Course 212: Michaelmas Term 2000
Part I: Limits and Continuity, Open and Closed Sets, Metric Spaces

David R. Wilkins

Contents

1 Review of Real Analysis

1.1 The Real Number System

The system of real numbers, denoted by $\mathbb{R}$, is an ordered set on which are defined appropriate operations of addition and multiplication. The system of real numbers is fully characterized by an axiom system consisting of the 15 axioms (listed below) which describe the algebraic structure and ordering of the real numbers, together with one further axiom, known as the *Least Upper Bound Axiom*, which distinguishes the real number system from other number systems such as the rational number system. The 15 axioms describing the algebraic and ordering properties of the real number system are as follows:

1. if x and y are real numbers then their sum $x + y$ is also a real number,

2. (*the Commutative Law for addition*) $x + y = y + x$ for all real numbers x and y,

3. (*the Associative Law for addition*) $(x + y) + z = x + (y + z)$ for all real numbers x, y and z,

4. there exists a (necessarily unique) real number, denoted by 0, with the property that $x + 0 = x = 0 + x$ for all real numbers x,

5. for each real numbers x there exists some (necessarily unique) real number $-x$ with the property that $x + (-x) = 0 = (-x) + x$,

6. if x and y are real numbers then their product xy is also a real number,

7. (*the Commutative Law for multiplication*) $xy = yx$ for all real numbers x and y,

8. (*the Associative Law for multiplication*) $(xy)z = x(yz)$ for all real numbers x, y and z,

9. there exists a (necessarily unique) real number, denoted by 1, with the property that $x1 = x = 1x$ for all real numbers x, and moreover $1 \neq 0$,

10. for each real numbers x satisfying $x \neq 0$ there exists some (necessarily unique) real number x^{-1} with the property that $xx^{-1} = 1 = x^{-1}x$,

11. (*the Distributive Law*) $x(y + z) = (xy) + (xz)$ for all real numbers x y and z,

12. (*the Trichotomy Law*) if x and y are real numbers then one and only one of the three statements $x < y$, $x = y$ and $y < x$ is true,

13. if x, y and z are real numbers and if $x < y$ and $y < z$ then $x < z$,

14. if x, y and z are real numbers and if $x < y$ then $x + z < y + z$,

15. if x and y are real numbers which satisfy $0 < x$ and $0 < y$ then $0 < xy$,

The operations of subtraction and division are defined in terms of addition and multiplication in the obvious fashion: $x - y = x + (-y)$ for all real numbers x and y, and $x/y = xy^{-1}$ provided that $y \neq 0$. The *absolute value* $|x|$ of a real number x is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0; \\ -x & \text{if } x < 0. \end{cases}$$

Note that $|x| \geq 0$ for all real numbers x and that $|x| = 0$ if and only if $x = 0$. Also $|x + y| \leq |x| + |y|$ and $|xy| = |x||y|$ for all real numbers x and y.

Let D be a subset of $\mathbb{R}$. A real number u is said to be an *upper bound* of the set D if $x \leq u$ for all $x \in D$. The set D is said to be *bounded above* if such an upper bound exists.

Definition Let D be some set of real numbers which is bounded above. A real number s is said to be the *least upper bound* (or *supremum*) of D (denoted by $\sup D$) if s is an upper bound of D and $s \leq u$ for all upper bounds u of D.

Example Indeed the real number 2 is the least upper bound of the sets $\{x \in \mathbb{R} : x \leq 2\}$ and $\{x \in \mathbb{R} : x < 2\}$. Note that the first of these sets contains its least upper bound, whereas the second set does not.

The axioms (1)–(15) listed above describing the algebraic and ordering properties of the real number system are not in themselves sufficient to fully characterize the real number system. (Indeed any property of real numbers that could be derived solely from these axioms would be equally valid for rational numbers.) We require in addition the following axiom:—

> *the Least Upper Bound Axiom*: if D is any non-empty subset of $\mathbb{R}$ which is bounded above then there exists a *least upper bound* $\sup D$ for the set D.

A *lower bound* of a set D of real numbers is a real number l with the property that $l \leq x$ for all $x \in D$. A set D of real numbers is said to be *bounded below* if such a lower bound exists. If D is bounded below, then there exists a greatest lower bound (or *infimum*) $\inf D$ of the set D. Indeed $\inf D = -\sup\{x \in \mathbb{R} : -x \in D\}$.

Remark We have simply listed above a complete set of axioms for the real number system. We have not however proved the existence of a system of real numbers satisfying these axioms. There are in fact several constructions of the real number system: one of the most popular of these is the representation of real numbers as *Dedekind sections* of the set of rational numbers. For an account of the this construction, and for a proof that these axioms are sufficient to characterize the real number system, see chapters 27–29 of *Calculus*, by M. Spivak.

1.2 Infinite Sequences of Real Numbers

An *infinite sequence* of real numbers is a sequence of the form $a_1, a_2, a_3, \ldots$, where each a_n is a real number. (More formally, one can view an infinite sequence of real numbers as a function from $\mathbb{N}$ to $\mathbb{R}$ which sends each natural number n to some real number a_n.)

Definition A sequence $a_1, a_2, a_3, \ldots$ of real numbers is said to *converge* to some real number l if and only if the following criterion is satisfied:

> given any real number ε satisfying $\varepsilon > 0$, there exists some natural number N such that $|a_n - l| < \varepsilon$ for all n satisfying $n \geq N$.

If the sequence $a_1, a_2, a_3, \ldots$ converges to the *limit* l then we denote this fact by writing '$a_n \to l$ as $n \to +\infty$', or by writing '$\lim_{n \to +\infty} a_n = l$'.

Example A straightforward application of the definition of convergence shows that $1/n \to 0$ as $n \to +\infty$. Indeed suppose that we are given any real number ε satisfying $\varepsilon > 0$. If we pick some natural number N large enough to satisfy $N > 1/\varepsilon$ then $|1/n| < \varepsilon$ for all natural numbers n satisfying $n \geq N$, as required.

Example We show that $(-1)^n/n^2 \to 0$ as $n \to +\infty$. Indeed, given any real number ε number satisfying $\varepsilon > 0$, we can find some natural number N satisfying $N^2 > 1/\varepsilon$. If $n \geq N$ then $|(-1)^n/n^2| < \varepsilon$, as required.

