let (x_n) be a Cauchy sequence in X. Then $(f(x_n))$ is a Cauchy sequence in Y.*

Proof. Let $\epsilon > 0$. Then there exists $\delta > 0$ such that

$$a, b \in X, d_X(a, b) < \delta \qquad \Rightarrow \qquad d_Y(f(a), f(b)) < \epsilon. \tag{4.20}$$

Since (x_n) is a Cauchy sequence, there exists $N \in \mathbb{N}$ such that

$$p, q > N \qquad \Rightarrow \qquad d_X(x_p, x_q) < \delta.$$

Combining this with (4.20) yields

$$p, q > N \qquad \Rightarrow \qquad d_Y(f(x_p), f(x_q)) < \epsilon.$$

It follows that $(f(x_n)$ is a Cauchy sequence in Y. ■

Theorem 4.11 *Let X and Y be metric spaces and suppose that Y is complete. Let A be a dense subset of X. Let $f : A \longrightarrow Y$ be a uniformly continuous map. Then there is a unique uniformly continuous mapping $\tilde{f} : X \longrightarrow Y$ such that $\tilde{f}(a) = f(a)$ for all $a \in A$.*

Proof. Let $x \in X$. Since A is dense in X there exists a sequence (a_n) in A converging to x. Now, (a_n) is a Cauchy sequence in A and so, by Lemma 4.4, $(f(a_n))$ is a Cauchy sequence in Y. Since Y is complete, this converges to some element y of Y. We will define

$$\tilde{f}(x) = y.$$

We need to prove that $\tilde{f}$ is well-defined. Suppose that (b_n) is another sequence in A converging to x. Then $(f(b_n))$ must converge to some element z of Y. We need to show that $y = z$. To see this we mix the two sequences (a_n) and (b_n) by $x_{2n-1} = a_n$, $x_{2n} = b_n$. The sequence (x_n) converges to x and so $(f(x_n))$ is convergent in Y. But since $(f(x_n))$ has one subsequence converging to y and another converging to z it follows that $y = z$ as required.

If $x \in A$, then we can always take $a_n = x$ for all $n \in \mathbb{N}$. It follows that $\tilde{f}(x) = f(x)$ for all $x \in A$.

Next we show that $\tilde{f}$ is uniformly continuous. Let $\epsilon > 0$. The there exists $\delta > 0$ such that

$$a, b \in A, d(a, b) < \delta \qquad \Rightarrow \qquad d(f(a), f(b)) < \tfrac{1}{2}\epsilon. \tag{4.21}$$

Now, let x and x' be points of X such that $d(x, x') < \frac{1}{2}\delta$. Let us find sequences (a_n) and (a'_n) in A converging to x and x' respectively. Then, there exists $N \in \mathbb{N}$ such that

$$n > N \qquad \Rightarrow \qquad d(a_n, x) < \tfrac{1}{4}\delta \text{ and } d(a'_n, x') < \tfrac{1}{4}\delta. \tag{4.22}$$

An application of the extended triangle inequality, (4.21) and (4.22) now yields

$$n > N \qquad \Rightarrow \qquad d(f(a_n), f(a'_n)) < \tfrac{1}{2}\epsilon.$$

Letting now $n \longrightarrow \infty$ and using the fact that d_Y is continuous on $Y \times Y$ we see that

$$d(\tilde{f}(x), \tilde{f}(x')) \leq \tfrac{1}{2}\epsilon < \epsilon,$$

since the sequences $(f(a_n))$ and $(f(a'_n))$ converge to $\tilde{f}(x)$ and $\tilde{f}(x')$ respectively. This establishes the uniform continuity of $\tilde{f}$.

The final step of the proof is to show that $\tilde{f}$ is unique. Let g be another continuous extension of f. Since A is dense in X and $\tilde{f}$ and g are both continuous functions that agree on A, we can use Proposition 2.20 to deduce that $g = \tilde{f}$. ■

Example We show that in Theorem 4.11, one cannot replace uniform continuity by continuity. Let $X = [0,1]$, $A =]0,1]$ and $Y = [-1,1]$. Let

$$f(a) = \sin\frac{1}{a}$$

for all $a \in A$. Then all the hypotheses of Theorem 4.11 are met, except for the uniform continuity of f. The function f is of course continuous on A. We leave it as an exercise to show that f does not extend continuously to X. □

Example Let V be a normed vector space. Suppose that T is a continuous linear map from ℓ^1 to V. Define $v_n = T(e_n) \in V$ where e_n denotes the sequence in ℓ^1 that has a 1 in the n-th place and 0 everywhere else. We see that

$$\|v_n\| \le \|T\|_{\mathrm{op}}\|e_n\|_{\ell^1} = \|T\|_{\mathrm{op}}.$$

Thus

$$\sup_{n\in\mathbb{N}} \|v_n\| \le \|T\|_{\mathrm{op}}. \tag{4.23}$$

Conversely suppose that $C = \sup_{n\in\mathbb{N}}\|v_n\| < \infty$. Let F denote the set of all finitely supported sequences in ℓ^1. Then we can define a map

$$T_0 : F \to V$$

by $T_0(\sum t_n e_n) = \sum t_n v_n$. Here the sums involved are finite sums. It is easy to check that $\|T_0(t)\| \le C\|t\|$ for all $t \in F$ and it follows that T_0 is uniformly continuous on the dense subset F of ℓ^1. Thus by Theorem 4.11 T_0 can be extended to a uniformly continuous mapping T on the whole of ℓ^1. In this particular example this is no big deal, because it is also straightforward to define T directly. For $t \in \ell^1$, we can take

$$T(t) = \sum_{n=1}^{\infty} t_n v_n$$

an absolutely convergent infinite sum in the complete space V. □

Example The previous example features two possible approaches to defining an operator T, one direct and one involving extension. There do exist analogous situations where the direct approach is unavailable. For instance let $(f_j)_{j=1}^{\infty}$ be an orthonormal set in ℓ^2. Then, for $t = (t_j) \in F$, we can define $T_0(t) = T_0(\sum t_n e_n) = \sum t_n f_n$. Here we view T_0 as a mapping from F to ℓ^2. A simple calculation

$$\|\sum t_n f_n\|_2^2 = \sum_{m,n} t_m t_n \langle f_m, f_n\rangle = \sum_n t_n^2 \|f_n\|_2^2 = \sum_n t_n^2 = \|t\|_2^2$$

shows that T_0 is an isometry. It follows that T_0 extends to an isometry $T : \ell^2 \to \ell^2$. It is important to realise that for general $t \in \ell^2$ the sum

$$\sum_{n=1}^{\infty} t_n f_n$$

does not converge absolutely in ℓ^2. It does converge in ℓ^2 norm, but taking this route is essentially repeating the argument of Theorem 4.11. □

4.5 Completions

Definition *Let X be a metric space. Then a* ***completion*** *(Y, j) of X is a complete metric space Y, together with an isometric inclusion $j : X \to Y$ such that $j(X)$ is dense in Y.*

The completion of a metric space is unique in the following sense.

THEOREM 4.12 *Let (Y, j) and (Z, k) be completions of X. Then there is a surjective isometry $\alpha : Y \to Z$ such that $k = \alpha \circ j$.*

Proof. Let us define $\beta : j(X) \to Z$ by $\beta(j(x)) = k(x)$. Since j is an injective mapping from X onto $j(X)$, β is well defined. Also β is an isometry since j and k are. Now apply Theorem 4.11 to define $\alpha : Y \to Z$ a continuous map. Obviously we have $k = \alpha \circ j$.

It remains to show that α is an isometry. We consider two mappings from $Y \times Y$ to $\mathbb{R}^+$.

$$\begin{aligned} (y_1, y_2) &\mapsto d_Y(y_1, y_2) \\ (y_1, y_2) &\mapsto d_Z(\alpha(y_1), \alpha(y_2)) \end{aligned}$$

These mappings are continuous by Theorem 2.10 (page 17) and since the metric is itself continuous (page 21). Since j and k are isometries, they agree on the subset $j(X)\times j(X)$ of $Y \times Y$. By Proposition 2.21, $j(X)\times j(X)$ is dense in $Y \times Y$. Finally by Proposition 2.20 the two mappings agree everywhere on $Y \times Y$. This says that α is an isometry. ■

With the issue of uniqueness of completions out of the way, we deal with the more difficult question of existence.

THEOREM 4.13 *Every metric space possesses a completion.*

In the proof that we give below, we unashamedly use the completeness of $\mathbb{R}$. We take the point of view that an understanding of $\mathbb{R}$ is needed to define the metric space concept in the first place. There is another proof in the literature using equivalence classes of Cauchy sequences which avoids this issue.

Proof. We will assume that X is a bounded metric space and construct a completion. At the end of the proof we will discuss the modifications that are necessary to dispense with the boundedness hypothesis.

Let $j : X \to C(X, \mathbb{R})$ be given by

$$(j(x_1))(x_2) = d(x_1, x_2).$$

One needs to stand back a moment to ponder this notation. Since $x_1 \in X$, $j(x_1) \in C(X, \mathbb{R})$, that is, $j(x_1)$ is itself a mapping from X to $\mathbb{R}$. For $x_2 \in X$, the notation $(j(x_1))(x_2) \in \mathbb{R}$ then stands for the image of x_2 by the mapping $j(x_1)$.

It follows from the continuity of the metric (page 21) that $j(x_1)$ is continuous. Since X is a bounded metric space, $j(x_1)$ is bounded. Next we observe

$$\|j(x_1) - j(x_2)\| = \sup_{x_3 \in X} |d(x_1, x_3) - d(x_2, x_3)|.$$

Two applications of the triangle inequality show that

$$|d(x_1, x_3) - d(x_2, x_3)| \leq d(x_1, x_2) \qquad \forall x_3 \in X,$$

while taking $x_3 = x_2$ shows that

$$\sup_{x_3 \in X} |d(x_1, x_3) - d(x_2, x_3)| \geq d(x_1, x_2).$$

Thus j is an isometry. Since $C(X, \mathbb{R})$ is complete, the closed subset $\mathrm{cl}(j(X))$ is also complete by Proposition 4.4 (page 52). Let $Y = \mathrm{cl}(j(X))$ and consider j just as a mapping from X to Y. Then (Y, j) is a completion of X.

In the case that X is unbounded, select any point x_0 from X. Now set $(j(x_1))(x_2) = d(x_1, x_2) - d(x_0, x_2)$. The key observation is that since

$$|d(x_1, x_2) - d(x_0, x_2)| \leq d(x_0, x_1)$$

the function $j(x_1)$ is actually bounded. The rest of the proof follows the same line. ■

4.6 Extension of Continuous Functions

LEMMA 4.14 *Let X be a metric space. Let E_0 and E_1 be disjoint closed subsets of X. Then there exists a continuous function $f : X \longrightarrow [0,1]$ such that $f^{-1}(\{0\}) = E_0$ and $f^{-1}(\{1\}) = E_1$.*

Proof. We define
$$f(x) = \frac{\mathrm{dist}_{E_0}(x)}{\mathrm{dist}_{E_0}(x) + \mathrm{dist}_{E_1}(x)}.$$
We observe that $x \longmapsto \mathrm{dist}_{E_0}(x)$ and $x \longmapsto \mathrm{dist}_{E_1}(x)$ are both continuous functions on X. Furthermore $\mathrm{dist}_{E_0}(x) + \mathrm{dist}_{E_1}(x) > 0$ for all $x \in X$ by Proposition 2.23 (page 25) and since E_0 and E_1 are disjoint. It follows that $f : X \longrightarrow [0,1]$ is continuous. Clearly $f(x) = 0$ if and only if $\mathrm{dist}_{E_0}(x) = 0$ which occurs if and only if $x \in E_0$ again by Proposition 2.23. Similarly $f(x) = 1$ if and only if $x \in E_1$. ■

We do not need it right at the moment, but there is a simple looking Corollary of Lemma 4.14 that is difficult to establish without using the distance to a subset function.

COROLLARY 4.15 *Let X be a metric space. Let E_0 and E_1 be disjoint closed subsets of X. Then there exist U_0 and U_1 disjoint open subsets of X such that $E_j \subseteq U_j$ for $j = 0, 1$.*

Proof. Let f be the function of Lemma 4.14 and simply set $U_0 = f^{-1}([0, \frac{1}{3}[)$ and $U_1 = f^{-1}(]\frac{2}{3}, 1])$. Of course, U_0 is open in X since $[0, \frac{1}{3}[$ is (relatively) open in $[0,1]$. Similarly for U_1. The sets U_0 and U_1 obviously satisfy the remaining properties. ■

With this diversion out of the way, we can now continue to the main order of business.

THEOREM 4.16 (TIETZE EXTENSION THEOREM) *Let X be a metric space and let E be a closed subset of X. Let $g : E \longrightarrow [-1,1]$ be a continuous mapping. Then there exists a continuous mapping $f : X \longrightarrow [-1,1]$ extending g. Explicitly, this means that*
$$f(x) = g(x) \qquad \forall x \in E.$$

Proof. Let us denote $g_0 = g$. We start by constructing a continuous map f_0. Let $E_0 = g_0^{-1}([-1, -\frac{1}{3}])$ and let $E_1 = g_0^{-1}([\frac{1}{3}, 1])$. Since E is closed in X and E_0 and E_1 are closed in E, it follows that E_0 and E_1 are also closed in X. By a straightforward variant of Lemma 4.14, there is a mapping $f_0 : X \longrightarrow [-\frac{1}{3}, \frac{1}{3}]$ such that $f_0^{-1}(\{-\frac{1}{3}\}) = E_0$ and $f_0^{-1}(\{\frac{1}{3}\}) = E_1$. We claim that
$$|g_0(x) - f_0(x)| \le \tfrac{2}{3} \qquad \forall x \in E.$$
There are three cases.

- $x \in E_0$. Then $f_0(x) = -\frac{1}{3}$ and $g_0(x) \in [-1, -\frac{1}{3}]$.
- $x \in E_1$. Then $f_0(x) = \frac{1}{3}$ and $g_0(x) \in [\frac{1}{3}, 1]$.
- $x \in E \setminus (E_0 \cup E_1)$. Then $f_0(x), g_0(x) \in]-\frac{1}{3}, \frac{1}{3}[$.

Now define
$$g_1(x) = g_0(x) - f_0(x) \qquad \forall x \in E.$$
Then $\|g_1\|_\infty \le \frac{2}{3}$. We repeat the above argument at $\frac{2}{3}$ scale to define f_1. We then proceed inductively in the obvious way. We thus obtain continuous functions $f_n : X \longrightarrow [-2^n \cdot 3^{-(n+1)}, 2^n \cdot 3^{-(n+1)}]$ satisfying
$$\|g_n - f_n\|_\infty \le 2^n \cdot 3^{-n},$$

where $g_n = g_{n-1} - f_{n-1}|_E$.

It is easy to see that $\|g_n\|_\infty$ tends to 0 as $n \to \infty$. Hence we may write g as the telescoping sum

$$g = \sum_{n=1}^{\infty}(g_{n-1} - g_n) = \sum_{n=0}^{\infty} f_n|_E.$$

But the function f given by the uniformly convergent sum

$$f = \sum_{n=0}^{\infty} f_n,$$

also converges to a continuous function taking values in $[-1, 1]$ and evidently

$$f|_E = \sum_{n=0}^{\infty} f_n|_E = g,$$

as required. ■

4.7 Baire's Theorem

The following result has a number of key applications that cannot be approached in any other way.