Example The infinite sequence $a_1, a_2, a_3, \ldots$ defined by $a_n = n$ is not convergent. To prove this formally, we suppose that it were the case that $\lim_{n \to +\infty} a_n = l$ for some real number l, and derive from this a contradiction. On setting $\varepsilon = 1$ (say) in the formal definition of convergence, we would deduce that there would exist some natural number N such that $|a_n - l| < 1$ for all $n \geq N$. But then $a_n < l + 1$ for all $n \geq N$, which is impossible. Thus the sequence cannot converge.

Example The infinite sequence $u_1, u_2, u_3, \ldots$ defined by $u_n = (-1)^n$ is not convergent. To prove this formally, we suppose that it were the case that $\lim_{n \to +\infty} u_n = l$ for some real number l. On setting $\varepsilon = \frac{1}{2}$ in the criterion for convergence, we would deduce the existence of some natural number N such that $|u_n - l| < \frac{1}{2}$ for all $n \geq N$. But then

$$|u_n - u_{n+1}| \leq |u_n - l| + |l - u_{n+1}| < \tfrac{1}{2} + \tfrac{1}{2} = 1$$

for all $n \geq N$, contradicting the fact that $u_n - u_{n+1} = \pm 2$ for all n. Thus the sequence cannot converge.

Definition We say that an infinite sequence $a_1, a_2, a_3, \ldots$ of real numbers is *bounded above* if there exists some real number B such that $a_n \leq B$ for all n. Similarly we say that this sequence is *bounded below* if there exists some real number A such that $a_n \geq A$ for all n. A sequence is said to be *bounded* if it is bounded above and bounded below, so that there exist real numbers A and B such that $A \leq a_n \leq B$ for all n.

Lemma 1.1 *Every convergent sequence of real numbers is bounded.*

Proof Let $a_1, a_2, a_3, \ldots$ be a sequence of real numbers converging to some real number l. On applying the formal definition of convergence (with $\varepsilon = 1$), we deduce the existence of some natural number N such that $|a_n - l| < 1$ for all $n \geq N$. But then $A \leq a_n \leq B$ for all n, where A is the minimum of $a_1, a_2, \ldots, a_{N-1}$ and $l - 1$, and B is the maximum of $a_1, a_2, \ldots, a_{N-1}$ and $l + 1$. ∎

Proposition 1.2 *Let $a_1, a_2, a_3, \ldots$ and $b_1, b_2, b_3, \ldots$ be convergent infinite sequences of real numbers. Then the sum, difference and product of these sequences are convergent, and*

$$\begin{aligned}
\lim_{n \to +\infty} (a_n + b_n) &= \lim_{n \to +\infty} a_n + \lim_{n \to +\infty} b_n, \\
\lim_{n \to +\infty} (a_n - b_n) &= \lim_{n \to +\infty} a_n - \lim_{n \to +\infty} b_n, \\
\lim_{n \to +\infty} (a_n b_n) &= \left(\lim_{n \to +\infty} a_n \right) \left(\lim_{n \to +\infty} b_n \right).
\end{aligned}$$

If in addition $b_n \neq 0$ for all n and $\lim_{n\to+\infty} b_n \neq 0$, then the quotient of the sequences (a_n) and (b_n) is convergent, and

$$\lim_{n\to+\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to+\infty} a_n}{\lim_{n\to+\infty} b_n}.$$

Proof Throughout this proof let $l = \lim_{n\to+\infty} a_n$ and $m = \lim_{n\to+\infty} b_n$.

First we prove that $a_n + b_n \to l + m$ as $n \to +\infty$. Let ε be any given real number satisfying $\varepsilon > 0$. We must show that there exists some natural number N such that $|a_n + b_n - (l + m)| < \varepsilon$ whenever $n \geq N$. Now $a_n \to l$ as $n \to +\infty$, and therefore, given any $\varepsilon_1 > 0$, there exists some natural number N_1 with the property that $|a_n - l| < \varepsilon_1$ whenever $n \geq N_1$. In particular, there exists a natural number N_1 with the property that $|a_n - l| < \frac{1}{2}\varepsilon$ whenever $n \geq N_1$. (To see this, let $\varepsilon_1 = \frac{1}{2}\varepsilon$.) Similarly there exists some natural number N_2 such that $|b_n - m| < \frac{1}{2}\varepsilon$ whenever $n \geq N_2$. Let N be the maximum of N_1 and N_2. If $n \geq N$ then

$$\begin{aligned} |a_n + b_n - (l + m)| &= |(a_n - l) + (b_n - m)| \leq |a_n - l| + |b_n - m| \\ &< \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon. \end{aligned}$$

Thus $a_n + b_n \to l + m$ as $n \to +\infty$.

Let c be some real number. We show that $cb_n \to cm$ as $n \to +\infty$. The case when $c = 0$ is trivial. Suppose that $c \neq 0$. Let $\varepsilon > 0$ be given. Then there exists some natural number N such that $|b_n - m| < \varepsilon/|c|$ whenever $n \geq N$. But then $|cb_n - cm| = |c||b_n - m| < \varepsilon$ whenever $n \geq N$. Thus $cb_n \to cm$ as $n \to +\infty$.

If we combine this result, for $c = -1$, with the previous result, we see that $-b_n \to -m$ as $n \to +\infty$, and therefore $a_n - b_n \to l - m$ as $n \to +\infty$.

Next we show that if $u_1, u_2, u_3, \ldots$ and $v_1, v_2, v_3, \ldots$ are infinite sequences, and if $u_n \to 0$ and $v_n \to 0$ as $n \to +\infty$, then $u_n v_n \to 0$ as $n \to +\infty$. Let $\varepsilon > 0$ be given. Then there exist natural numbers N_1 and N_2 such that $|u_n| < \sqrt{\varepsilon}$ whenever $n \geq N_1$ and $|v_n| < \sqrt{\varepsilon}$ whenever $n \geq N_2$. Let N be the maximum of N_1 and N_2. If $n \geq N$ then $|u_n v_n| < \varepsilon$. We deduce that $u_n v_n \to 0$ as $n \to +\infty$.