THEOREM 4.17 (BAIRE'S CATEGORY THEOREM) *Let X be a complete metric space. Let A_k be a sequence of closed subsets of X with $\operatorname{int}(A_k) = \emptyset$. Then*

$$X \setminus \bigcup_{k=1}^{\infty} A_k \text{ is dense in } X. \tag{4.24}$$

In particular if X is nonempty we have

$$\bigcup_{k=1}^{\infty} A_k \neq X.$$

Proof. We suppose that (4.24) fails. Then there exist $x_0 \in X$ and $t > 0$ such that

$$U(x_0, t) \subseteq \bigcup_{k=1}^{\infty} A_k. \tag{4.25}$$

We construct a sequence (x_n) in X. Let $t_0 = \frac{1}{2}t$. We choose $x_1 \in X \setminus A_1$ such that $d(x_1, x_0) < t_0$. This is possible since otherwise we would have $U(x_0, t_0) \subseteq A_1$ contradicting the fact that $\operatorname{int}(A_1) = \emptyset$. Now define $t_1 = \min(\frac{1}{2}t_0, \frac{1}{4}\operatorname{dist}_{A_1}(x_1)) > 0$. Next find $x_2 \in X \setminus A_2$ such that $d(x_2, x_1) < t_1$. If this is not possible then $U(x_1, t_1) \subseteq A_2$ contradicting the hypothesis $\operatorname{int}(A_2) = \emptyset$. Now define $t_2 = \min(\frac{1}{2}t_1, \frac{1}{4}\operatorname{dist}_{A_2}(x_2)) > 0$ and then find $x_3 \in X \setminus A_3$ such that $d(x_3, x_2) < t_2$. Continuing in this way, we obtain a sequence (x_n) in X and a sequence (t_n) of strictly positive reals such that

$$\begin{aligned} t_n &= \min(\tfrac{1}{2}t_{n-1}, \tfrac{1}{4}\operatorname{dist}_{A_n}(x_n)) > 0 \qquad n = 1, 2, \ldots \\ x_n &\notin A_n \qquad n = 1, 2, \ldots \\ d(x_{n+1}, x_n) &< t_n \qquad n = 0, 1, 2, \ldots \end{aligned}$$

Since $d(x_{n+1}, x_n) < t_n \leq 2^{-n}t_0 = 2^{-n-1}t$, we see that (x_n) is a Cauchy sequence. The detailed justification of this is similar to one found in the proof of Theorem 4.9. It follows from the completeness of X that (x_n) converges to some limit point $x \in X$.

We next show that $x \notin A_k$ for all $k \in \mathbb{N}$. Indeed, by passing to the limit in the extended triangle inequality, we obtain for each $k = 1, 2, \ldots$

$$d(x, x_k) \le \sum_{n=k}^{\infty} d(x_{n+1}, x_n) < \sum_{n=k}^{\infty} t_n \le \sum_{n=k}^{\infty} 2^{-2-n+k} \mathrm{dist}_{A_k}(x_k) = \tfrac{1}{2}\mathrm{dist}_{A_k}(x_k).$$

It follows that $x \notin A_k$, as required.

Finally, following the same line, we find

$$d(x, x_0) \le \sum_{n=0}^{\infty} d(x_{n+1}, x_n) < \sum_{n=0}^{\infty} t_n \le \sum_{n=0}^{\infty} 2^{-1-n} t = t,$$

so that $x \in U(x_0, t)$. This contradicts (4.25) and completes the proof of the Theorem. ∎

4.8 Complete Normed Spaces

Complete normed spaces are also called ***Banach Spaces***. The following result is very basic and could have been left as an exercise for the reader.

THEOREM 4.18 *Let V and W be normed vector spaces and suppose that W is complete. Then $\mathcal{CL}(V, W)$ is a complete normed vector space with the operator norm.*

Proof. Let (T_n) be a cauchy sequence in $\mathcal{CL}(V, W)$. First consider a fixed point v of V. Then we have

$$\|T_p(v) \Leftrightarrow T_q(v)\| \le \|T_p \Leftrightarrow T_q\|_{\mathrm{op}} \|v\|.$$

It follows easily that $(T_n(v))$ is a Cauchy sequence in W. Since W is complete, there is an element $w \in W$ such that $T_n(v) \Leftrightarrow\rightarrow w$ as $n \Leftrightarrow\rightarrow \infty$. We now allow v to vary and define a mapping T by $T(v) = w$. We leave the reader to check that the mapping T is a *linear* mapping from V to W. Now let $\epsilon > 0$ then by the Cauchy condition there exist $N \in \mathbb{N}$ such that

$$p, q > N \qquad \Rightarrow \qquad \|T_p \Leftrightarrow T_q\|_{\mathrm{op}} \le \epsilon$$

or equivalently that

$$p, q > N, v \in V \qquad \Rightarrow \qquad \|T_p(v) \Leftrightarrow T_q(v)\| \le \epsilon \|v\|. \tag{4.26}$$

Now let q tend to infinity in (4.26). We find that

$$p > N, v \in V \qquad \Rightarrow \qquad \|T_p(v) \Leftrightarrow T(v)\| \le \epsilon \|v\|,$$

or equivalently that (T_p) converges to T in $\mathcal{CL}(V, W)$. This also shows that $T \in \mathcal{CL}(V, W)$. ∎

Rather more interesting is the following Proposition.

PROPOSITION 4.19 *Let V be a complete normed vector space and N a closed linear subspace. Then the quotient space $Q = V/N$ is a complete normed space with the norm defined by* (3.38).

Proof. This Proposition is included because it illustrates a method of proof not seen elsewhere in these notes. The key idea is the use of **rapidly convergent subsequences**. We denote by π the canonical projection mapping $\pi : V \Leftrightarrow\rightarrow Q$.

Let (q_n) be a Cauchy sequence in Q. Applying the Cauchy condition with $\epsilon = 2^{-k}$, we find n_k such that

$$\ell, m \ge n_k \quad \Rightarrow \quad \|q_\ell \Leftrightarrow q_m\| < 2^{-k}. \tag{4.27}$$

In particular, taking $l = n_k$, $m = n_{k+1}$, we have

$$\|q_{n_k} \Leftrightarrow q_{n_{k+1}}\| < 2^{-k}. \tag{4.28}$$

We now proceed to find lifts of the q_{n_k}. Let v_1 be any element of V with $\pi(v_1) = q_{n_1}$. Now by (4.28) and the definition of the quotient norm, there exists $u_{k+1} \in V$ such that $\|u_{k+1}\| < 2^{-k}$ and $\pi(u_{k+1}) = q_{n_{k+1}} \Leftrightarrow q_{n_k}$. We now define $v_2 = v_1 + u_2$, $v_3 = v_2 + u_3$, etc., so that we now have

$$\|v_k \Leftrightarrow v_{k+1}\| < 2^{-k} \tag{4.29}$$

and $\pi(v_k) = q_{n_k}$ for $k \in \mathbb{N}$. It is easy see to that (4.29) forces (v_k) to be a Cauchy sequence in V as in the proof of the Contraction Mapping Theorem. Since V is complete, we can infer the existence of $v \in V$ such that (v_k) converges to v.

Since π is continuous, it follows that the subsequence (q_{n_k}) converges to the element $\pi(v)$ of Q. Furthermore, we can have the estimate

$$\|q_{n_k} \Leftrightarrow \pi(v)\| \le 2^{-(k-1)}. \tag{4.30}$$

Combining this with (4.27) we find

$$\ell \ge n_k \quad \Rightarrow \quad \|q_\ell \Leftrightarrow \pi(v)\| < 3 \cdot 2^{-k},$$

from which it follows that the original sequence (q_n) converges to $\pi(v)$. ■

THEOREM 4.20 (OPEN MAPPING THEOREM) *Let U and V be complete normed spaces and let $T : U \Leftrightarrow\!\rightarrow V$ be a continuous surjective linear map. Then there is a constant $\epsilon > 0$ such that for every $v \in V$ with $\|v\| \le 1$, there exist $u \in U$ with $\|u\| \le \epsilon$ such that $T(u) = v$.*

The reason for the terminology is that the statement that T is an open mapping (see page 94 for the definition) is equivalent to the conclusion of the Theorem.

Proof. There are two separate ideas in the proof. The first is to use the Baire Category Theorem and the second involves iteration.

Let B_n denote $\{u : u \in U, \|u\| \le n\}$, the closed n-ball in U. Then, since T is onto, we have

$$V = \bigcup_{n \in \mathbb{N}} T(B_n).$$

We can't use this directly in the Baire Category Theorem because we don't know that the $T(B_n)$ are closed. We take the easiest way around this difficulty and write simply

$$V = \bigcup_{n \in \mathbb{N}} \mathrm{cl}(T(B_n)).$$

By the Baire Category Theorem (page 60), there exists $n \in \mathbb{N}$ such that $\mathrm{cl}(T(B_n))$ has nonempty interior. This means that there exists $v \in V$ and $t > 0$ such that $U_V(v, t) \subseteq \mathrm{cl}(T(B_n))$. By symmetry, it follows that $U_V(\Leftrightarrow v, t) \subseteq \mathrm{cl}(T(B_n))$. We claim that $U_V(0_V, t) \subseteq \mathrm{cl}(T(B_n))$. Let $w \in U_V(0_V, t)$. Then, we can find two sequences (x_k) and (y_k) in B_n such that $(T(x_k))$ converges to $w + v$ and $(T(y_k))$ converges to $w \Leftrightarrow v$. It follows that the sequence $(T(\frac{1}{2}(x_k + y_k)))$ converges w. This establishes the claim.

Now, let v be a generic element of V with $\|v\| < t$. Then $v \in \mathrm{cl}(T(B_n))$. Hence, there exists $u_0 \in B_n$ such that $\|v \Leftrightarrow T(u_0)\| < \frac{1}{2}t$. We repeat the argument, but rescaled by a factor of $\frac{1}{2}$ and applied to $v \Leftrightarrow T(u_0)$. Thus, there is an element $u_1 \in U$ with $\|u_1\| < \frac{1}{2}n$ and such that $\|v \Leftrightarrow T(u_0) \Leftrightarrow T(u_1)\| < \frac{1}{4}t$. Continuing in this way leads to elements $u_k \in U$ with $\|u_k\| < n\, 2^{-k}$ such that

$$\|v \Leftrightarrow \sum_{k=0}^{\ell} T(u_k)\| < t\, 2^{-\ell-1}.$$

Using now the fact that U is complete (the completeness of V is needed to apply Baire's Theorem), we find that $T(u) = v$ where

$$u = \sum_{k=0}^{\infty} u_k \in U$$

is given by an absolutely convergent series and has norm bounded by $2n$. Rescaling gives the required result. ■

Corollary 4.21 *Let V be a vector space with two norms $\|\ \|_1$ and $\|\ \|_2$, both of which make V complete. Suppose that there is a constant C such that*

$$\|v\|_2 \le C\|v\|_1 \qquad \forall v \in V.$$

Then $\|\ \|_1$ and $\|\ \|_2$ are equivalent norms.

Proof. Apply the Open Mapping Theorem in case that T is the identity mapping from $(V, \|\ \|_1)$ to $(V, \|\ \|_2)$. ■

It is possible to construct an infinite dimensional vector space with two *incomparable* norms both of which render it complete.

Proposition 4.22 *Let V_1 and V_2 be vector spaces and suppose that $\|\ \|$ is a norm on $V = V_1 \oplus V_2$ which renders V complete. Let P_1 and P_2 be the linear projection operators corresponding to the direct sum. Then the following are equivalent:*

- *P_1 and P_2 are continuous.*
- *V_1 and V_2 are closed in V.*

Proof. Suppose that P_1 is continuous and that (v_n) is a sequence in V_1 which converges to some element v of V. Then $(P_1(v_n))$ converges to $P_1(v)$. But, since $P_1(v_n) = v_n$ and by the uniqueness of limits (Proposition 2.3), $v = P_1(v)$ or equivalently $v \in V_1$. This shows that V_1 is closed in V. Of course since $P_2 = I - P_1$, if P_1 is continuous, so is P_2 and similarly we find that V_2 is closed.

The converse is much harder. We can identify V_2 with the quotient space V/V_1. Since V_1 is closed, we have a natural quotient norm

$$\|v_2\|_Q = \inf_{v_1 \in V_1} \|v_1 + v_2\|$$

on V_2 as well as the restriction of the given norm. We see that V_2 is complete in both norms, by Proposition 4.19 and since V_2 is closed in V. Clearly, $\|v_2\|_Q \le \|v_2\|$ for all $v_2 \in V_2$ so that the two norms are comparable. Hence, by Corollary 4.21, we see that there exists a finite constant C such that

$$\|v_2\| \le C \inf_{v_1 \in V_1} \|v_1 + v_2\| \qquad \forall v_2 \in V_2. \tag{4.31}$$

A moment's thought shows that (4.31) is equivalent to

$$\|P_2(v)\| \le C\|v\| \qquad \forall v \in V,$$

so that P_2 is continuous as required. One shows that P_1 is continuous by a similar argument, or by applying the formula $P_1 = I - P_2$. ■

5

Compactness

Compactness is one of the most important concepts in mathematical analysis. It is a topological form of finiteness. The formal definition is quite involved.

DEFINITION *A metric space X is said to be* **compact** *if, whenever V_α are open sets for every α in some index set I such that $\cup_{\alpha \in I} V_\alpha = X$, then there exists a finite subset $F \subseteq I$ such that $\cup_{\alpha \in F} V_\alpha = X$.*

PROPOSITION 5.1 *Every compact metric space is bounded.*

Proof. Let X be a compact metric space. If X is empty, then it is bounded. If not, select a point $x_0 \in X$. Now we have

$$X = \bigcup_{n \in \mathbb{N}} U(x_0, n).$$

By compactness, there is a finite subset $F \subseteq \mathbb{N}$ such that

$$X = \bigcup_{n \in F} U(x_0, n).$$

Let N be the largest integer in F. Such an integer exists because F is finite and non-empty. Then $X = U(x_0, N)$ and it follows that X is bounded. ■

PROPOSITION 5.2 *Every finite metric space is compact.*

Proof. Let V_α be open subsets of X for every α in some index set I such that $\cup_{\alpha \in I} V_\alpha = X$. Since X is finite, let us enumerate it as $X = \{x_1, x_2, \ldots, x_n\}$. For each $j = 1, \ldots, n$ there exists $\alpha_j \in I$ such that $x_j \in V_{\alpha_j}$. Let us set $F = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Then clearly $\cup_{\alpha \in F} V_\alpha = X$. ■

5.1 Compact Subsets

We also use the term compact to describe subsets of a metric space.

DEFINITION *A subset K of a metric space X is* **compact** *iff it is compact when viewed as a metric space in the restriction metric.*

It follows immediately from the definition and Proposition 5.1 that compact subsets are necessarily bounded.

We need a direct way to describe which subsets are compact.

PROPOSITION 5.3 *Let X be a metric space and let $K \subseteq X$. Then the following two conditions are equivalent.*

- *K is a compact subset of X.*
- *Whenever V_α are open sets of X for every α in some index set I such that $\cup_{\alpha \in I} V_\alpha \supseteq K$, then there exists a finite subset $F \subseteq I$ such that $\cup_{\alpha \in F} V_\alpha \supseteq K$.*

Proof. Suppose that the first statement holds. Let us assume that I is an index set which labels open subsets V_α of X such that $\cup_{\alpha \in I} V_\alpha \supseteq K$. Then, by Theorem 2.27 (page 27) the subsets $K \cap V_\alpha$ are open subsets of K. We clearly have $\cup_{\alpha \in I}(K \cap V_\alpha) = K$. Since K is compact, by the definition, there exists a finite subset $F \subseteq I$ such that $\cup_{\alpha \in F}(K \cap V_\alpha) = K$. It follows immediately that $\cup_{\alpha \in F} V_\alpha \supseteq K$.