We can apply this result with $u_n = a_n - l$ and $v_n = b_n - m$ for all natural numbers n. Using the results we have already obtained, we see that

$$\begin{aligned} 0 &= \lim_{n\to+\infty} (u_n v_n) = \lim_{n\to+\infty} (a_n b_n - a_n m - l b_n + lm) \\ &= \lim_{n\to+\infty} (a_n b_n) - m \lim_{n\to+\infty} a_n - l \lim_{n\to+\infty} b_n + lm = \lim_{n\to+\infty} (a_n b_n) - lm. \end{aligned}$$

Thus $a_n b_n \to lm$ as $n \to +\infty$.

Next we show that if $w_1, w_2, w_3, \ldots$ is an infinite sequence of non-zero real numbers, and if $w_n \to 1$ as $n \to +\infty$ then $1/w_n \to 1$ as $n \to +\infty$. Let $\varepsilon > 0$ be given. Let ε_0 be the minimum of $\frac{1}{2}\varepsilon$ and $\frac{1}{2}$. Then there exists some natural number N such that $|w_n - 1| < \varepsilon_0$ whenever $n \geq N$. Thus if $n \geq N$ then $|w_n - 1| < \frac{1}{2}\varepsilon$ and $\frac{1}{2} < w_n < \frac{3}{2}$. But then

$$\left| \frac{1}{w_n} - 1 \right| = \left| \frac{1 - w_n}{w_n} \right| = \frac{|w_n - 1|}{|w_n|} \leq 2|w_n - 1| < \varepsilon.$$

We deduce that $1/w_n \to 1$ as $n \to +\infty$. If we apply this result with $w_n = b_n/m$, where $m \neq 0$, we deduce that $m/b_n \to 1$, and thus $1/b_n \to 1/m$ as $n \to +\infty$. The result we have already obtained for products of sequences then enables us to deduce that $a_n/b_n \to l/m$ as $n \to +\infty$. ∎

Example On using Proposition 1.2, we see that

$$\lim_{n \to +\infty} \frac{6n^2 - 4n}{3n^2 + 7} = \lim_{n \to +\infty} \frac{6 - (4/n)}{3 + (7/n^2)} = \frac{\lim\limits_{n \to +\infty} (6 - (4/n))}{\lim\limits_{n \to +\infty} (3 + (7/n^2))} = \frac{6}{3} = 2.$$

1.3 Monotonic Sequences

An infinite sequence $a_1, a_2, a_3, \ldots$ of real numbers is said to be *strictly increasing* if $a_{n+1} > a_n$ for all n, *strictly decreasing* if $a_{n+1} < a_n$ for all n, *non-decreasing* if $a_{n+1} \geq a_n$ for all n, or *non-increasing* if $a_{n+1} \leq a_n$ for all n. A sequence satisfying any one of these conditions is said to be *monotonic*; thus a monotonic sequence is either non-decreasing or non-increasing.

Theorem 1.3 *A non-decreasing sequence of real numbers that is bounded above is convergent. Similarly a non-increasing sequence of real numbers that is bounded below is convergent.*

Proof Let $a_1, a_2, a_3, \ldots$ be a non-decreasing sequence of real numbers that is bounded above. It follows from the Least Upper Bound Axiom that there exists a least upper bound l for the set $\{a_n : n \in \mathbb{N}\}$. We claim that the sequence converges to l.

Let $\varepsilon > 0$ be given. We must show that there exists some natural number N such that $|a_n - l| < \varepsilon$ whenever $n \geq N$. Now $l - \varepsilon$ is not an upper bound for the set $\{a_n : n \in \mathbb{N}\}$ (since l is the least upper bound), and therefore there must exist some natural number N such that $a_N > l - \varepsilon$. But then $l - \varepsilon < a_n \leq l$ whenever $n \geq N$, since the sequence is non-decreasing and bounded above by l. Thus $|a_n - l| < \varepsilon$ whenever $n \geq N$. Therefore $a_n \to l$ as $n \to +\infty$, as required.

If the sequence $a_1, a_2, a_3, \ldots$ is non-increasing and bounded below then the sequence $-a_1, -a_2, -a_3, \ldots$ is non-decreasing and bounded above, and is therefore convergent. It follows that the sequence $a_1, a_2, a_3, \ldots$ is also convergent. ∎

Example Let $a_1 = 2$ and

$$a_{n+1} = a_n - \frac{a_n^2 - 2}{2a_n}$$

for all natural numbers n. Now

$$a_{n+1} = \frac{a_n^2 + 2}{2a_n} \text{ and } a_{n+1}^2 = a_n^2 - (a_n^2 - 2) + \left(\frac{a_n^2 - 2}{2a_n}\right)^2 = 2 + \left(\frac{a_n^2 - 2}{2a_n}\right)^2.$$

It therefore follows by induction on n that $a_n > 0$ and $a_n^2 > 2$ for all natural numbers n. But then $a_{n+1} < a_n$ for all n, and thus the sequence $a_1, a_2, a_3, \ldots$ is decreasing and bounded below. It follows from Theorem 1.3 that this sequence converges to some real number α. Also $a_n > 1$ for all n (since $a_n > 0$ and $a_n^2 > 2$), and therefore $\alpha \geq 1$. But then, on applying Proposition 1.2, we see that

$$\alpha = \lim_{n \to +\infty} a_{n+1} = \lim_{n \to +\infty} \left(a_n - \frac{a_n^2 - 2}{2a_n}\right) = \alpha - \frac{\alpha^2 - 2}{2\alpha}.$$

Thus $\alpha^2 = 2$, and so $\alpha = \sqrt{2}$.

1.4 Subsequences and the Bolzano-Weierstrass Theorem

Let $a_1, a_2, a_3, \ldots$ be an infinite sequence of real numbers. A *subsequence* of this sequence is a sequence of the form $a_{n_1}, a_{n_2}, a_{n_3}, \ldots$, where $n_1, n_2, n_3, \ldots$ are natural numbers satisfying $n_1 < n_2 < n_3 < \cdots$. Thus, for example, $a_2, a_4, a_6, \ldots$ and $a_1, a_4, a_9, \ldots$ are subsequences of the given sequence.

Theorem 1.4 (Bolzano-Weierstrass) *Every bounded sequence of real numbers has a convergent subsequence.*

Proof Let $a_1, a_2, a_3, \ldots$ be a bounded sequence of real numbers, and let

$$S = \{n \in \mathbb{N} : a_n \geq a_k \text{ for all } k \geq n\}$$

(i.e., S is the set of all natural numbers n with the property that a_n is greater than or equal to all the succeeding members of the sequence).

First let us suppose that the set S is infinite. Arrange the elements of S in increasing order so that $S = \{n_1, n_2, n_3, n_4, \ldots\}$, where $n_1 < n_2 < n_3 < n_4 < \cdots$. It follows from the manner in which the set S was defined that $a_{n_1} \geq a_{n_2} \geq a_{n_3} \geq a_{n_4} \geq \cdots$. Thus $a_{n_1}, a_{n_2}, a_{n_3}, \ldots$ is a non-increasing subsequence of the original sequence $a_1, a_2, a_3, \ldots$. This subsequence is bounded below (since the original sequence is bounded). It follows from Theorem 1.3 that $a_{n_1}, a_{n_2}, a_{n_3}, \ldots$ is a convergent subsequence of the original sequence.

Now suppose that the set S is finite. Choose a natural number n_1 which is greater than every natural number belonging to S. Then n_1 does not belong to S. Therefore there must exist some natural number n_2 satisfying $n_2 > n_1$ such that $a_{n_2} > a_{n_1}$. Moreover n_2 does not belong to S (since n_2 is greater than n_1 and n_1 is greater than every natural number belonging to S). Therefore there must exist some natural number n_3 satisfying $n_3 > n_2$ such that $a_{n_3} > a_{n_2}$. We can continue in this way to construct (by induction on j) a strictly increasing subsequence $a_{n_1}, a_{n_2}, a_{n_3}, \ldots$ of our original sequence. This increasing subsequence is bounded above (since the original sequence is bounded) and thus is convergent, by Theorem 1.3. This completes the proof of the Bolzano-Weierstrass Theorem. ∎

1.5 Cauchy's Criterion for Convergence

Definition A sequence $x_1, x_2, x_3, \ldots$ of real numbers is said to be a *Cauchy sequence* if the following condition is satisfied:

> for every real number ε satisfying $\varepsilon > 0$ there exists some natural number N such that $|x_m - x_n| < \varepsilon$ for all natural numbers m and n satisfying $m \geq N$ and $n \geq N$.

Lemma 1.5 *Every Cauchy sequence of real numbers is bounded.*

Proof Let $x_1, x_2, x_3, \ldots$ be a Cauchy sequence. Then there exists some natural number N such that $|x_n - x_m| < 1$ whenever $m \geq N$ and $n \geq N$. In particular, $|x_n| \leq |x_N| + 1$ whenever $n \geq N$. Therefore $|x_n| \leq R$ for all n, where R is the maximum of the real numbers $|x_1|, |x_2|, \ldots, |x_{N-1}|$ and $|x_N| + 1$. Thus the sequence is bounded, as required. ∎

The following important result is known as *Cauchy's Criterion for convergence*, or as the *General Principle of Convergence.*

Theorem 1.6 (Cauchy's Criterion for Convergence) *An infinite sequence of real numbers is convergent if and only if it is a Cauchy sequence.*

Proof First we show that convergent sequences are Cauchy sequences. Let $x_1, x_2, x_3, \ldots$ be a convergent sequence, and let $l = \lim_{n \to +\infty} x_j$. Let $\varepsilon > 0$ be given. Then there exists some natural number N such that $|x_n - l| < \frac{1}{2}\varepsilon$ for all $n \geq N$. Thus if $m \geq N$ and $n \geq N$ then $|x_m - l| < \frac{1}{2}\varepsilon$ and $|x_n - l| < \frac{1}{2}\varepsilon$, and hence

$$|x_m - x_n| = |(x_m - l) - (x_n - l)| \leq |x_m - l| + |x_n - l| < \varepsilon.$$

Thus the sequence $x_1, x_2, x_3, \ldots$ is a Cauchy sequence.

Conversely we must show that any Cauchy sequence $x_1, x_2, x_3, \ldots$ is convergent. Now Cauchy sequences are bounded, by Lemma 1.5. The sequence $x_1, x_2, x_3, \ldots$ therefore has a convergent subsequence $x_{n_1}, x_{n_2}, x_{n_3}, \ldots$, by the Bolzano-Weierstrass Theorem (Theorem 1.4). Let $l = \lim_{j \to +\infty} x_{n_j}$. We claim that the sequence $x_1, x_2, x_3, \ldots$ itself converges to l.

Let $\varepsilon > 0$ be given. Then there exists some natural number N such that $|x_n - x_m| < \frac{1}{2}\varepsilon$ whenever $m \geq N$ and $n \geq N$ (since the sequence is a Cauchy sequence). Let j be chosen large enough to ensure that $n_j \geq N$ and $|x_{n_j} - l| < \frac{1}{2}\varepsilon$. Then

$$|x_n - l| \leq |x_n - x_{n_j}| + |x_{n_j} - l| < \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon$$

whenever $n \geq N$, and thus $x_n \to l$ as $n \to +\infty$, as required. ∎

1.6 Limits of Functions of a Real Variable

Definition Let D be a subset of the set $\mathbb{R}$ of real numbers, and let s be a real number (which may or may not belong to D). We say that s is a *limit point* of D if, given any $\delta > 0$, there exists some $x \in D$ which satisfies $0 < |x - s| < \delta$.

We now define the *limit* of a real-valued function at any limit point of the domain of that function.

Definition Let $f: D \to \mathbb{R}$ be a real-valued function defined over some subset D of $\mathbb{R}$, and let s be a limit point of D. A real number l is said to be the *limit* of the function f as x tends to s in D if, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that $|f(x) - l| < \varepsilon$ for all $x \in D$ satisfying $0 < |x - s| < \delta$.

If l is the limit of $f(x)$ as x tends to s, for some s, then we denote this fact either by writing '$f(x) \to l$ as $x \to s$' or by writing '$\lim_{x \to s} f(x) = l$'.

Note that $\lim_{x \to s} f(x) = l$ if and only if $\lim_{h \to 0} f(s + h) = l$: this follows directly from the definition given above.