For the converse, we suppose that the second condition holds. We need to show that K is compact. Let U_α be open subsets of K for every $\alpha \in I$ such that $\cup_{\alpha \in I} U_\alpha = K$. The, according to Theorem 2.27 there exist open subsets V_α of X such that $K \cap V_\alpha = U_\alpha$. We clearly have $\cup_{\alpha \in I} V_\alpha \supseteq K$ so that we may apply the second condition to infer the existence of F finite $F \subseteq I$ such that $\cup_{\alpha \in F} V_\alpha \supseteq K$. But it is easy to verify that

$$K \cap \left(\bigcup_{\alpha \in F} V_\alpha \right) = \bigcup_{\alpha \in F} (K \cap V_\alpha) = \bigcup_{\alpha \in F} U_\alpha,$$

and it follows that $\cup_{\alpha \in F} U_\alpha = K$ as required. We have just verified the compactness of K as a metric space. ■

PROPOSITION 5.4 *Compact subsets are necessarily closed.*

Proof. Let K be a compact subset of a metric space X. Let us suppose that K is not closed. We will provide a contradiction. Let $x \in \text{cl}(K) \setminus K$. Then $x \notin \text{int}(X \setminus K)$. Thus for every $\epsilon > 0$,

$$U(x, \epsilon) \cap K \neq \emptyset \tag{5.1}$$

holds. On the other hand, since $x \notin K$ we have

$$K \subseteq \bigcup_{n \in \mathbb{N}} \left(X \setminus B(x, \tfrac{1}{n})\right),$$

since for every $y \in K$, $d(x, y) > 0$. Since $B(x, t)$ is closed, $X \setminus B(x, t)$ is open. We therefore apply the compactness criterion for subsets to find F a finite subset of $\mathbb{N}$, such that

$$K \subseteq \bigcup_{n \in F} \left(X \setminus B(x, \tfrac{1}{n})\right).$$

If K is empty then K is closed and we are done. If not, F is a non-empty finite subset of $\mathbb{N}$ which therefore possesses a maximal element N. It follows that K is disjoint from $B(x, \frac{1}{N})$ contradicting (5.1). ■

THEOREM 5.5 *Every closed subset of a compact metric space is compact.*

Proof. Let X be a compact metric space. Let K be a closed subset of X. Let V_α be open sets of X for every α in some index set I such that $\cup_{\alpha \in I} V_\alpha \supseteq K$. We will show the existence of a finite subset F of I such that $\cup_{\alpha \in F} V_\alpha \supseteq K$.

To do this, we extend the index set by one index. Let $J = I \cup \{\beta\}$. Define $V_\beta = X \setminus K$ an open subset of X since K is closed and by Theorem 2.5 (page 13). We have

$$\begin{aligned}\bigcup_{\alpha \in J} V_\alpha &= \left(\bigcup_{\alpha \in I} V_\alpha\right) \cup (X \setminus K) \\ &\supseteq K \cup (X \setminus K) = X.\end{aligned}$$

By the compactness of X there is a finite subset G of J such that $\cup_{\alpha \in G} V_\alpha = X$. We can assume without loss of generality that $\beta \in G$ and define $F = G \cap I$. It is then clear that F is a finite subset of I and that

$$\bigcup_{\alpha \in F} V_\alpha \supseteq K.$$ ■

5.2 The Finite Intersection Property

DEFINITION *Let X be a set and let C_α be a subset of X for every α in some index set I. Then we say that the family $(C_\alpha)_{\alpha \in I}$ has the* ***finite intersection property*** *iff*

$$\bigcap_{\alpha \in F} C_\alpha \neq \emptyset$$

for every finite subset F of I.

The following Proposition is an immediate consequence of the definition of compactness and Theorem 2.5 (page 13).

PROPOSITION 5.6 *Let X be a metric space. Then the following two statements are equivalent.*

- *X is compact.*
- *Whenever $(C_\alpha)_{\alpha \in I}$ is a family of closed subsets of X having the finite intersection property, then $\cap_{\alpha \in I} C_\alpha \neq \emptyset$.*

This reformulation of compactness is often very useful.

5.3 Other Formulations of Compactness

In this section we look at some conditions which are equivalent to or very nearly equivalent to compactness. The first of these is countable compactness. Countable compactness is a technical device and is never used in practice.

DEFINITION *A metric space X is said to be* ***countably compact*** *if, whenever V_n are open sets for every $n \in \mathbb{N}$ such that $\cup_{n \in \mathbb{N}} V_n = X$, then there exists a finite subset $F \subseteq \mathbb{N}$ such that $\cup_{n \in F} V_n = X$.*

We say that a subset K of a metric space is countably compact if it is countably compact when viewed as a metric space in its own right. There is a formulation of the concept of countably compact subset entirely analogous to that given for compact subset in Proposition 5.3 (page 65).

Clearly, a metric space that is compact is also countably compact. The converse is true in the context of metric spaces, but false in the setting of topological spaces. Towards the converse, we have the following Proposition.

Proposition 5.7 *Let X be a separable, countably compact metric space. Then X is compact.*

Proof. Let V_α be open sets of X for every α in some index set I satisfying $\cup_{\alpha\in I}V_\alpha = X$. We aim to find a finite subset F of I such that $\cup_{\alpha\in F}V_\alpha = X$. Observe that if for some $\alpha \in I$, we have $V_\alpha = X$, then we are done. Thus we may assume that for each $\alpha \in I$, the set $X \setminus V_\alpha$ is non-empty.

Let S be a countable dense subset of X. We observe that $S \cap V_\alpha$ is dense in V_α. We now use the proof of Theorem 2.26 (page 26), which shows that

$$V_\alpha = \bigcup_{s\in S\cap V_\alpha} U(s, \tfrac{1}{2}\mathrm{dist}_{X\setminus V_\alpha}(s)). \tag{5.2}$$

Now, let us define the subset Q_α of $S \times \mathbb{Q}$ by

$$Q_\alpha = \{(s,t); s\in S, t\in\mathbb{Q}, t>0, U(s,t)\subseteq V_\alpha\} \tag{5.3}$$

Then, by (5.2) we have

$$V_\alpha = \bigcup_{(s,t)\in Q_\alpha} U(s,t). \tag{5.4}$$

We also define $Q = \cup_{\alpha\in I}Q_\alpha$. Since Q is a subset of $S\times\mathbb{Q}$ it is necessarily countable. Then we have by (5.4)

$$X = \bigcup_{(s,t)\in Q} U(s,t). \tag{5.5}$$

The key idea of the proof is to replace the union $\cup_{\alpha\in I}V_\alpha$ with the countable union $\cup_{(s,t)\in Q}U(s,t)$. We now apply the countable compactness of X to (5.5). We obtain a finite subset $R \subset Q$ such that

$$X = \bigcup_{(s,t)\in R} U(s,t). \tag{5.6}$$

For each $(s,t)\in R$ there exists $\alpha\in I$ such that $(s,t)\in Q_\alpha$. Let F be the finite subset of I having these α as members. Then by (5.3) we have

$$\bigcup_{(s,t)\in R} U(s,t) \subseteq \bigcup_{\alpha\in F} V_\alpha. \tag{5.7}$$

Finally, combining (5.6) and (5.7) gives the desired conclusion. ■

While countable compactness is merely a means to an end, sequential compactness is a very useful tool. It is equivalent to compactness in metric spaces (but not in topological spaces) and can be used as a replacement for compactness in nearly all situations.

Definition *Let X be a metric space. Then X is* ***sequentially compact*** *iff every sequence (x_n) in X possesses a convergent subsequence. A subset of a metric space is sequentially compact iff it is a sequentially compact metric space in the restriction metric.*

Proposition 5.8 *Every compact metric space is sequentially compact.*

Proof. Let X be a compact metric space and let (x_n) be a sequence of points of X. Let $T_m = \{x_n; n \ge m\}$ for $m\in\mathbb{N}$. Then the closed sets $\mathrm{cl}(T_m)$ clearly have the finite intersection property. Hence $\cap_{m\in\mathbb{N}}\mathrm{cl}(T_m) \neq \emptyset$. Let x be a member of this set. We construct a subsequence of (x_n) that converges to x. Let (ϵ_n) be a sequence of strictly positive real numbers decreasing to zero. Then we define the natural subsequence (n_k) inductively. Since $x \in \mathrm{cl}(T_1)$, we choose $n_1 \in \mathbb{N}$ such that $d(x, x_{n_1}) < \epsilon_1$. Now assuming that n_k is already defined, we use the fact that $x \in \mathrm{cl}(T_{n_k+1})$ to find $n_{k+1} > n_k$ and such that $d(x, x_{n_{k+1}}) < \epsilon_{k+1}$. It is easy to see that the subsequence (x_{n_k}) converges to x. Since (x_n) was an arbitrary sequence of points of X it follows that X is sequentially compact. ∎

LEMMA 5.9 *Let (x_n) be a sequence in the cell $[a, b]^d$ in $\mathbb{R}^d$. Then (x_n) possesses a subsequence which converges to some point of $\mathbb{R}^d$.*

Proof. We define the natural subsequence (n_k) inductively. Let $n_1 = 1$. Let $C_1 = [a, b]^d$ the original cell. Let $c = \frac{1}{2}(a + b)$. The we divide up $[a, b]$ as $[a, c] \cup [c, b]$. Taking the n-th product, this divides the cell C_1 up into 2^d cells with half the linear size of C_1. We select one of these cells C_2 with the property that the set

$$R_2 = \{n; n \in \mathbb{N}, n > n_1, x_n \in C_2\}$$

is infinite. It cannot happen that all of the 2^d cells fail to have this property, for then the set $\{n; n \in \mathbb{N}, n > n_1\}$ would be finite. We choose $n_2 \in R_2$.

To understand the general step of the inductive process, suppose that C_k, R_k and $n_k \in R_k$ have been chosen. Then as before we divide C_k into 2^d cells of half the linear size. We select one of these cells C_{k+1} with the property that

$$R_{k+1} = \{n; n \in R_k, n > n_k, x_n \in C_{k+1}\}$$

is infinite. We choose $n_{k+1} \in R_{k+1}$.

Let $\epsilon > 0$. Then there exists $K \in \mathbb{N}$ such that $\mathrm{diam}(C_K) < \epsilon$. It follows that

$$\begin{aligned} p, q \geq K \qquad &\Rightarrow \qquad n_p, n_q \in R_K, \\ &\Rightarrow \qquad d(x_{n_p}, x_{n_q}) < \epsilon. \end{aligned}$$

In words, this just says that the subsequence (x_{n_k}) is a Cauchy sequence. But since $\mathbb{R}^d$ is complete, it will converge to some point of $\mathbb{R}^d$. ∎

THEOREM 5.10 (BOLZANO–WEIERSTRASS THEOREM) *Every closed bounded subset of $\mathbb{R}^d$ is sequentially compact.*

Proof. Let K be a closed bounded subset of $\mathbb{R}^d$ and suppose that (x_n) is a sequence of points of K. Since K is bounded it is contained in some cell $[a, b]^d$ of $\mathbb{R}^d$. Then according to Lemma 5.9, (x_n) possesses a subsequence (x_{n_k}) convergent to some point x of $\mathbb{R}^d$. But since K is closed, it follows that $x \in K$. This shows that K is sequentially compact. ∎

THEOREM 5.11 (HEINE–BOREL THEOREM) *A subset K of $\mathbb{R}^d$ is compact iff it is closed and bounded.*

Proof. We have already seen that a compact subset of a metric space is necessarily closed and bounded. It therefore remains only to show that if K is closed and bounded then it is compact. Since $\mathbb{R}^d$ is separable, it follows from Theorem 2.24 (page 25) that K is also. Hence by Proposition 5.7 it is enough to show that K is countably compact. We will establish the condition for a subset to be countably compact analogous to that given by Proposition 5.3 (page 65). Thus, let V_k be open subsets of $\mathbb{R}^d$ for $k \in \mathbb{N}$. Suppose that $\cup_{k \in \mathbb{N}} V_k \supseteq K$. We will show that there exists a finite subset F of $\mathbb{N}$ such that

$$\bigcup_{k \in F} V_k \supseteq K. \tag{5.8}$$

Suppose not. Then, for every k, the set $F = \{1, 2, \ldots, k\}$ fails to satisfy (5.8). So, there is a point $x_k \in K$ with

$$x_k \notin \bigcup_{\ell=1}^{k} V_\ell. \tag{5.9}$$

Now by the Bolzano–Weierstrass Theorem, the sequence (x_k) has a subsequence which converges to some element $x \in K$. By hypothesis, there exists $m \in \mathbb{N}$ such that $x \in V_m$. Since V_m is open, some tail of the subsequence lies entirely inside V_m. This follows from Proposition 2.4 (page 12). Therefore, there exists $k \in \mathbb{N}$ with $k \geq m$ and $x_k \in V_m$. This contradiction with (5.9) establishes the result. ■

COROLLARY 5.12 *Every closed bounded cell*

$$\prod_{j=1}^{d} [a_j, b_j]$$

in $\mathbb{R}^d$ *is compact.*

EXAMPLE The star space X based on an *infinite* set S provides an example of a complete bounded metric space that is *not* compact. See page 40 for the definition of the star space and page 53 for the proof of completeness. Let (s_n) be a sequence of distinct elements of S. Then define a sequence (x_n) of X by $x_n = 0\langle c\rangle + 1\langle s_n\rangle$ where c denotes the centre of X. Then it is immediate that

$$d_X(x_m, x_n) = \begin{cases} 2 & \text{if } m \neq n \\ 0 & \text{if } m = n \end{cases}$$

It follows that (x_n) possesses no convergent subsequence. Thus X is not sequentially compact. □

EXAMPLE Another example of a bounded complete space that is not compact is the unit ball of ℓ^1. The sequence of coordinate vectors (e_n) does not possess a convergent subsequence. □

EXAMPLE The orthogonal groups provide examples of interesting compact spaces. An $n \times n$ matrix U is said to be ***orthogonal*** iff $U'U = I$. Here we have denoted U' the transpose of U. The set of all orthogonal $n \times n$ matrices is usually denoted $O(n)$. It is well known to be a group under matrix multiplication. We view $O(n)$ as a subset of the vector space $M(n, n, \mathbb{R})$ of all $n \times n$ matrices which is a n^2 dimensional real vector space. The equations $U'U = I$ can be rewritten as

$$\sum_{\ell=1}^{n} u_{\ell j} u_{\ell k} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \tag{5.10}$$

showing that $O(n)$ is a closed subset of $M(n, n, \mathbb{R})$. The first case in (5.10) can again be rewritten as

$$\sum_{\ell=1}^{n} u_{\ell j}^2 = 1 \qquad j = 1, 2, \ldots, n$$

showing that $|u_{\ell j}| \leq 1$ for all $j = 1, 2, \ldots, n$ and all $\ell = 1, 2, \ldots, n$. Thus $O(n)$ is a bounded subset of $M(n, n, \mathbb{R})$. The Heine–Borel Theorem can now be used to conclude that $O(n)$ is compact. □

5.4 Preservation of Compactness by Continuous Mappings

One of the most important properties of compactness is that it is preserved by continuous mappings.

THEOREM 5.13 *Let X and Y be metric spaces. Suppose that X is compact. Let $f : X \longrightarrow Y$ be a continuous surjection. Then Y is also compact.*

Proof. We work directly from the definition. Let V_α be open sets of Y for every α in some index set I such that $\cup_{\alpha \in I} V_\alpha = Y$. Then, by Theorem 2.8 (page 16), $f^{-1}(V_\alpha)$ are open subsets of X. We have

$$X = \bigcup_{\alpha \in I} f^{-1}(V_\alpha).$$

We now apply the compactness of X to deduce the existence of a finite subset F of I such that

$$X = \bigcup_{\alpha \in F} f^{-1}(V_\alpha).$$

We wish to deduce that

$$Y = \bigcup_{\alpha \in F} V_\alpha. \tag{5.11}$$

Let $y \in Y$, then since f is surjective, there exists $x \in X$ such that $f(x) = y$. There exists $\alpha \in F$ such that $x \in f^{-1}(V_\alpha)$. It follows that $y = f(x) \in V_\alpha$. This verifies (5.11). ■

There is a formulation of this result in terms of compact subsets which is probably used more frequently.

COROLLARY 5.14 *Let X and Y be metric spaces. Let $f : X \longrightarrow Y$ be a continuous mapping. Let K be a compact subset of X. Then $f(K)$ is a compact subset of Y.*

Proof. By definition, K is a compact metric space in its own right. Since $f|_K$ can be regarded as a continuous mapping from K onto $f(K)$, it follows that $f(K)$ is compact, when viewed as a metric space with the metric obtained by restriction from Y. Hence by Theorem 5.13, $f(K)$ is a compact subset of Y. ■

Theorem 5.13 has important consequences because of the application to real-valued functions. We need the following result.

PROPOSITION 5.15 *The supremum of every non-empty compact subset K of $\mathbb{R}$ belongs to K.*

Of course, the same result applies to the infimum.

Proof. Let K be a compact non-empty subset of $\mathbb{R}$. Since K is bounded and non-empty, it possesses a supremum x. For every $n \in \mathbb{N}$, the number $x - \frac{1}{n}$ is not an upper bound for K. Thus there exists $x_n \in K$ and $x_n > x - \frac{1}{n}$. On the other hand, x is an upper bound for K so that $x_n \le x$. It follows that $|x - x_n| < \frac{1}{n}$ so that (x_n) converges to x. Since $x_n \in K$ and K is closed, it follows that $x \in K$. ■

THEOREM 5.16 *Let X be a non-empty compact metric space and let $f : X \longrightarrow \mathbb{R}$ be continuous. Then f attains its maximum value.*

Proof. By Theorem 5.13, $f(X)$ is a non-empty compact subset of $\mathbb{R}$ which therefore contains its supremum. Hence, there exists $x_0 \in X$ such that

$$f(x_0) = \sup_{x \in X} f(x),$$

as required. ■

One of the most significant applications of this result involves norms on finite-dimensional spaces.

COROLLARY 5.17 *Let V be a finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$. Then any two norms on V are equivalent.*

Proof. We give the proof for a finite-dimensional real vector space. The complex case is similar. Let us select a basis $(e_1, e_2, \ldots, e_n)$ of V. We define a norm $\|\ \|_1$ on V by

$$\|\sum_{j=1}^{n} t_j e_j\|_1 = \sum_{j=1}^{n} |t_j|.$$

Then for any other norm $\|\ \|_V$ on V it will be shown that $\|\ \|_1$ and $\|\ \|_V$ are equivalent. We have

$$\|\sum_{j=1}^{n} t_j e_j\|_V \leq \sum_{j=1}^{n} |t_j| \|e_j\|_V \leq C \sum_{j=1}^{n} |t_j| = C \|\sum_{j=1}^{n} t_j e_j\|_1, \tag{5.12}$$

where

$$C = \max_{j=1}^{n} \|e_j\|_V.$$

For the converse inequality we will need to use the Heine–Borel Theorem which was proved with respect to the infinity norm

$$\|\sum_{j=1}^{n} t_j e_j\|_\infty = \max_{j=1}^{n} |t_j|.$$

This is not a problem because

$$\max_{j=1}^{n} |t_j| \leq \sum_{j=1}^{n} |t_j| \leq n \max_{j=1}^{n} |t_j|.$$

so that $\|\ \|_1$ and $\|\ \|_\infty$ are equivalent. It follows that the unit sphere S for the norm $\|\ \|_1$ is compact for the metric of the $\|\ \|_1$ norm. Explicitly we have

$$S = \{\sum_{j=1}^{n} t_j e_j ; \sum_{j=1}^{n} |t_j| = 1\}.$$

By (5.12), $v \longmapsto \|v\|_V$ is continuous as a map

$$(V, \|\ \|_1) \quad \longrightarrow \quad \mathbb{R}.$$

It follows that this function attains its minimum value on S. Thus, if we let

$$c = \inf_{v \in S} \|v\|_V, \tag{5.13}$$

there actually exists $u \in S$ such that $\|u\|_V = c$. Since u cannot be the zero vector, it follows that $c > 0$. Rescaling (5.12) now yields

$$c\|v\|_1 \leq \|v\|_V,$$

for all $v \in V$. ∎

5.5 Compactness and Uniform Continuity

One of the most important applications of compactness is to uniform continuity. This is used heavily in all areas of approximation.

THEOREM 5.18 *Let X be a compact metric space and let Y be a metric space. Let $f : X \longrightarrow Y$ be a continuous mapping, then f is uniformly continuous.*

Proof. Let $\epsilon > 0$. We apply the continuity of f. At each point $x \in X$ there exist a number $\delta_x > 0$ such that

$$f(U(x, \delta_x)) \subseteq U(f(x), \tfrac{1}{2}\epsilon). \tag{5.14}$$

We can now write

$$X = \bigcup_{x \in X} U(x, \tfrac{1}{2}\delta_x).$$

Applying the compactness of X there is a finite subset $F \subseteq X$ such that

$$X = \bigcup_{x \in F} U(x, \tfrac{1}{2}\delta_x). \tag{5.15}$$

Now let $\delta = \min_{x \in F} \frac{1}{2}\delta_x$. We claim that this δ works in the definition of uniform continuity. Let z_1 and z_2 be points of X satisfying $d(z_1, z_2) < \delta$. By (5.15), there exists $x \in F$ such that $z_1 \in U(x, \frac{1}{2}\delta_x)$. Now, using the triangle inequality we have

$$d(x, z_2) \le d(x, z_1) + d(z_1, z_2) < \tfrac{1}{2}\delta_x + \delta \le \delta_x,$$

so that both z_1 and z_2 lie in $U(x, \delta_x)$. It now follows from (5.14) that $f(z_1)$ and $f(z_2)$ both lie in $U(f(x), \frac{1}{2}\epsilon)$. It then follows again by the triangle inequality that $d(f(z_1), f(z_2)) < \epsilon$ as required. ■

The following Theorem is a typical application of the use of uniform continuity in approximation theory.

THEOREM 5.19 (BERNSTEIN APPROXIMATION THEOREM) *Let $f : [0,1] \longrightarrow \mathbb{R}$ be a continuous function. Define the nth* ***Bernstein polynomial*** *by*

$$B_n(f, x) = \sum_{k=0}^{n} {}^nC_k f(\tfrac{k}{n}) x^k (1 - x)^{n-k}.$$

Then $(B_n(f, \cdot))$ converges uniformly to f on $[0,1]$.

Sketch proof. We leave the proof of the following three identities to the reader

$$1 = \sum_{k=0}^{n} {}^nC_k x^k (1 - x)^{n-k}, \tag{5.16}$$

$$nx = \sum_{k=0}^{n} k \; {}^nC_k x^k (1 - x)^{n-k}, \tag{5.17}$$

$$n(n - 1)x^2 = \sum_{k=0}^{n} k(k - 1) \; {}^nC_k x^k (1 - x)^{n-k}. \tag{5.18}$$

Then it is easy to see that

$$\begin{aligned} \sum_{k=0}^{n} \left(x - \frac{k}{n} \right)^2 {}^nC_k x^k (1 - x)^{n-k} &= \sum_{k=0}^{n} \left(x^2 - \frac{2k}{n} x + \frac{k^2 - k}{n^2} + \frac{k}{n^2} \right) {}^nC_k x^k (1 - x)^{n-k}, \\ &= x^2 - 2x^2 + \frac{n(n - 1)}{n^2} x^2 + \frac{n}{n^2} x, \\ &= \frac{1}{n} x(1 - x). \end{aligned}$$

by applying (5.16), (5.17) and (5.18). Then for $\delta > 0$ we obtain a Tchebychev inequality

$$\begin{aligned}
\sum_{|x-\frac{k}{n}|>\delta} \delta^2 \; {}^nC_k x^k (1-x)^{n-k} &\leq \sum_{|x-\frac{k}{n}|>\delta} \left(x - \frac{k}{n}\right)^2 \; {}^nC_k x^k (1-x)^{n-k} \\
&\leq \frac{1}{n} x(1-x).
\end{aligned}$$

We are now ready to study the approximation. Since f is continuous on the compact set $[0,1]$ it is also uniformly continuous. Furthermore, by Corollary 5.14 (page 70), f is bounded. Thus we have

$$\begin{aligned}
f(x) - B_n(f,x) &= f(x) - \sum_{k=0}^{n} {}^nC_k f(\tfrac{k}{n}) x^k (1-x)^{n-k}, \\
&= \sum_{k=0}^{n} {}^nC_k (f(x) - f(\tfrac{k}{n})) x^k (1-x)^{n-k},
\end{aligned}$$

and

$$\begin{aligned}
|f(x) - B_n(f,x)| &\leq \sum_{k=0}^{n} {}^nC_k |f(x) - f(\tfrac{k}{n})| x^k (1-x)^{n-k}, \\
&\leq E_1 + E_2,
\end{aligned}$$

where

$$\begin{aligned}
E_1 &= \sum_{|x-\frac{k}{n}|>\delta} {}^nC_k |f(x) - f(\tfrac{k}{n})| x^k (1-x)^{n-k}, \\
&\leq 2\|f\|_\infty \delta^{-2} \frac{1}{n} x(1-x), \\
&\leq \frac{1}{2n} \|f\|_\infty \delta^{-2},
\end{aligned}$$

and

$$\begin{aligned}
E_2 &= \sum_{|x-\frac{k}{n}|\leq\delta} {}^nC_k |f(x) - f(\tfrac{k}{n})| x^k (1-x)^{n-k}, \\
&\leq \sum_{k=0}^{n} {}^nC_k \omega_f(\delta) x^k (1-x)^{n-k}, \\
&= \omega_f(\delta).
\end{aligned}$$

Let now $\epsilon > 0$. Then, using the uniform continuity of f, choose $\delta > 0$ so small that $\omega_f(\delta) < \frac{1}{2}\epsilon$. Then, with δ now fixed, select N so large that $\frac{1}{2N}\|f\|_\infty \delta^{-2} < \frac{1}{2}\epsilon$. It follows that

$$\sup_{0\leq x\leq 1} |f(x) - B_n(f,x)| \leq \epsilon \qquad \forall n \geq N,$$

as required for uniform convergence of the Bernstein polynomials to f. ■

5.6 Compactness and Uniform Convergence

There are also some applications of compactness to establish uniform convergence.

Proposition 5.20 (Dini's Theorem) *Let X be a compact metric space and suppose that (f_n) is a sequence of real-valued continuous functions on X decreasing to 0. That is*

$$f_n(x) \geq f_{n+1}(x) \qquad \forall n \in \mathbb{N},\ \forall x \in X$$

and for each fixed $x \in X$,

$$f_n(x) \stackrel{n\to\infty}{\Leftrightarrow} 0.$$

Then (f_n) converges to 0 uniformly.

Proof. Obviously $f_n(x) \geq 0$ for all $n \in \mathbb{N}$ and $x \in X$. Let $\epsilon > 0$. Then for each $x \in X$ there exists N_x such that

$$n \geq N_x \qquad \Rightarrow \qquad f_n(x) < \tfrac{1}{2}\epsilon.$$

Now for all $x \in X$ there exists $\delta_x > 0$ such that

$$d(z,x) < \delta_x \qquad \Rightarrow \qquad |f_{N_x}(z) \Leftrightarrow f_{N_x}(x)| < \tfrac{1}{2}\epsilon.$$

We can now write

$$X = \bigcup_{x \in X} U(x, \delta_x).$$

Applying the compactness of X there is a finite subset $F \subseteq X$ such that

$$X = \bigcup_{x \in F} U(x, \delta_x).$$

Now let $N = \max_{x \in F} N_x$. We will show that $n \geq N$ implies that $f_n(z) < \epsilon$ simultaneously for all $z \in X$.

To verify this, let $z \in X$. Then we find $x \in F$ such that $d(z,x) < \delta_x$. It follows from this that $|f_{N_x}(z) \Leftrightarrow f_{N_x}(x)| < \frac{1}{2}\epsilon$. But, combining this with $f_{N_x}(x) < \frac{1}{2}\epsilon$ we obtain $f_{N_x}(z) < \epsilon$. Finally, since the sequence $(f_n(z))$ is decreasing we have $f_n(z) < \epsilon$ for all $n \geq N$. ∎

5.7 Equivalence of Compactness and Sequential Compactness

We begin with a Theorem whose proof parallels the proof of the Heine–Borel Theorem (page 68).

Theorem 5.21 *Every separable sequentially compact space is compact.*

Proof. Let X be a separable sequentially compact metric space. We must show that X is compact. By Proposition 5.7 it is enough to show that X is countably compact. Thus, let V_k be open subsets of X for $k \in \mathbb{N}$. Suppose that $\cup_{k \in \mathbb{N}} V_k = X$. We will show that there exists a finite subset F of $\mathbb{N}$ such that

$$\bigcup_{k \in F} V_k = X. \tag{5.19}$$

Suppose not. Then, for every k, the set $F = \{1, 2, \ldots, k\}$ fails to satisfy (5.19). So, there is a point $x_k \in X$ with

$$x_k \notin \bigcup_{\ell=1}^{k} V_\ell. \tag{5.20}$$

Now by sequential compactness, the sequence (x_k) has a subsequence which converges to some element $x \in X$. By hypothesis, there exists $m \in \mathbb{N}$ such that $x \in V_m$. Since V_m is open, some tail of the subsequence lies entirely inside V_m. Therefore, there exists $k \in \mathbb{N}$ with $k \geq m$ and $x_k \in V_m$. This contradiction with (5.20) establishes the result. ∎

DEFINITION *A metric space X is said to be* ***totally bounded*** *iff for every $\epsilon > 0$ there exists a finite subset F of X such that*

$$\bigcup_{x \in F} U(x, \epsilon) = X. \tag{5.21}$$

THEOREM 5.22 *If X is a sequentially compact metric space, then X is totally bounded.*

Proof. Let $\epsilon > 0$ and suppose that (5.21) fails for every finite subset F of X. We will obtain a contradiction with the sequential compactness of X.

We define a sequence (x_n) inductively. Let x_1 be any point of X. We observe that X cannot be empty since then (5.21) holds with $F = \emptyset$. Now assume that $x_1, \ldots, x_n$ have been defined. We choose x_{n+1} such that

$$x_{n+1} \in X \setminus \bigcup_{k=1}^{n} U(x_k, \epsilon).$$

Once again, it is the failure of (5.21), this time with $F = \{x_1, \ldots, x_n\}$ which guarantees the existence of x_{n+1}.

Since X is sequentially compact, the sequence (x_n) possesses a subsequence convergent to some point x of X. Hence there exists N such that $x_N \in U(x, \epsilon)$. Thus $x \in U(x_N, \epsilon)$. Using the fact that $U(x_N, \epsilon)$ is open and hence a neighbourhood of x, and the convergence of the subsequence to x we see that there exists $n > N$ with $x_n \in U(x_N, \epsilon)$. But this contradicts the definition of x_n. ∎

An extension of this result will be needed later.