Lemma 1.7 *Let $f\colon D \to \mathbb{R}$ be a real-valued function defined over some subset D of $\mathbb{R}$, and let s be a limit point of D. Then the limit $\lim_{x\to s} f(x)$, if it exists, is unique.*

Proof Suppose that $\lim_{x\to s} f(x) = l$ and $\lim_{x\to s} f(x) = m$. We must show that $l = m$. Let $\varepsilon > 0$ be given. Then there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that $|f(x) - l| < \varepsilon$ whenever $x \in D$ satisfies $0 < |x - s| < \delta_1$ and $|f(x) - m| < \varepsilon$ whenever $x \in D$ satisfies $0 < |x - s| < \delta_2$. Choose $x \in D$ satisfying $0 < |x - s| < \delta$, where δ is the minimum of δ_1 and δ_2. (This is possible since s is a limit point of D.) Then $|f(x) - l| < \varepsilon$ and $|f(x) - m| < \varepsilon$, and hence

$$|l - m| \leq |l - f(x)| + |f(x) - m| < 2\varepsilon$$

by the Triangle Inequality. Since $|l - m| < 2\varepsilon$ for all $\varepsilon > 0$, we conclude that $l = m$, as required. ∎

Example We show that $\lim_{x\to 0} \frac{1}{4}x^2 = 0$. Let $\varepsilon > 0$ be given. Suppose that we choose $\delta = 2\sqrt{\varepsilon}$, for example. If $0 < |x| < \delta$ then $|\frac{1}{4}x^2| < \frac{1}{4}\delta^2 = \varepsilon$, as required.

Example We show that $\lim_{x\to 0} 3x\cos(1/x) = 0$. Let $\varepsilon > 0$ be given. Choose $\delta = \frac{1}{3}\varepsilon$. Then $\delta > 0$. Moreover if $0 < |x| < \delta$ then $|3x\cos(1/x)| < \varepsilon$, as required.

Proposition 1.8 *Let $f\colon D \to \mathbb{R}$ and $g\colon D \to \mathbb{R}$ be functions defined over some subset D of $\mathbb{R}$. Let s be a limit point of D. Suppose that $\lim_{x\to s} f(x)$ and $\lim_{x\to s} g(x)$ exist. Then $\lim_{x\to s} (f(x) + g(x))$, $\lim_{x\to s} (f(x) - g(x))$ and $\lim_{x\to s} (f(x)g(x))$ exist, and*

$$\begin{aligned}
\lim_{x\to s} (f(x) + g(x)) &= \lim_{x\to s} f(x) + \lim_{x\to s} g(x), \\
\lim_{x\to s} (f(x) - g(x)) &= \lim_{x\to s} f(x) - \lim_{x\to s} g(x), \\
\lim_{x\to s} (f(x)g(x)) &= \lim_{x\to s} f(x) \lim_{x\to s} g(x).
\end{aligned}$$

If in addition $g(x) \neq 0$ for all $x \in D$ and $\lim_{x\to s} g(x) \neq 0$, then $\lim_{x\to s} f(x)/g(x)$ exists, and

$$\lim_{x\to s} \frac{f(x)}{g(x)} = \frac{\lim_{x\to s} f(x)}{\lim_{x\to s} g(x)}.$$

Proof Let $l = \lim_{x \to s} f(x)$ and $m = \lim_{x \to s} g(x)$.

First we prove that $\lim_{x \to s} f(x) + g(x) = l + m$. Let $\varepsilon > 0$ be given. We must prove that there exists some $\delta > 0$ such that $|f(x) + g(x) - (l + m)| < \varepsilon$ for all $x \in D$ satisfying $0 < |x - s| < \delta$. Now there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that $|f(x) - l| < \frac{1}{2}\varepsilon$ for all $x \in D$ satisfying $0 < |x - s| < \delta_1$, and $|g(x) - m| < \frac{1}{2}\varepsilon$ for all $x \in D$ satisfying $0 < |x - s| < \delta_2$, since $l = \lim_{x \to s} f(x)$ and $m = \lim_{x \to s} g(x)$. Let δ be the minimum of δ_1 and δ_2. If $x \in D$ satisfies $0 < |x - s| < \delta$ then $|f(x) - l| < \frac{1}{2}\varepsilon$ and $|g(x) - m| < \frac{1}{2}\varepsilon$, and hence

$$|f(x) + g(x) - (l + m)| \leq |f(x) - l| + |g(x) - m| < \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon$$

This shows that $\lim_{x \to s} (f(x) + g(x)) = l + m$.

Let c be some real number. We show that $\lim_{x \to s} (cg(x)) = cm$. The case when $c = 0$ is trivial. Suppose that $c \neq 0$. Let $\varepsilon > 0$ be given. Then there exists some $\delta > 0$ such that $|g(x) - m| < \varepsilon/|c|$ whenever $0 < |x - s| < \delta$. But then $|cg(x) - cm| = |c||g(x) - m| < \varepsilon$ whenever $0 < |x - s| < \delta$. Thus $\lim_{x \to s} (cg(x)) = cm$.

If we combine this result, for $c = -1$, with the previous result, we see that $\lim_{x \to s} (-g(x)) = -m$, and therefore $\lim_{x \to s} (f(x) - g(x)) = l - m$.

Next we show that if $p\colon D \to \mathbb{R}$ and $q\colon D \to \mathbb{R}$ are functions with the property that $\lim_{x \to s} p(x) = \lim_{x \to s} q(x) = 0$, then $\lim_{x \to s} (p(x)q(x)) = 0$. Let $\varepsilon > 0$ be given. Then there exist real numbers $\delta_1 > 0$ and $\delta_2 > 0$ such that $|p(x)| < \sqrt{\varepsilon}$ whenever $0 < |x - s| < \delta_1$ and $|q(x)| < \sqrt{\varepsilon}$ whenever $0 < |x - s| < \delta_2$. Let δ be the minimum of δ_1 and δ_2. If $0 < |x - s| < \delta$ then $|p(x)q(x)| < \varepsilon$. We deduce that $\lim_{x \to s} (p(x)q(x)) = 0$.