DEFINITION *A subset Y of a metric space X is said to be* ***totally bounded*** *iff for every $\epsilon > 0$ there exists a finite subset F of Y such that*

$$\bigcup_{y \in F} U(y, \epsilon) \supseteq Y.$$

It is almost immediate from the definition that if Y is a totally bounded subset, then so is $\mathrm{cl}(Y)$. In essence this is because of the inclusion chain

$$\bigcup_{y \in F} U(y, 2\epsilon) \supseteq \bigcup_{y \in F} B(y, \epsilon) \supseteq \bigcup_{y \in F} U(y, \epsilon) \supseteq Y.$$

It follows from this that

$$\bigcup_{y \in F} U(y, 2\epsilon) \supseteq \mathrm{cl}(Y).$$

The proof of the following result follows that of Theorem 5.22 so closely that we leave the details to the reader.

THEOREM 5.23 *Let Y be a subset of a metric space X. If $\mathrm{cl}(Y)$ is sequentially compact, then Y is a totally bounded subset of X*

Another very easy result is the following.

PROPOSITION 5.24 *Every totally bounded metric space is separable.*

Sketch proof. Choose a sequence (ϵ_n) of strictly positive reals decreasing to zero. For each $n \in \mathbb{N}$ apply the total boundedness condition to obtain a finite subset F_n of X such that

$$\bigcup_{x \in F_n} U(x, \epsilon_n) = X.$$

We leave the reader to show that

$$\bigcup_{n \in \mathbb{N}} F_n$$

is a countable dense subset of X. ∎

We can now establish the converse to Proposition 5.8 (page 67).

COROLLARY 5.25 *Every sequentially compact metric space is compact.*

Proof. This is an immediate consequence of Theorem 5.21, Theorem 5.22 and Proposition 5.24.

5.8 Compactness and Completeness

PROPOSITION 5.26 *A sequentially compact metric space is complete.*

Proof. Let (x_n) be a Cauchy sequence in a sequentially compact metric space X. Then there is a subsequence (x_{n_k}) converging to some element $x \in X$. We use the convergence of this subsequence and the Cauchy condition to establish the convergence of the original sequence to x.

Let $\epsilon > 0$. Then, by the Cauchy condition, there exists N such that

$$p, q > N \qquad \Rightarrow \qquad d(x_p, x_q) < \tfrac{1}{2}\epsilon.$$

By convergence of the subsequence we also have

$$k > K \qquad \Rightarrow \qquad d(x_{n_k}, x) < \tfrac{1}{2}\epsilon.$$

Let us choose $k = \max(N, K) + 1$. Then taking $q = n_k$ and using the triangle inequality, we obtain

$$p > N \qquad \Rightarrow \qquad d(x_p, x) < \epsilon.$$

as required to establish convergence. ∎

THEOREM 5.27 *A complete totally bounded metric space is sequentially compact.*

Proof. This proof uses the famous diagonal subsequence argument. Let X be a complete totally bounded metric space. Let (x_k) be a sequence in X. Let (ϵ_k) be a sequence of strictly positive reals decreasing to zero. For each $k \in \mathbb{N}$ we apply the total boundedness condition to obtain a finite subset F_k of X such that

$$\bigcup_{x \in F_k} U(x, \epsilon_k) = X.$$

We extract subsequences inductively.

$$\begin{array}{ccccc}
x_{n_{1,1}} & x_{n_{1,2}} & x_{n_{1,3}} & x_{n_{1,4}} & \cdots \\
x_{n_{2,1}} & x_{n_{2,2}} & x_{n_{2,3}} & x_{n_{2,4}} & \cdots \\
x_{n_{3,1}} & x_{n_{3,2}} & x_{n_{3,3}} & x_{n_{3,4}} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_{n_{k,1}} & x_{n_{k,2}} & x_{n_{k,3}} & x_{n_{k,4}} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

The first subsequence $(x_{n_{1,\ell}})$ is a subsequence of (x_n) contained in a set $U(s_1, \epsilon_1)$ for some $s_1 \in F_1$. This uses the fact that the sequence (x_n) cannot meet all such $U(s_1, \epsilon_1)$ in a finite set. The same reasoning allows us to extract the second subsequence $(x_{n_{2,\ell}})$ from $(x_{n_{1,\ell}})$ so that $(x_{n_{2,\ell}})$ is contained in a set $U(s_2, \epsilon_2)$ for some $s_2 \in F_2$. We continue in this way.

We now consider the ***diagonal subsequence*** (x_{m_k}) defined by

$$x_{m_k} = x_{n_{k,k}}.$$

The crucial observation is that, for each k, the tail sequence $x_{m_k}, x_{m_{k+1}}, x_{m_{k+2}}, \ldots$ is itself a subsequence of $x_{n_{k,1}}, x_{n_{k,2}}, x_{n_{k,3}}, x_{n_{k,4}}, \ldots$. Thus, for each $k \in \mathbb{N}$ the tail sequence $x_{m_k}, x_{m_{k+1}}, x_{m_{k+2}}, \ldots$ lies in $U(s_k, \epsilon_k)$ and hence has diameter less than $2\epsilon_k$. It follows immediately that (x_{m_k}) is a Cauchy sequence and hence convergent in X. ■

5.9 Equicontinuous Sets

Throughout this section, K denotes a compact metric space. We denote by $C(K)$ the space of bounded real-valued continuous functions on K. All the proofs presented here also work for complex valued functions. We consider $C(K)$ as a normed space with the uniform norm. We have already observed that $C(K)$ is complete with this norm — see the example following Proposition 4.7 (page 53).

DEFINITION *Let $F \subseteq C(K)$. We say that F is* **equicontinuous** *iff for all $\epsilon > 0$, there exists $\delta > 0$ such that*

$$d_K(x, y) < \delta \qquad \Rightarrow \qquad |f(x) \Leftrightarrow f(y)| < \epsilon \qquad \forall f \in F.$$

It is more or less clear that a set F is equicontinuous iff there is a "modulus of continuity function" that works simultaneously for all functions in the set F. Explicitly we have the following Lemma, the proof of which is left as an exercise to the reader.

LEMMA 5.28 *Let $F \subseteq C(K)$. Then F is equicontinuous iff there is a function $\omega : [0, \infty[\Leftrightarrow\!\rightarrow [0, \infty[$ satisfying $\omega(0) = 0$ and continuous at 0, such that*

$$|f(x) \Leftrightarrow f(y)| \le \omega(d_K(x, y)) \qquad \forall x, y \in K,\ \forall f \in F.$$

The key result of this section is the following.

THEOREM 5.29 (ASCOLI–ARZELA THEOREM) *Let $F \subseteq C(K)$. Then the following are equivalent statements.*

- *F has compact closure in $C(K)$.*
- *F is bounded in $C(K)$ and F is equicontinuous.*

Proof. We assume first that F has compact closure in $C(K)$. Then according to Proposition 5.1 (page 64), F is bounded in $C(K)$. We show that F is equicontinuous. Suppose not. Then there exists $\epsilon > 0$, two sequences (x_n) and (y_n) in K and a sequence (f_n) in F such that $(d(x_n, y_n))$ converges to 0 and $|f_n(x_n) \Leftrightarrow f_n(y_n)| \ge \epsilon$. Now using the hypothesis that F has compact closure, we see that (f_n) has a subsequence convergent in $C(K)$. We denote the limit function by f. Then there exists $n \in \mathbb{N}$ such that

$$\|f \Leftrightarrow f_n\|_\infty \le \tfrac{1}{3}\epsilon.$$

It now follows that $|f(x_n) \Leftrightarrow f(y_n)| \geq \frac{1}{3}\epsilon$, contradicting the uniform continuity of f. The function f is uniformly continuous by virtue of Theorem 5.18 (page 72).

The real work of the proof is contained in the converse. Let us assume that F is bounded and equicontinuous. It suffices to show that F is totally bounded. For then we will have that $\mathrm{cl}(F)$ is totally bounded (by a remark on page 75) and complete, since $C(K)$ is complete. It is then enough to apply Theorem 5.27 and Corollary 5.25 to deduce that $\mathrm{cl}(F)$ is compact in $C(K)$.

Let $\epsilon > 0$. Then using the equicontinuity, we can find $\delta > 0$ such that

$$d(x, y) < \delta, f \in F \qquad \Rightarrow \qquad |f(x) \Leftrightarrow f(y)| < \tfrac{1}{3}\epsilon. \tag{5.22}$$

Now using the total boundedness of K, we can find $N \in \mathbb{N}$ and $x_1, x_2, \ldots x_N \in K$ such that

$$\bigcup_{n=1}^{N} U(x_n, \delta) = K. \tag{5.23}$$

Since F is bounded, the set

$$\{(f(x_1), f(x_2), \ldots, f(x_N)); f \in F\}$$

is a bounded subset of $\mathbb{R}^N$ and hence is totally bounded in $\mathbb{R}^N$ by the Heine–Borel Theorem and Theorem 5.23. Hence there exists $M \in \mathbb{N}$ and functions $f_1, f_2, \ldots, f_M$ in F such that for all $f \in F$ there exists m with $1 \leq m \leq M$ such that

$$\sup_{n=1}^{N} |f(x_n) \Leftrightarrow f_m(x_n)| \leq \tfrac{1}{3}\epsilon. \tag{5.24}$$

Combining now (5.22), (5.23) and (5.24) we have, for an arbitrary point $x \in K$ and n chosen such that $d(x, x_n) < \delta$,

$$|f(x) \Leftrightarrow f_m(x)| \leq |f(x) \Leftrightarrow f(x_n)| + |f(x_n) \Leftrightarrow f_m(x_n)| + |f_m(x) \Leftrightarrow f_m(x_n)| < \epsilon.$$

It follows that F is totally bounded in $C(K)$. ■

5.10 The Stone–Weierstrass Theorem

One of the key approximation theorems in analysis is the Stone–Weierstrass Theorem. Let K be a compact metric space and recall that the notation $C(K)$ stands for the space of real-valued continuous functions on K. The space $C(K)$ is a vector space under pointwise operations, a fact that we have already heavily used. It is also a linear associative algebra — this means that $C(K)$ is closed under pointwise multiplication and that it satisfies the axioms for a ring.

DEFINITION *A subset $A \subseteq C(K)$ is said to be a* ***subalgebra*** *of $C(K)$ if it is closed under both the linear and multiplicative operations. Explicitly, this means*

- $f_1, f_2 \in A, t_1, t_2 \in \mathbb{R} \Longrightarrow t_1 f_1 + t_2 f_2 \in A.$
- $f_1, f_2 \in A \Longrightarrow f_1 \cdot f_2 \in A.$

where the function combinations are defined by

$$(t_1 f_1 + t_2 f_2)(x) = t_1 f_1(x) + t_2 f_2(x) \qquad \forall x \in K$$

and

$$(f_1 \cdot f_2)(x) = f_1(x) f_2(x) \qquad \forall x \in K.$$

Definition *A subalgebra A of $C(K)$ is said to be* **unital** *iff the constant function* $\mathbf{1}$ *which is identically equal to 1 belongs to A. A subalgebra A is said to* **separating** *iff whenever x_1 and x_2 are two distinct points of K, there exists a function $f \in A$ such that $f(x_1) \neq f(x_2)$.*

If A is both unital and separating, then whenever x_1 and x_2 are two distinct points of K and a_1 and a_2 are given real numbers, there exists $f \in A$ such that $f(x_1) = a_1$ and $f(x_2) = a_2$.

Theorem 5.30 (Stone–Weierstrass Theorem) *Let K be a compact metric space and suppose that A be a unital separating subalgebra of $C(K)$. Then A is dense in $C(K)$ for the standard uniform metric on $C(K)$.*

Before we can prove this result, we need to develop some preliminary ideas.

Lemma 5.31 *If A is a unital separating subalgebra of $C(K)$, then so is its uniform closure $\mathrm{cl}(A)$.*

Proof. Obviously $\mathrm{cl}(A)$ is unital and separating because $A \subseteq \mathrm{cl}(A)$. It remains to check that $\mathrm{cl}(A)$ is a subalgebra. This is routine. For instance, to show that $f \cdot g \in \mathrm{cl}(A)$ whenever $f, g \in \mathrm{cl}(A)$, we find a sequence (f_n) in A converging uniformly to f and a sequence (g_n) converging uniformly to g. Clearly

$$\begin{aligned} \|f \cdot g - f_n \cdot g_n\| &= \|(f - f_n) \cdot g + f_n \cdot (g - g_n)\| \\ &\leq \|(f - f_n) \cdot g\| + \|f_n \cdot (g - g_n)\| \\ &= \|f - f_n\| \|g\| + \|f_n\| \|g - g_n\| \\ &\longrightarrow 0 \end{aligned}$$

so that $(f_n \cdot g_n)$ converges uniformly to $f \cdot g$. The proof that $\mathrm{cl}(A)$ is closed under linear operations is similar. ■

The upthrust of Lemma 5.31 is that the Stone–Weierstrass Theorem can be reformulated in the following way.

Theorem 5.32 *Let K be a compact metric space and suppose that A be a uniformly closed unital separating subalgebra of $C(K)$. Then $A = C(K)$.*

Lemma 5.33 *Let $a > 0$. Then there exists a sequence (p_n) of real polynomials such that $p_n(x) \longrightarrow |x|$ uniformly on $[-a, a]$.*

This is a consequence of Theorem 5.19 after some elementary rescaling, but it is also possible to give a proof *à la main*.

Proof. There are various pitfalls in designing a strategy for the proof. For instance, taking the p_n to be the partial sums of a fixed power series is doomed to failure. The proof has to be fairly subtle.

Without loss of generality one may take $a = 1$. Let us define the function

$$f_n(x) = \sqrt{\frac{2}{\pi}} n e^{-\frac{1}{2}n^2x^2}.$$

The key facts about this function are that it is positive, that

$$\int_{-\infty}^{\infty} f_n(x)dx = 2$$

and that for large values of n, the graph of f_n has a "spike" near 0. Let ϵ_n be a sequence of strictly positive numbers converging to 0. Let r_n be an even polynomial such that

$$\sup_{-1 \leq x \leq 1} |f_n(x) - r_n(x)| \leq \epsilon_n.$$

We can easily construct r_n by truncating the power series expansion of f_n. If ϵ_n is suitably small, r_n will have this same spiky behaviour. Following this philosophy, we can expect the polynomial q_n given by

$$q_n(s) = \int_0^s r_n(t)dt \tag{5.25}$$

to approximate the "signum" function, and the second primitive p_n

$$p_n(x) = \int_0^x q_n(s)ds \tag{5.26}$$

should approximate the "modulus" function. One can actually obtain p_n directly from r_n by

$$p_n(x) = \int_0^x (x \Leftrightarrow t) r_n(t)dt.$$

Clearly, from (5.25) and (5.26) p_n is an even polynomial function. Since both $|x|$ and $p_n(x)$ are even in x we need only estimate

$$\sup_{0\le x\le 1} |x \Leftrightarrow p_n(x)|. \tag{5.27}$$

Clearly

$$\left| \int_0^x (x \Leftrightarrow t) r_n(t)dt \Leftrightarrow \int_0^x (x \Leftrightarrow t) f_n(t)dt \right| \le \epsilon_n$$

for all $x \in [0,1]$ so that it is enough to show that

$$\sup_{0\le x\le 1} \left| x \Leftrightarrow \int_0^x (x \Leftrightarrow t) f_n(t)dt \right|$$

tends to zero as n tends to infinity. We have

$$\begin{aligned} \sup_{0\le x\le 1} \left| x \Leftrightarrow \int_0^x (x \Leftrightarrow t) f_n(t)dt \right| &= \sup_{0\le x\le 1} \left| \int_0^\infty x f_n(t)dt \Leftrightarrow \int_0^x (x \Leftrightarrow t) f_n(t)dt \right| \\ &\le \left\{ \sup_{0\le x\le 1} x \int_x^\infty f_n(t)dt \right\} + \int_0^\infty t f_n(t)dt \end{aligned} \tag{5.28}$$

The second term in (5.28) is independent of x and tends to zero like $\frac{1}{n}$ so we concentrate on the first term which, after making a change of variables in the integral, can be rewritten

$$\sqrt{\frac{2}{\pi}} \sup_{0\le x\le 1} x \int_{nx}^\infty e^{-\frac{1}{2}t^2} dt \tag{5.29}$$

But (5.29) also tends to 0 as n tends to ∞ since

$$x \int_{nx}^\infty e^{-\frac{1}{2}t^2} dt \le \begin{cases} n^{-\frac{1}{2}} \int_0^\infty e^{-\frac{1}{2}t^2} dt & \text{if } 0 \le x \le n^{-\frac{1}{2}}, \\ \int_{\sqrt{n}}^\infty e^{-\frac{1}{2}t^2} dt & \text{if } n^{-\frac{1}{2}} \le x \le 1. \end{cases}$$

These estimates show that (5.27) tends to 0 as n tends to ∞ as required. ■

Lemma 5.34 *Let A be a uniformly closed unital subalgebra of $C(K)$. Let $f, g \in A$. Then the functions $\max(f,g)$ and $\min(f,g)$ are also in A.*

Proof. Since we have the identities

$$\max(f,g) = \tfrac{1}{2}(f + g + |f \Leftrightarrow g|)$$

and

$$\min(f,g) = \tfrac{1}{2}(f + g \Leftrightarrow |f \Leftrightarrow g|)$$

it is enough to establish that if $h \in A$ then $|h| \in A$. For then, taking $h = f \Leftrightarrow g$ the result follows. Since h is a continuous function defined on a compact space, it is bounded and hence it takes values in $[\Leftrightarrow a, a]$ for some $a > 0$. It now follows from Lemma 5.33 that the sequence of functions $(p_n \circ h)$ converges uniformly to $|h|$. Each function $p_n \circ h$ is in A since A is a unital subalgebra of $C(K)$. Hence, since A is also uniformly closed it follows that $|h| \in A$. ■

Proof of the Stone–Weierstrass Theorem. We start with a function $f \in C(K)$ that we wish to approximate and a positive number ϵ which is the allowed uniform error. Let x be an arbitrary point of K which we fix for the moment. Now let y be another arbitrary point of K which we allow to vary. Since A is both unital and separating, we can find a function $h_{x,y} \in A$ such that

$$h_{x,y}(x) = f(x)$$

and

$$h_{x,y}(y) = f(y).$$

Since both f and $h_{x,y}$ are continuous at y there is an open neighbourhood $V_{x,y}$ of y such that $h_{x,y}(z) \Leftrightarrow f(z) < \epsilon$ for all $z \in V_{x,y}$. Clearly we have for each fixed x

$$K = \bigcup_{y \in K} V_{x,y},$$

and hence by the compactness of K, there exist $m \in \mathbb{N}$ and $y_1, \ldots, y_m \in K$ such that

$$K = \bigcup_{k=1}^{m} V_{x,y_k}.$$

It is worth pointing out that m and the points $y_1, \ldots, y_m$ depend on x, but it would be too cumbersome to express this fact notationally.

The function

$$g_x = \min_{k=1}^{m} h_{x,y_k}$$

is in $\mathrm{cl}(A)$ because of Lemma 5.34 and has the following properties

$$g_x(x) = f(x)$$

and

$$g_x(z) < f(z) + \epsilon \qquad \forall z \in K. \tag{5.30}$$

We note that (5.30) holds since, for all $z \in K$ there exists k with $1 \le k \le m$ such that $z \in V_{x,y_k}$. We then have

$$g_x(z) \le h_{x,y_k}(z) < f(z) + \epsilon.$$

Dependence on y has now been eliminated, and we now allow x to vary. Since f and g_x are continuous at x, there is an open neighbourhood U_x of x such that

$$g_x(z) > f(z) \Leftrightarrow \epsilon \qquad \forall z \in U_x \tag{5.31}$$

Clearly we have

$$K = \bigcup_{x \in K} U_x,$$

and hence by the compactness of K, there exist $\ell \in \mathbb{N}$ and $x_1, \ldots, x_\ell \in K$ such that

$$K = \bigcup_{j=1}^{\ell} U_{x_j}.$$

The function

$$g = \max_{j=1}^{\ell} g_{x_j}$$

is in $\mathrm{cl}(A)$ applying Lemma 5.34 again. We check that both the inequalities

$$f(z) - \epsilon < g(z) < f(z) + \epsilon \qquad \forall z \in K$$

hold. The inequality on the right holds because of (5.30). For the inequality on the left, let $z \in K$. Then there exists j with $1 \le j \le \ell$ such that $z \in U_{x_j}$. We have, using (5.31)

$$g(z) \ge g_{x_j}(z) > f(z) - \epsilon.$$

We have shown that $\mathrm{cl}(A)$ is dense in $C(K)$ and hence we conclude that $\mathrm{cl}(A) = C(K)$ as required. ■

EXAMPLE Let $K = [-1, 1]$ and let A be the algebra of (restrictions of) polynomial functions. Then A is clearly a unital separating subalgebra of $C(K)$ and is therefore (uniformly) dense in $C(K)$. Of course this example contains Lemma 5.33 as a special case. □

EXAMPLE Let $K = [0, 1] \times [0, 1]$ the unit square. If f and g are continuous functions on $[0, 1]$, we can make a new function on K by

$$(f \otimes g)(s, t) = f(s) \cdot g(t). \tag{5.32}$$

It is easy to see that the set A of all finite sums of such functions is a unital separating subalgebra of $C(K)$. In fact, as linear spaces, we have

$$A \cong C([0, 1]) \otimes C([0, 1])$$

the tensor product of $C([0, 1])$ with itself. This is the reason for using the $\otimes$ notation in (5.32). The Stone–Weierstrass Theorem shows that A is dense in $C(K)$. There are many quite difficult problems associated with this example, for instance it is true, but not immediately obvious that A is a proper subalgebra of $C(K)$. □

The Stone–Weierstrass Theorem allows a number of extensions. First of all, there is an extension to complex-valued continuous functions.

THEOREM 5.35 *Let K be a compact metric space and suppose that A be a unital separating self-adjoint subalgebra of $C(K, \mathbb{C})$. Then A is dense in $C(K, \mathbb{C})$ for the uniform metric.*

Here, the condition that A is ***self-adjoint*** means that $\overline{f} \in A$ whenever $f \in A$. The function $\overline{f}$ is defined by

$$\overline{f}(z) = \overline{f(z)} \qquad \forall z \in K.$$

Proof. The key observation is that if $f \in A$ then $\Re f = \frac{1}{2}(f + \overline{f})$ is also in A. It is now easy to see that

$$\Re A = \{\Re f; f \in A\} = A \cap C(K, \mathbb{R})$$

is a unital separating subalgebra of $C(K, \mathbb{R})$. Thus applying the standard Stone–Weierstrass Theorem we see that $\Re A$ is dense in $C(K, \mathbb{R})$. Since $\Re A \subseteq A$ the result follows immediately. ■

Example Perhaps the most interesting application is to trigonometric polynomials. Let $\mathbb{T}$ denote the quotient group $\mathbb{R}/2\pi\mathbb{Z}$. Here we are viewing $2\pi\mathbb{Z}$ as an additive subgroup of $\mathbb{R}$ considered as an abelian group. We can think of $\mathbb{T}$ as the reals modulo 2π. Topologically, $\mathbb{T}$ is a circle, so it is called the circle group. In particular, $\mathbb{T}$ is compact, if for instance it is given the metric

$$d_{\mathbb{T}}(\dot{t}, \dot{s}) = \inf_{n \in \mathbb{Z}} |2n\pi + t - s|$$

for $t, s \in \mathbb{R}$ and $\dot{t}, \dot{s}$ denoting the corresponding points of $\mathbb{T}$. A ***trigonometric polynomial*** is a complex-valued function p on $\mathbb{R}$ given by a finite sum

$$p(t) = \sum_{n=-N}^{N} a_n e^{int} \qquad t \in \mathbb{R}$$

where $a_n \in \mathbb{C}$. The function p is 2π-periodic when viewed as a function on $\mathbb{R}$ and hence it may be considered as a function on $\mathbb{T}$. It is straightforward to see that the set A of all trigonometric polynomials is a unital self-adjoint separating subalgebra of $C(\mathbb{T}, \mathbb{C})$. It follows that A is dense in $C(\mathbb{T}, \mathbb{C})$. This result is very important in the theory of Fourier series, although it is normally approached from a rather different angle. □

The second major extension of the Stone–Weierstrass Theorem involves ***one-point compactifications***. A complete treatment is outside the scope of these notes. It is however possible to give the general idea.

Given a space such as $\mathbb{R}$ it is possible to add a ***point at infinity*** designated ∞, and define a new metric $\tilde{d}$ on the resulting space. Consider the map $f : \mathbb{R} \longrightarrow \mathbb{R}^2$ defined by

$$f(u) = (\frac{1 - u^2}{1 + u^2}, \frac{2u}{1 + u^2})$$

which actually maps into the unit circle S^1 in $\mathbb{R}^2$. The only point of the unit circle which is not in the image of f is the point $(-1, 0)$. We treat this point as if it were $f(\infty)$. We define

$$\tilde{d}(u, v) = \|f(u) - f(v)\|$$

where $\| \ \|$ is the standard Euclidean norm on $\mathbb{R}^2$ and

$$\tilde{d}(\infty, v) = \tilde{d}(v, \infty) = \|(-1, 0) - f(v)\|.$$

Together with the required $\tilde{d}(\infty, \infty) = 0$, this clearly defines a metric on $\mathbb{R} \cup \{\infty\}$, because it is really just the Euclidean distance on the unit circle. If we restrict the metric $\tilde{d}$ to $\mathbb{R}$ we obtain a metric which is topologically equivalent to, but not uniformly equivalent to the standard metric on $\mathbb{R}$. Clearly, $(\mathbb{R} \cup \{\infty\}, \tilde{d})$ is a compact metric space because the unit circle is a compact subset of $\mathbb{R}^2$. This space is called the one-point compactification of $\mathbb{R}$. Similar constructions lead to one-point compactifications of many other spaces. For instance, the one-point compactification of $\mathbb{R}^n$ can be identified to the n-sphere S^n in $\mathbb{R}^{n+1}$. There is also a very natural two-point compactification of $\mathbb{R}$ which can be denoted $[-\infty, \infty]$.

Definition *Let $f : \mathbb{R} \longrightarrow \mathbb{R}$. We say that f **possesses a limit** a **at infinity** if and only if for all $\epsilon > 0$, there exists $A > 0$ such that $|f(x) - a| < \epsilon$ whenever $|x| > A$. We also say that f **vanishes at infinity** if and only if f possesses the limit 0 at infinity.*

The set of all continuous real-valued functions on $\mathbb{R}$ vanishing at infinity will be denoted $C_0(\mathbb{R})$. It is easy to see that $C_0(\mathbb{R})$ is a uniformly closed subalgebra of $C(\mathbb{R})$. The following Proposition is left as an exercise.

Proposition 5.36 *A continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ extends to a continuous function $\tilde{f} : \mathbb{R} \cup \{\infty\} \longrightarrow \mathbb{R}$ if and only if f possesses a limit at infinity.*

THEOREM 5.37 *Let A be a separating subalgebra of $C_0(\mathbb{R})$ that separates from infinity. Explicitly, this last statement means that for all $x \in \mathbb{R}$ there exists $f \in A$ such that $f(x) \neq 0$. Then A is uniformly dense in $C_0(\mathbb{R})$.*

Proof. Let B denote the set $\{f + \lambda \mathbf{1}; f \in A, \lambda \in \mathbb{R}\}$, an algebra of continuous functions on $\mathbb{R}$ which possess a limit at infinity. The set of lifts $\tilde{B} = \{\tilde{g}; g \in B\}$ is then a unital subalgebra of $C(\mathbb{R} \cup \{\infty\})$. The hypotheses on A guarantee that $\tilde{B}$ separates the points of $\mathbb{R} \cup \{\infty\}$. Thus by the Stone–Weierstrass Theorem, $\tilde{B}$ is uniformly dense in $C(\mathbb{R} \cup \{\infty\})$. It follows easily that A is uniformly dense in $C_0(\mathbb{R})$. ■

EXAMPLE This example is on $[0, \infty[$ rather than $\mathbb{R}$. Let A consist of all functions $f : [0, \infty[\longrightarrow \mathbb{R}$ of the type

$$f(x) = \sum_{j=1}^{n} a_j e^{-\lambda_j x} \qquad (x \geq 0)$$

where $a_j \in \mathbb{R}$ and $\lambda_j > 0$. Such a function is clearly continuous and vanishes at infinity. The set A is clearly a separating algebra of functions which also separates from infinity. Hence A is uniformly dense in $C_0([0, \infty[)$. This result can be used to yield the Uniqueness Theorem for Laplace transforms. □

6

Connectedness

From the intuitive point of view, a metric space is connected if it is in one piece.

Definition *A **splitting** of a metric space X is a partition*

$$\begin{aligned} X &= X_1 \cup X_2, \\ \emptyset &= X_1 \cap X_2, \end{aligned} \tag{6.1}$$

*where X_1 and X_2 are open subsets of X. The splitting (6.1) is said to be **trivial** iff either $X_1 = X$ and $X_2 = \emptyset$ or $X_1 = \emptyset$ and $X_2 = X$. A metric space X is **connected** iff every splitting of X is trivial.*

In a splitting, the subsets X_1 and X_2, being complements of each other, are also closed.

Proposition 6.1 *Every closed interval $[a, b]$ of $\mathbb{R}$ is connected.*

Proof. If $b < a$ then $[a, b] = \emptyset$. If $a = b$ then $[a, b]$ is a singleton. In either case, all partitions of $[a, b]$, whether they are splittings or not, are trivial.

Let us suppose that $a < b$ and that

$$\begin{aligned} [a, b] &= X_1 \cup X_2, \\ \emptyset &= X_1 \cap X_2, \end{aligned} \tag{6.2}$$

is a splitting of $[a, b]$. We may assume without loss of generality that $a \in X_1$, for if not, it suffices to interchange the sets X_1 and X_2. Let us suppose that $X_2 \neq \emptyset$. Then we define

$$c = \inf X_2.$$

We claim that $a < c$. Since X_1 is open, there exists $\epsilon > 0$ such that $U(a, \epsilon) \subseteq X_1$. This means that $[a, a + \epsilon[\subseteq X_1$ or equivalently that $X_2 \subseteq [a + \epsilon, b]$. It follows that $c \geq a + \epsilon$.

Exactly the same argument shows that if $c \in X_1$ and $c < b$ then there exists $\epsilon > 0$ such that $[c, c + \epsilon[\subseteq X_1$. But since by definition of c, we have $[a, c[\subseteq X_1$, we conclude that $[a, c + \epsilon[\subseteq X_1$ or equivalently $X_2 \subseteq [c + \epsilon, b]$ contradicting the definition of c. On the other hand, if $c = b$ and $c \in X_1$, then X_2 must be empty.

Hence, it must be the case that $c \in X_2$. But then there exists $\epsilon > 0$ such that $]c \Leftrightarrow \epsilon, c + \epsilon[\subseteq X_2$ which also contradicts the definition of c.

We are therefore forced to concluded that the supposition $X_2 \neq \emptyset$ is false. It follows that the splitting (6.2) is trivial. ■

6.1 Connected Subsets

So far we have discussed connectedness for metric spaces. We now extend the concept to subsets in the usual way.

DEFINITION *Let A be a subset of a metric space X. Then A is a* **connected subset** *iff A is a connected metric space in the restriction metric inherited from X.*

When we disentangle this definition using Theorem 2.27 (page 27) we obtain the following complicated Proposition.