We can apply this result with $p(x) = f(x) - l$ and $q(x) = g(x) - m$ for all $x \in D$. Using the results we have already obtained, we see that

$$\begin{aligned} 0 &= \lim_{x \to s}(p(x)q(x)) = \lim_{x \to s}(f(x)g(x) - f(x)m - lg(x) + lm) \\ &= \lim_{x \to s}(f(x)g(x)) - m \lim_{x \to s} f(x) - l \lim_{x \to s} g(x) + lm = \lim_{x \to s}(f(x)g(x)) - lm. \end{aligned}$$

Thus $\lim_{x \to s} (f(x)g(x)) = lm$.

Next we show that if $h\colon D \to R$ is a function that is non-zero throughout D, and if $\lim_{x \to s} h(x) \to 1$ then $\lim_{x \to s} (1/h(x)) = 1$. Let $\varepsilon > 0$ be given. Let ε_0 be the minimum of $\frac{1}{2}\varepsilon$ and $\frac{1}{2}$. Then there exists some $\delta > 0$ such that $|h(x) - 1| < \varepsilon_0$ whenever $0 < |x - s| < \delta$. Thus if $0 < |x - s| < \delta$ then $|h(x) - 1| < \frac{1}{2}\varepsilon$ and $\frac{1}{2} < h(x) < \frac{3}{2}$. But then

$$\left| \frac{1}{h(x)} - 1 \right| = \left| \frac{h(x) - 1}{h(x)} \right| = \frac{|h(x) - 1|}{|h(x)|} < 2|h(x) - 1| < \varepsilon.$$

We deduce that $\lim_{x \to s} 1/h(x) = 1$. If we apply this result with $h(x) = g(x)/m$, where $m \neq 0$, we deduce that $\lim_{x \to s} m/g(x) = 1$, and thus $\lim_{x \to s} 1/g(x) = 1/m$. The result we have already obtained for products of functions then enables us to deduce that $\lim_{x \to s} (f(x)/g(x)) \to l/m$.

1.7 Continuous Functions of a Real Variable.

Definition Let D be a subset of $\mathbb{R}$, and let $f\colon D \to \mathbb{R}$ be a real-valued function on D. Let s be a point of D. The function f is said to be *continuous* at s if, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that $|f(x) - f(s)| < \varepsilon$ for all $x \in D$ satisfying $|x - s| < \delta$. If f is continuous at every point of D then we say that f is continuous on D.

Example Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x > 0; \\ 0 & \text{if } x \leq 0. \end{cases}$$

The function f is not continuous at 0. To prove this formally we note that when $0 < \varepsilon \leq 1$ there does not exist any $\delta > 0$ with the property that $|f(x) - f(0)| < \varepsilon$ for all x satisfying $|x| < \delta$ (since $|f(x) - f(0)| = 1$ for all $x > 0$).

Example Let $f\colon \mathbb{R} \to \mathbb{R}$ be the function defined by

$$f(x) = \begin{cases} \sin \dfrac{1}{x} & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$$

We show that this function is not continuous at 0. Suppose that ε is chosen to satisfy $0 < \varepsilon < 1$. No matter how small we choose δ, where $\delta > 0$, we can always find $x \in \mathbb{R}$ for which $|x| < \delta$ and $|f(x) - f(0)| \geq \varepsilon$. Indeed, given any $\delta > 0$, we can choose some integer n large enough to ensure that $0 < x_n < \delta$, where x_n satisfies $1/x_n = (4n+1)\pi/2$. Moreover $f(x_n) = 1$. This shows that the criterion defining the concept of continuity is not satisfied at $x = 0$.

Example Let $f\colon \mathbb{R} \to \mathbb{R}$ be the function defined by

$$f(x) = \begin{cases} 3x \sin \dfrac{1}{x} & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$$

We claim that the function f is continuous at 0. To prove this, we must apply the definition of continuity directly. Suppose we are given any real number ε satisfying $\varepsilon > 0$. If $\delta = \frac{1}{3}\varepsilon$ then $|f(x)| \leq 3|x| < \varepsilon$ for all real numbers x satisfying $|x| < \delta$, as required.

The following lemma describes the relationship between limits and continuity.

Lemma 1.9 *Let D be a subset of $\mathbb{R}$, and let $s \in D$.*

(i) *Suppose that s is a limit point of D. Then a function $f: D \to \mathbb{R}$ with domain D is continuous at s if and only if $\lim_{x \to s} f(x) = f(s)$;*

(ii) *Suppose that s is not a limit point of D. Then every function $f: D \to \mathbb{R}$ with domain D is continuous at s.*

Proof If s is a limit point of D belonging to D then the required result follows immediately on comparing the formal definition of the limit of a function with the formal definition of continuity (since the condition $|f(x) - f(s)| < \varepsilon$ is automatically satisfied for any $\varepsilon > 0$ when $x = s$).

Suppose that s is not a limit point of D. Then there exists some $\delta > 0$ such that the only element x of D satisfying $|x - s| < \delta$ is s itself. The definition of continuity is therefore satisfied trivially at s by any function $f: D \to \mathbb{R}$ with domain D. ∎

Given functions $f: D \to \mathbb{R}$ and $g: D \to \mathbb{R}$ defined over some subset D of $\mathbb{R}$, we denote by $f + g$, $f - g$, $f \cdot g$ and $|f|$ the functions on D defined by

$$(f + g)(x) = f(x) + g(x), \qquad (f - g)(x) = f(x) - g(x),$$

$$(f \cdot g)(x) = f(x)g(x), \qquad |f|(x) = |f(x)|.$$

Proposition 1.10 *Let $f: D \to \mathbb{R}$ and $g: D \to \mathbb{R}$ be functions defined over some subset D of $\mathbb{R}$. Suppose that f and g are continuous at some point s of D. Then the functions $f + g$, $f - g$ and $f \cdot g$ are also continuous at s. If moreover the function g is everywhere non-zero on D then the function f/g is continuous at s.*

Proof This result follows directly from Proposition 1.8, using the fact that a function $f: D \to \mathbb{R}$ is continuous at a limit point s of D belonging to D if and only if $\lim_{x \to s} f(x) = f(s)$ (Lemma 1.9). ∎

Remark Proposition 1.10 can also be proved directly from the formal definition of continuity by a straightforward adaptation of the proof of Proposition 1.8. Indeed suppose that $f: D \to \mathbb{R}$ and $g: D \to \mathbb{R}$ are continuous at s, where $s \in D$. We show that $f + g$ is continuous at s. Let $\varepsilon > 0$ be given. Then there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that $|f(x) - f(s)| < \frac{1}{2}\varepsilon$ for all

$x \in D$ satisfying $|x - s| < \delta_1$, and $|g(x) - g(s)| < \frac{1}{2}\varepsilon$ for all $x \in D$ satisfying $|x - s| < \delta_2$. Let δ be the minimum of δ_1 and δ_2. If $|x - s| < \delta$ then

$$|f(x) + g(x) - (f(s) + g(s))| < |f(x) - f(s)| + |g(x) - g(s)| < \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon,$$

showing that $f + g$ is continuous at s. The proof of Proposition 1.8 can be adapted in a similar fashion to show that $f - g$, $f \cdot g$ and f/g are also continuous at s.