PROPOSITION 6.2 *Let X be a metric space and let $A \subseteq X$. Then the following two conditions are equivalent.*

- *A is a connected subset of X.*
- *Whenever V_1 and V_2 are open subsets of X such that*

$$A \subseteq V_1 \cup V_2 \tag{6.3}$$

and

$$\emptyset = A \cap V_1 \cap V_2, \tag{6.4}$$

then either $A \subseteq V_1$ or $A \subseteq V_2$.

We leave the proof of this Proposition to the reader.

LEMMA 6.3 *Let X be a metric space and suppose that A is a connected subset of X. Then for every splitting*

$$X = V_1 \cup V_2, \tag{6.5}$$

$$\emptyset = V_1 \cap V_2, \tag{6.6}$$

of X we must have either $A \subseteq V_1$ or $A \subseteq V_2$.

Proof. It is immediate that (6.3) follows from (6.5) and (6.4) follows from (6.6). The conclusion follows immediately from the connectivity of A. ∎

Another result we can obtain from Proposition 6.2 is the following.

PROPOSITION 6.4 *Let A be a connected subset of a metric space X. Then $\mathrm{cl}(A)$ is also connected.*

Proof. We suppose that V_1 and V_2 are open subsets of X such that

$$\mathrm{cl}(A) \subseteq V_1 \cup V_2 \tag{6.7}$$

and

$$\emptyset = \mathrm{cl}(A) \cap V_1 \cap V_2$$

hold. Then *a fortiori* (6.3) and (6.4) also hold. Since A is connected, we deduce that either $A \subseteq V_1$ or $A \subseteq V_2$. Let us suppose that $A \subseteq V_1$ without loss of generality. We claim that

$$\mathrm{cl}(A) \cap V_2 = \emptyset. \tag{6.8}$$

We establish the claim by contradiction. If $x \in \mathrm{cl}(A) \cap V_2$, then we can find a sequence (x_j) in A converging to x. Since V_2 is open and $x \in V_2$, for j large enough, we will have

$$x_j \in A \cap V_2 \subseteq A \cap V_1 \cap V_2 = \emptyset,$$

a contradiction. The claim is established. But now by (6.7) and (6.8) we find that $\mathrm{cl}(A) \subseteq V_1$. Similarly, supposing that $A \subseteq V_2$ will lead to $\mathrm{cl}(A) \subseteq V_2$. ■

6.2 Connectivity of the Real Line

PROPOSITION 6.5 *Every interval in $\mathbb{R}$ is connected.*

Proof. Let I be an interval of $\mathbb{R}$. We view I as a metric space in its own right and show that it is connected. Suppose not. Then there is a non-trivial splitting of I. Let $a, b \in I$ be points of I on different sides of the splitting. Without loss of generality we may suppose that $a \leq b$. But, by Proposition 6.1 (page 85) the closed interval $[a, b]$ is a connected subset of I containing both a and b. Hence, by Lemma 6.3 a and b must lie on the same side of any splitting – a contradiction. ■

The converse is also true.

THEOREM 6.6 *Every non-empty connected subset of $\mathbb{R}$ is an interval.*

Proof. Let I be a connected subset of $\mathbb{R}$. Let $a = \inf I$ and $b = \sup I$ with the understanding that $a = \Leftrightarrow\infty$ if I is unbounded below and $b = \infty$ if I is unbounded above. It is the order completeness axiom that guarantees the existence of a and b. We claim that $]a, b[\subseteq I$. For if not, there exists $c \notin I$ satisfying $a < c < b$. But then taking $X = \mathbb{R}$, $A = I$, $V_1 =]\Leftrightarrow\infty, c[$ and $V_2 =]c, \infty[$ in Proposition 6.2 (page 86) shows that I is not connected. If $a \in \mathbb{R}$ then a may or may not be in I. If $b \in \mathbb{R}$ then b may or may not be in I. But in any event, I is an interval. ■

6.3 Connected Components

THEOREM 6.7 *Let X be a metric space. Let A and B be connected subsets of X with $A \cap B \neq \emptyset$. Then $A \cup B$ is connected.*

Proof. Let V_1 and V_2 are open subsets of X such that

$$A \cup B \subseteq V_1 \cup V_2$$

and

$$\emptyset = (A \cup B) \cap V_1 \cap V_2, \tag{6.9}$$

then we must show that either $A \cup B \subseteq V_1$ or $A \cup B \subseteq V_2$. It is clear from Lemma 6.3 and the fact that A is connected that either $A \subseteq V_1$ or $A \subseteq V_2$. Similarly, since B is connected, either $B \subseteq V_1$ or $B \subseteq V_2$. There are then 4 possibilities.

- $A \subseteq V_1$ and $B \subseteq V_1$.
- $A \subseteq V_1$ and $B \subseteq V_2$.
- $A \subseteq V_2$ and $B \subseteq V_1$.
- $A \subseteq V_2$ and $B \subseteq V_2$.

We show that the second and third cases are impossible. Suppose for instance that the second case holds. Let $x \in A \cap B$. Then $x \in V_1$ and $x \in V_2$. From this it follows that $x \in (A \cup B) \cap V_1 \cap V_2$ which contradicts

(6.9). The third case is impossible by similar reasoning. It follows that either the first case holds, so that $A \cup B \subseteq V_1$, or the fourth case holds and $A \cup B \subseteq V_2$. ∎

The next step is to discuss whether two points can be separated one from the other in a metric space. This turns out to be a key notion.

DEFINITION *Two elements x_1 and x_2 of a metric space X are* ***connected through*** *X iff there is a connected subset C of X such that $x_1, x_2 \in C$. In this circumstance we will write $x_1 \underset{X}{\sim} x_2$.*

THEOREM 6.8 *The relation $\underset{X}{\sim}$ is an equivalence relation on X.*

Proof. The symmetry condition is obvious because the definition of the relation is symmetric in x_1 and x_2. For the reflexivity, it suffices to take $C = \{x\}$ if $x_1 = x_2 = x$. All the work is in establishing the transitivity. Let $x_1, x_2, x_3 \in X$ and suppose that $x_1 \underset{X}{\sim} x_2$ and $x_2 \underset{X}{\sim} x_3$. Then by definition, there exist connected subsets A and B of X such that $x_1, x_2 \in A$ and $x_2, x_3 \in B$. Clearly, $x_2 \in A \cap B$. An application of Theorem 6.7 shows that $A \cup B$ is connected. Of course, $x_1, x_3 \in A \cup B$ so that $x_1 \underset{X}{\sim} x_3$. ∎

DEFINITION *Let X be a metric space. The equivalence classes of the relation $\underset{X}{\sim}$ are called* ***components****. For an element $x \in X$, the* ***component of*** *x means the equivalence class containing x. In an obvious way, the components of X are the maximal connected subsets of X.*

The following is an immediate consequence of Proposition 6.4.

PROPOSITION 6.9 *Let X be a metric space and let C be a component of X. Then C is closed in X.*

This is a good opportunity to prove the following Proposition.

PROPOSITION 6.10 *Every open subset of $\mathbb{R}$ is a countable disjoint union of open intervals.*

Proof. Let V be an open subset of $\mathbb{R}$. We consider V as a metric space in its own right. We can write V as a disjoint union of its components. Let U be a typical component of V. Then by the previous result, U is an interval. We claim that U is open. Let $x \in U$. Then $x \in V$ and, since V is open in $\mathbb{R}$, there exists $\epsilon > 0$ such that $]x \Leftrightarrow \epsilon, x + \epsilon[\subseteq V$. But $]x \Leftrightarrow \epsilon, x + \epsilon[$ is a connected set and hence must lie in the same component as x. This shows that $]x \Leftrightarrow \epsilon, x + \epsilon[\subseteq U$. Hence U is open. Finally, since each open interval must contain a rational number, select in each component a rational. Since $\mathbb{Q}$ is countably infinite, it is clear that the number of components of V is countable. ∎

How can we recognize components? In general it is not always easy. The following Lemma is sometimes useful.

LEMMA 6.11 *Let C be a nonempty subset of a metric space X which is simultaneously open, closed and connected. Then C is a component of X.*

Proof. Let C be a nonempty connected open closed subset of X. Let Y be the component of X containing C. Then, since $X = C \cup (X \setminus C)$ is a splitting of X and Y is connected we find that either $Y \subseteq C$ or $Y \subseteq X \setminus C$). Since C is nonempty, and $C \subseteq Y$ the second alternative is not possible. Hence $C = Y$ and C is a component. ∎

EXAMPLE Consider the subset $X = \{0\} \cup \{\frac{1}{n}; n \in \mathbb{N}\}$ of $\mathbb{R}$. Clearly each of the singletons $\{\frac{1}{n}\}$ is relatively open and relatively closed in X and also connected. Hence each set $\{\frac{1}{n}\}$ is a component of X. When these are removed from X we are left just with $\{0\}$ which is clearly connected. Hence $\{0\}$ is also a component. □

EXAMPLE A variation on the preceding example is

$$X = \bigcup_{j=0}^{\infty} I_j$$

where $I_0 = [-1, 0]$ and $I_j = [\frac{1}{2j+1}, \frac{1}{2j}]$ for $j \in \mathbb{N}$. Each of the intervals I_j with $j \in \mathbb{N}$ is open, closed and connected in X and hence a component. The remaining set I_0 is clearly connected and hence it too must be a component. □

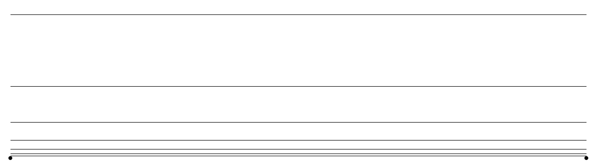

FIGURE 5: *Example of distinct components that cannot be split.*

EXAMPLE In $\mathbb{R}^2$ let $a = (-1, 0)$, $b = (1, 0)$. Let I_k be the closed line segment joining $(-1, 2^{-k})$ to $(1, 2^{-k})$. Finally, let

$$A = \{a, b\} \cup (\bigcup_{k=1}^{\infty} I_k).$$

The components of A are $\{a\}$, $\{b\}$ and I_k for $k \in \mathbb{N}$. What are the splittings of A? Suppose that one of the splitting sets A_1 contains a. Then, since A_1 is open (in A), it also contains a tail of the sequence $(-1, 2^{-k})$. But, since each I_k is connected, it also contains a tail of the I_k and in particular a tail of the sequence $(1, 2^{-k})$. Finally, since A_1 is closed (in A), $b \in A_1$. It is now not difficult to see that

$$A_2 = \bigcup_{k \in F} I_k$$

and

$$A_1 = A \setminus A_2,$$

where F is some finite subset of $\mathbb{N}$. Certainly, the points a and b lie on the same side of every splitting, despite the fact that $\{a\}$ and $\{b\}$ are distinct components. This shows that the converse of Lemma 6.3 is false. The subset $\{a\} \cup \{b\}$ of A lies on the same side of every splitting of A, but is not a connected subset of A. □

EXAMPLE Let E be the Cantor set in $\mathbb{R}$ (page 13). The components of the Cantor set are all singletons. Let a and b be distinct points of the Cantor set. Suppose without loss of generality that $a < b$. Let $\epsilon = b - a > 0$. Now select n such that $3^{-n} < \frac{1}{2}\epsilon$. The intervals of E_n have length 3^{-n} so clearly, a and b must belong to different constituent subintervals of E_n. Thus we may select $c \notin E_n$ with $a < c < b$. Now apply Proposition 6.2 with $V_1 =]-\infty, c[$ and $V_2 =]c, \infty[$. Since $a \in V_1$ and $b \in V_2$ it follows that there can be no connected subset of E that contains both a and b. □

Definition *A metric space is said to be* ***totally disconnected*** *iff every component is a singleton.*

6.4 Compactness and Connectedness

There is a subtle interplay between compactness and connectedness.

Proposition 6.12 *Let K be a compact metric space and let C be a component in K. Let $\mathcal{V}$ be the collection of all simultaneously open and closed subsets of K containing C. Then*

$$C = \bigcap_{V \in \mathcal{V}} V$$

Proof. Let us define $D = \cap_{V \in \mathcal{V}} V$. We will show that D is a connected subset of K. Certainly D is a closed subset of K because it is an intersection of closed sets. If it is not connected we can write

$$D = D_1 \cup D_2, \qquad \emptyset = D_1 \cap D_2 \tag{6.10}$$

where D_1 and D_2 are non-empty subsets of D simultaneously open and closed in D. Since D is closed in K and D_j is closed in D it follows that D_j is closed in K for $j = 1, 2$. Since D_1 and D_2 are also disjoint, it is therefore possible to separate them with sets U_1 and U_2 open in K (see Corollary 4.15). We have

$$U_1 \cap U_2 = \emptyset \qquad \text{and} \qquad D_j \subseteq U_j,\ j = 1, 2.$$

Then

$$\begin{aligned} K \setminus (U_1 \cup U_2) &\subseteq K \setminus (D_1 \cup D_2) \\ &= K \setminus D \\ &= \bigcup_{V \in \mathcal{V}} (K \setminus V) \end{aligned}$$

Since $K \setminus (U_1 \cup U_2)$ is a closed subset of K it is compact and hence we can find a finite subset $\mathcal{F} \subseteq \mathcal{V}$ such that

$$K \setminus (U_1 \cup U_2) \subseteq \bigcup_{V \in \mathcal{F}} (K \setminus V). \tag{6.11}$$

Let us define

$$W = \bigcap_{V \in \mathcal{F}} V.$$

A moment's thought convinces us that $W \in \mathcal{V}$. We can restate (6.11) as $W \subseteq U_1 \cup U_2$. Clearly $W \cap U_1 = W \cap (K \setminus U_2)$ is the intersection of two closed sets and hence closed (as well as open). Similarly, $W \cap U_2$ is simultaneously open and closed. Thus the three sets $W \cap U_1$, $W \cap U_2$ and $K \setminus W$ form a three way splitting of K. Since C is connected it lies entirely in one of the sets of the splitting. Since $C \subseteq W$ we can assume without loss of generality that $C \subseteq W \cap U_1$. But then $W \cap U_1 \in \mathcal{V}$ and it follows that $D \subseteq U_1$. It follows that in the splitting (6.10), $D_2 = \emptyset$. The contradiction shows that D is connected. Finally by the maximality of C among connected subsets of K we see that $D = C$. ■

6.5 Preservation of Connectedness by Continuous Mappings

One of the most important properties of connectedness is that it is preserved by continuous mappings.

Theorem 6.13 *Let X and Y be metric spaces. Suppose that X is connected. Let $f : X \longrightarrow Y$ be a continuous surjection. Then Y is also connected.*

Proof. Let

$$\begin{aligned} Y &= Y_1 \cup Y_2, \\ \emptyset &= Y_1 \cap Y_2, \end{aligned}$$

be a splitting of Y. Then it is easy to see that

$$\begin{aligned} X &= f^{-1}(Y_1) \cup f^{-1}(Y_2), \\ \emptyset &= f^{-1}(Y_1) \cap f^{-1}(Y_2), \end{aligned}$$

is a splitting of X. But since X is connected, the second splitting must be trivial. It follows that the first splitting is also trivial. ■

Corollary 6.14 *Let X and Y be metric spaces and let $f : X \longrightarrow Y$ be a continuous mapping. Let A be a connected subset of X. Then the direct image $f(A)$ is a connected subset of Y.*

Corollary 6.15 (Intermediate Value Theorem) *Let $f : [a, b] \longrightarrow \mathbb{R}$ be a continuous mapping. Let x be between $f(a)$ and $f(b)$. Then $x \in f([a, b])$.*

Proof. Let us suppose without loss of generality that $f(a) \le f(b)$. Then the statement that x is between $f(a)$ and $f(b)$ implies that $f(a) \le x \le f(b)$. By the previous Corollary, $f([a, b])$ is a connected subset of $\mathbb{R}$. By Theorem 6.6 (page 87), $f([a, b])$ is an interval. Since this set contains $f(a)$ and $f(b)$, it also must contain the interval $[f(a), f(b)]$. It follows that x is in the direct image of f. ■

The following definition is long overdue.