Proposition 1.11 *Let $f\colon D \to \mathbb{R}$ and $g\colon E \to \mathbb{R}$ be functions defined on D and E respectively, where D and E are subsets of $\mathbb{R}$ satisfying $f(D) \subset E$. Let s be an element of D. Suppose that the function f is continuous at s and that the function g is continuous at $f(s)$. Then the composition $g \circ f$ of f and g is continuous at s.*

Proof Let $\varepsilon > 0$ be given. Then there exists some $\eta > 0$ such that $|g(u) - g(f(s))| < \varepsilon$ for all $u \in E$ satisfying $|u - f(s)| < \eta$. But then there exists some $\delta > 0$ such that $|f(x) - f(s)| < \eta$ for all $x \in D$ satisfying $|x - s| < \delta$. Thus if $|x - s| < \delta$ then $|g(f(x)) - g(f(s))| < \varepsilon$. Hence $g \circ f$ is continuous at s. ∎

Lemma 1.12 *Let $f\colon D \to \mathbb{R}$ be a function defined on some subset D of $\mathbb{R}$, and let $a_1, a_2, a_3, \ldots$ be a sequence of real numbers belonging to D. Suppose that $a_n \to s$ as $n \to +\infty$, where $s \in D$, and that f is continuous at s. Then $f(a_n) \to f(s)$ as $n \to +\infty$.*

Proof Let $\varepsilon > 0$ be given. Then there exists some $\delta > 0$ such that $|f(x) - f(s)| < \varepsilon$ for all $x \in D$ satisfying $|x - s| < \delta$. But then there exists some positive integer N such that $|a_n - s| < \delta$ for all n satisfying $n \geq N$. Thus $|f(a_n) - f(s)| < \varepsilon$ for all $n \geq N$. Hence $f(a_n) \to f(s)$ as $n \to +\infty$. ∎

Proposition 1.13 *Let $f\colon D \to \mathbb{R}$ and $g\colon E \to \mathbb{R}$ be functions defined on D and E respectively, where D and E are subsets of $\mathbb{R}$ satisfying $f(D) \subset E$. Let s be a limit point of D, and let l be an element of E. Suppose that $\lim_{x \to s} f(x) = l$ and that the function g is continuous at l. Then $\lim_{x \to s} g(f(x)) = g(l)$.*

Proof Let $\varepsilon > 0$ be given. Then there exists some $\eta > 0$ such that $|g(u) - g(l)| < \varepsilon$ for all $u \in E$ satisfying $|u - l| < \eta$. But then there exists $\delta > 0$ such that $|f(x) - l| < \eta$ for all $x \in D$ satisfying $0 < |x - s| < \delta$. Thus if $0 < |x - s| < \delta$ then $|g(f(x)) - g(l)| < \varepsilon$. Hence $\lim_{x \to s} g(f(x)) \to g(l)$. ∎

1.8 The Intermediate Value Theorem

Theorem 1.14 (The Intermediate Value Theorem) *Let a and b be real numbers satisfying $a < b$, and let $f\colon [a,b] \to \mathbb{R}$ be a continuous function defined on the interval $[a,b]$. Let c be a real number which lies between $f(a)$ and $f(b)$ (so that either $f(a) \leq c \leq f(b)$ or else $f(a) \geq c \geq f(b)$.) Then there exists some $s \in [a,b]$ for which $f(s) = c$.*

Proof We first prove the result in the special case in which $c = 0$ and $f(a) \leq 0 \leq f(b)$. We must show that there exists some $s \in [a,b]$ for which $f(s) = 0$. Let S be the subset of $[a,b]$ defined by $S = \{x \in [a,b] : f(x) \leq 0\}$. The set S is non-empty and bounded above (since $a \in S$ and b is an upper bound for the set S). Therefore there exists a least upper bound $\sup S$ for the set S. Let $s = \sup S$. Then $a \leq s \leq b$, since $a \in S$ and $S \subset [a,b]$. We show that $f(s) = 0$.

Now if it were the case that $f(s) \neq 0$. An application of the definition of continuity (with $0 < \varepsilon \leq |f(s)|$) shows that there exists some $\delta > 0$ such that $f(x)$ has the same sign as $f(s)$ for all $x \in [a,b]$ satisfying $|x - s| < \delta$. (Thus if $f(s) > 0$ then $f(x) > 0$ whenever $|x - s| < \delta$, or if $f(s) < 0$ then $f(x) < 0$ whenever $|x - s| < \delta$.)

In particular, suppose that it were the case that $f(s) < 0$. Then $s < b$ (since $f(b) \geq 0$ by hypothesis), and hence $f(x) < 0$, and thus $x \in S$, for some $x \in [a,b]$ satisfying $s < x < s + \delta$. But this would contradict the definition of s.

Next suppose that it were the case that $f(s) > 0$. Then $s > a$ (since $f(a) \leq 0$ by hypothesis), and hence $f(x) > 0$, and thus $x \notin S$, for all $x \in [a,b]$ satisfying $x > s - \delta$. But then $f(x) > 0$ for all $x \geq s - \delta$, and thus $s - \delta$ would be an upper bound of the set S, which also contradicts the definition of s. The only remaining possibility is that $f(s) = 0$, which is what we are seeking to prove.