Definition *A metric space X is* ***discrete*** *iff every subset of X is open. A subset A of a metric space X is discrete iff it is discrete as a metric space with the restriction metric.*

Because of Theorem 2.2 a metric space is discrete iff its singletons are open.

Example The subset $\mathbb{Z}$ is a discrete subset of $\mathbb{R}$. Indeed, given $n \in \mathbb{Z}$ we have $\{n\} = \mathbb{Z} \cap]n - \frac{1}{2}, n + \frac{1}{2}[$ showing that $\{n\}$ is open in $\mathbb{Z}$. □

Obviously a discrete metric space is totally disconnected. The Cantor set is a space that is totally disconnected but not discrete. The following Proposition involves an important technique.

Proposition 6.16 *Let X be a connected metric space. Let Y be a discrete metric space. Let $f : X \longrightarrow Y$ be a continuous mapping. Then f is a constant mapping.*

Proof. By Theorem 6.13, the direct image $f(X)$ is connected and hence contained in a single component of Y. But the only components of Y are singletons. Hence f is a constant mapping. ■

It will be noted that the same conclusion would hold if Y we replace the hypothesis Y discrete by Y totally disconnected. However this extension is seldom used in practice.

6.6 Path Connectedness

There is a strong form of connectedness called path connectedness which is sometimes useful.

Definition *A metric space X is* ***path connected*** *if for every pair of points $x_0, x_1 \in X$ there is a path from x_0 to x_1. Such a* ***path*** *is a continuous mapping $f : [0,1] \longrightarrow X$ such that $f(0) = x_0$ and $f(1) = x_1$.*

Theorem 6.17 *Every path connected space is connected.*

Proof. Let X be a path connected metric space. We show that X has but one component. Let $x_0, x_1 \in X$, we will show that x_0 and x_1 lie in the same component of X. Let f be a path joining x_0 to x_1. Then, by Theorem 6.13, the underlying set $f([0,1])$ of the path is a connected set to which both x_0 and x_1 belong. Thus x_0 and x_1 belong to the same component of X. ■

We leave the following Lemma as an exercise for the reader. It is an easy application of the Glueing Theorem (page 29).

Lemma 6.18 *Let X be a metric space and let $x_0, x_1, x_2 \in X$. Suppose that there is a path joining x_0 and x_1. Suppose that there is a path joining x_1 and x_2. Then there is a path joining x_0 and x_2.*

Theorem 6.19 *Let V be a connected open subset of $\mathbb{R}^d$. Then V is path connected.*

Proof. If V is empty, there is nothing to show. Let $x_0 \in V$. Then, by Lemma 6.18, it is enough to show that for every $x \in V$ there is a path in V joining x_0 to x. Let W be the set of all such points x. Then by considering the constant path with value x_0 we see that $x_0 \in W$.
***Claim:* W *is open in* V** Let $x_1 \in W$. Then, since V is open, there exists $\epsilon > 0$ such that $U(x_1, \epsilon) \subseteq V$. In particular, for every point x of $U(x_1, \epsilon)$ the line segment joining x_1 to x lies in V. The function $g : [0,1] \longrightarrow V$ given by

$$g(t) = (1 - t)x_1 + tx \qquad \forall t \in [0,1]$$

is a path from x_1 to x lying entirely in V. But, since $x_1 \in W$ there is a path in V from x_0 to x_1 so by another application of Lemma 6.18, there is a path in V from x_0 to x. Hence $U(x_1, \epsilon) \subseteq W$. Thus W is open in V.
***Claim:* W *is closed in* V** Let $x_1 \in V \setminus W$. Then, repeating the previous argument, there exists $\epsilon > 0$ such that $U(x_1, \epsilon) \subseteq V$ and furthermore for every point x of $U(x_1, \epsilon)$ there is a path from x_1 to x lying entirely in V. It again follows from Lemma 6.18 that if $x \in W$ then $x_1 \in W$. But $x_1 \notin W$, so it follows that $U(x_1, \epsilon) \subseteq V \setminus W$. Hence $V \setminus W$ is open in V. It follows that W is closed in V.
Since V is connected and W is a non-empty open closed subset of V it follows that $W = V$. ■

We remark that the proof actually shows that V is connected then any two points x_0 and x of V can be joined by a path consisting of finitely many line segments — that is a ***piecewise linear*** path. Also, there is nothing special about $\mathbb{R}^d$ here, the same proof would work in any real normed vector space.

Example A standard example of a space that is connected but not path connected is the subset $A = Y \cup S$ of $\mathbb{R}^2$ where Y is the line segment

$$Y = \{(0, y); -1 \leq y \leq 1\}$$

lying in the y-axis and S is the union of the following line segments.

- From $(2^{-n}, -1)$ to $(2^{-n}, 1)$ for $n \in \mathbb{Z}^+$.
- From $(2^{-(n+1)}, -1)$ to $(2^{-n}, -1)$ for $n \in \mathbb{Z}^+$, n even.
- From $(2^{-(n+1)}, 1)$ to $(2^{-n}, 1)$ for $n \in \mathbb{Z}^+$, n odd.

The sets Y and S are shown on the left in Figure 6. It is clear that both Y and S are path connected and hence connected. Thus, the only possible non-trivial splitting of A is $A = Y \cup S$ (or its reversal) and it

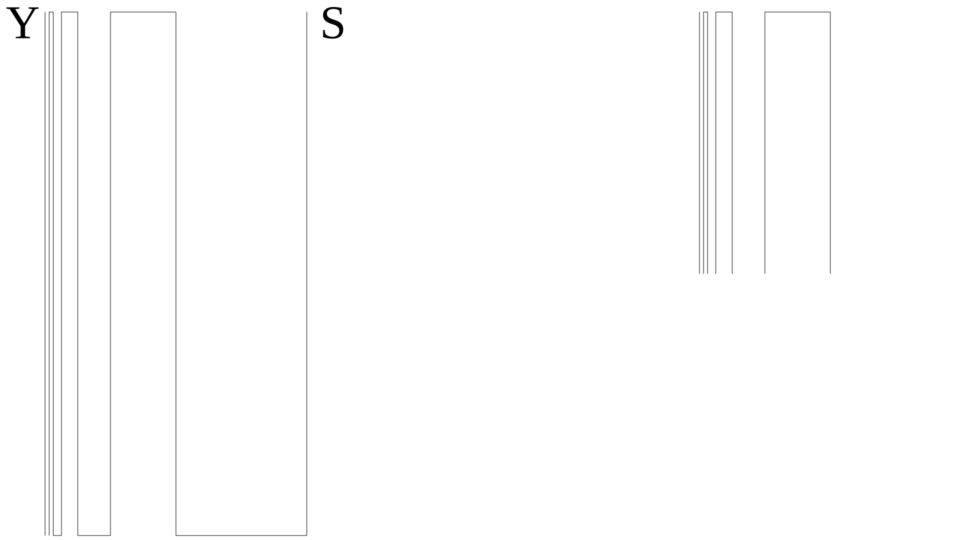

FIGURE 6: *A connected space that is not path connected (left) and a horizontal swath cut through the same space (right).*

is easy to see that S is not closed in A since for example the sequence of points $((2^{-n}, 0))$ of S converges to the point $(0, 0)$ of Y. It follows that A is connected.

To see that A is not path connected is harder. We first study the connectivity of the intersection of A with a narrow horizontal strip. Let $a < b$ and suppose that $[\Leftrightarrow 1, 1]$ is not contained in $[a, b]$. Let H be the horizontal strip

$$H = \{(x, y); a \leq y \leq b\}.$$

In case $a = 0$, $b = 1$ the set $A \cap H$ is shown on the right in Figure 6. We leave the reader to show that $Y \cap H$ is a component in $A \cap H$. The method is similar to that used in the two examples following Lemma 6.11.

Let now $f : [0, 1] \Leftrightarrow\!\!\rightarrow A$ be a path from $(0, 0)$ to $(1, 0)$ lying in A. Define

$$t = \inf_{f(s) \in S} s.$$

Informally, t is the *first time* that the path jumps from Y to S. If $t = 0$ then $f(t) \in Y$. On the other hand, if $t > 0$ then $f(s) \in Y$ for $0 \leq s < t$ and it follows by continuity of f and the fact that Y is closed that $f(t) \in Y$. Since f is continuous at t there exists $\delta > 0$ such that

$$|s \Leftrightarrow t| \leq \delta \qquad \Rightarrow \qquad d(f(s), f(t)) \leq \tfrac{1}{2}.$$

Thus taking η to be the y-coordinate of $f(t)$ and setting $a = \eta \Leftrightarrow \frac{1}{2}$ and $b = \eta + \frac{1}{2}$ we see that the restriction of f to $[t \Leftrightarrow \delta, t + \delta]$ is a continuous mapping taking values in $A \cap H$. But since $f(t) \in Y \cap H$ and $Y \cap H$ is a component of $A \cap H$ we see that $f(s) \in Y \cap H$ for all s in $[t \Leftrightarrow \delta, t + \delta]$. This contradicts the definition of t. □

6.7 Separation Theorem for Convex Sets

In this section we tackle the Separation Theorem for Convex Sets stated on page 43. At first glance, there does not seem to be much of a connection between this topic and connectedness. To show the hidden connection we start with the following proposition.

Proposition 6.20 *Let C be an open convex subset of V a finite dimensional real normed vector space of dimension at least 2. Suppose that $0_V \notin C$. Then there exists a line L through 0_V (that is a one-dimensional linear subspace L of V) such that $L \cap C = \emptyset$.*

Before tackling the proof we need the following Lemma.

Lemma 6.21 *Let V be a real normed vector space of dimension at least 2 and let $S = \{v; v \in V, \|v\| = 1\}$ be its unit ball. Then S is a connected subset of V.*

Proof. Let $v_0, v_1 \in S$. Then consider

$$v(t) = \|(1 - t)v_0 + tv_1\|^{-1}((1 - t)v_0 + tv_1)$$

for $0 \le t \le 1$. If the vector $(1 - t)v_0 + tv_1$ is non-zero for all $t \in [0, 1]$ then $t \to v(t)$ is a continuous path from v_0 to v_1 lying in S. On the other hand, if $(1 - t)v_0 + tv_1 = 0_V$, then $(1 - t)v_0 = -tv_1$, and taking the norm of both sides leads to $(1 - t) = t$ so that $t = \frac{1}{2}$ and it follows that $v_1 = -v_0$. Thus, either there is a continuous path in joining v_0 to v_1 or v_0 and v_1 are antipodal points of S.

Now suppose that v_0, v_1 and v_2 are three *distinct* points of S. Then at most one of the pairs $\{v_0, v_1\}$, $\{v_1, v_2\}$ and $\{v_2, v_0\}$ is antipodal and it follows that we can connect v_0 and v_1 by a continuous path in S either directly, or by using Lemma 6.18. Finally it is easy to see that if $\dim(V) \ge 2$ then S possesses at least 3 distinct points and the proof is complete. Of course, in case $\dim(V) = 1$, then S is a 2 point disconnected set. ■

Proof of Proposition 6.20. We define the following subset A of S.

$$A = \{u; u \in S, \exists t > 0 \text{ such that } tu \in C\}.$$

Let $u \in A$, and let $t > 0$ be such that $tu \in C$. Since C is open in V there exists $s > 0$ such that $U(tu, s) \subseteq C$. Now let $v \in S$ be such that $\|v - u\| < st^{-1}$. Then clearly $\|tv - tu\| < s$, so that $tv \in C$ and $v \in A$. We have just shown that A is relatively open in S.

Now define

$$B = \{u; u \in S, -u \in A\}.$$

Then B is relatively open in S since A is. Also $A \cap B = \emptyset$ since if there exist $t_1 > 0$ and $t_2 < 0$ such that $t_1 u, t_2 u \in C$ then it follows from the convexity of C that $0_V \in C$ which is contrary to hypothesis.

If C is empty then the result is obvious. Hence we may assume that $C \ne \emptyset$ and it follows that both A and B are nonempty. The scenario $S = A \cup B$ is now ruled out by the connectivity of S. Hence there exists $u \in S \setminus (A \cup B)$, and the line L through 0_V and u does not meet C. The point tu is not in C since

- $u \notin A$ if $t > 0$,
- $u \notin B$ if $t < 0$,
- $tu = 0_V \notin C$ if $t = 0$.

■

Definition *Let X and Y be metric spaces and let $\varphi : X \to Y$. Then φ is an* **open mapping** *iff the direct image $\varphi(\Omega)$ is an open subset of Y for every open subset Ω of X.*

Proof of Theorem 3.9. Without loss of generality, we may suppose that $v = 0_V$. For this it suffices to apply a translation.

The proof is by induction on the dimension of V. If V is one-dimensional then the result is obvious, since an open convex set in $\mathbb{R}$ is just an open interval. Thus, we may suppose that $n \ge 2$, that the result is proved for vector spaces of dimension $n \Leftrightarrow 1$ and establish it in case $\dim(V) = n$. Since $n \geq 2$ we can apply Proposition 6.20 to find a one-dimensional subspace L of V that does not meet C. Consider now the quotient vector space $Q = V/L$ and let π be the canonical projection $\pi : V \Leftrightarrow\!\rightarrow Q$.

The direct image $\pi(C)$ is clearly a convex subset of Q. To see this, let $q_1, q_2 \in \pi(C)$. Then we can find lifts $v_1, v_2 \in C$ such that $\pi(v_j) = q_j$ for $j = 1, 2$. Let t_1, t_2 be nonnegative real numbers such that $t_1 + t_2 = 1$. Then, by the convexity of C we find that $t_1 v_1 + t_2 v_2 \in C$ and it follows that $t_1 q_1 + t_2 q_2 = \pi(t_1 v_1 + t_2 v_2) \in \pi(C)$.

We next claim that $\pi(C)$ is an open subset of Q. In fact, π is an open mapping. Let Ω be an arbitrary open subset of V and let $q_0 \in \pi(\Omega)$. Then, there exist $v_0 \in V$ such that $v_0 \in \Omega$ and $\pi(v_0) = q_0$. Since Ω is open in V, there exists $t > 0$ such that $U(v_0, t) \subseteq \Omega$. Now, let $q \in U(q_0, t)$. Since $\|q \Leftrightarrow q_0\|_Q < t$, and by the definition of the quotient norm (page 46), there exist $w \in V$ such that $\|w\|_V < t$ and $\pi(w) = q \Leftrightarrow q_0$. It now follows that $v = v_0 + w \in \Omega$ and that $\pi(v) = q$. Thus we have shown that $U(q_0, t) \subseteq \pi(\Omega)$. Since q_0 was an arbitrary point of $\pi(\Omega)$, it follows that $\pi(\Omega)$ is open in Q.

Finally, since $0_Q \notin \pi(C)$ (because $L \cap C = \emptyset$), we can apply the inductive hypothesis to obtain the existence of a linear form φ on Q such that

$$\varphi(q) < \varphi(0_Q) = 0 \qquad \forall q \in \pi(C).$$

It follows immediately that

$$\varphi \circ \pi(v) < 0 = \varphi \circ \pi(0_V) \qquad \forall v \in C.$$

Since $\varphi \circ \pi$ is a linear form on V, this completes the inductive step. ∎