The result in the general case follows from that in the case $c = 0$ by applying the result in this special case to the function $x \mapsto f(x) - c$ when $f(a) \leq c \leq f(b)$, and to the function $x \mapsto c - f(x)$ when $f(a) \geq c \geq f(b)$. ∎

Corollary 1.15 *Given any positive real number b and natural number n, there exists some positive real number a satisfying $a^n = b$.*

Proof Let $f(x) = x^n - b$, and let $c = \max(b, 1)$. Then $f(0) < 0$ and $f(c) \geq 0$. It follows from the Intermediate Value Theorem (Theorem 1.14) that $f(a) = 0$ for some real number a satisfying $0 < a \leq c$. But then $a^n = b$, as required. ∎

Corollary 1.16 *Let P be a polynomial of odd degree with real coefficients. Then P has at least one real root.*

Proof Let us write $P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$, where n is odd and $a_n \neq 0$. Without loss of generality we may suppose that $a_n > 0$ (after replacing the polynomial P by $-P$ if necessary). We claim that $P(K) > 0$ and $P(-K) < 0$ for some sufficiently large real number K. (Indeed if K is chosen large enough to ensure that $K > 1$ and $|a_n|K \geq 2n|a_j|$ for $j = 0, 1, \ldots, n-1$ then $|P(x) - a_n x^n| \leq \frac{1}{2}|a_n x^n|$ whenever $|x| \geq K$. But then $P(x)$ and $a_n x^n$ have the same sign whenever $|x| \geq K$. In particular, $P(K) > 0$ and $P(-K) < 0$.) It follows immediately from the Intermediate Value Theorem (Theorem 1.14) that there exists some $x_0 \in [-K, K]$ for which $P(x_0) = 0$. Thus the polynomial P has at least one real root. ∎

A function f is said to be *strictly increasing* on an interval I if $f(x_1) < f(x_2)$ for all $x_1, x_2 \in I$ satisfying $x_1 < x_2$.

The next theorem is useful in verifying the continuity of functions involving square roots, nth roots, inverse trigonometric and logarithm functions.

Theorem 1.17 *A continuous strictly increasing function $f\colon [a,b] \to \mathbb{R}$ on an interval $[a,b]$ has a well-defined continuous inverse $f^{-1}\colon [c,d] \to [a,b]$ defined over the interval $[c,d]$, where $c = f(a)$ and $d = f(b)$.*

Proof The function f is injective (one-to-one) on $[a,b]$, and maps $[a,b]$ into the interval $[c,d]$, since f is strictly increasing. Moreover it follows immediately from the Intermediate Value Theorem that f maps $[a,b]$ onto $[c,d]$. Thus f is a bijection from $[a,b]$ to $[c,d]$ and so has a well-defined inverse $f^{-1}\colon [c,d] \to [a,b]$.

Let v satisfy $c < v < d$, and let $u = f^{-1}(v)$ (so that $f(u) = v$). We show that f^{-1} is continuous at v. Let $\varepsilon > 0$ be given. Now $a < u < b$, and hence there exist real numbers u_- and u_+ in the interval $[a,b]$ satisfying $u - \varepsilon < u_- < u < u_+ < u + \varepsilon$. Let $v_- = f(u_-)$ and $v_+ = f(u_+)$, and let δ be the minimum of $v_+ - v$ and $v - v_-$. Then $\delta > 0$, and if $y \in [c,d]$ satisfies $|y - v| < \delta$ then $v_- < y < v_+$. But then $u_- < f^{-1}(y) < u_+$, and thus $|f^{-1}(y) - f^{-1}(v)| < \varepsilon$. We deduce that f^{-1} is continuous at v whenever $c < v < d$. A similar proof shows that f^{-1} is continuous at both c and d. ∎

1.9 Continuous Functions on Closed Bounded Intervals

Theorem 1.18 *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous real-valued function defined on the interval $[a,b]$. Then there exists a constant M with the property that $|f(x)| \leq M$ for all $x \in [a,b]$.*

We give two proofs of this theorem. The first uses the Bolzano-Weierstrass Theorem which states that every bounded sequence of real numbers possesses a convergent subsequence. The second makes use of the Least Upper Bound Axiom.

1st Proof Suppose that the function were not bounded on the interval $[a,b]$. Then there would exist a sequence $x_1, x_2, x_3, \ldots$ of real numbers in the interval $[a,b]$ such that $|f(x_n)| > n$ for all n. The bounded sequence $x_1, x_2, x_3, \ldots$ would possess a convergent subsequence $x_{n_1}, x_{n_2}, x_{n_3}, \ldots$, by the Bolzano-Weierstrass Theorem (Theorem 1.4). Moreover the limit l of this subsequence would belong to $[a,b]$. But then $f(x_{n_k}) \to f(l)$ as $k \to +\infty$, by Lemma 1.12. It follows that there exists some natural number N with the property that $|f(x_{n_k}) - f(l)| < 1$ whenever $k \geq N$, so that $|f(x_{n_k})| \leq |f(l)| + 1$ whenever $k \geq N$. But this is a contradiction, since $|f(x_{n_k})| > n_k$ for all k and n_k increases without limit as $k \to +\infty$. Thus the function f is indeed bounded on the closed interval $[a,b]$. ∎

2nd Proof Define $S = \{\tau \in [a,b] : f \text{ is bounded on } [a,\tau]\}$. Clearly $a \in S$ and $S \subset [a,b]$. Thus the set S is non-empty and bounded. It follows from the Least Upper Bound axiom that there exists a least upper bound for the set S. Let $s = \sup S$. Then $s \in [a,b]$. The function f is continuous at s. Therefore there exists some $\delta > 0$ such that $|f(x) - f(s)| < 1$, and thus $|f(x)| < |f(s)| + 1$, for all $x \in [a,b]$ satisfying $|x - s| < \delta$.

Now $s-\delta$ is not an upper bound for the set S and hence $s-\delta < \tau \leq s$ for some $\tau \in S$. But then the function f is bounded on $[a,\tau]$ (since $\tau \in S$) and on $[\tau, s]$ (since $|f(s)| + 1$ is an upper bound on this interval). We conclude that f is bounded on $[a,s]$, and thus $s \in S$. Moreover if it were the case that $s < b$ then the function f would be bounded on $[a,x]$, and thus $x \in S$, for all $x \in [a,b]$ satisfying $s < x < s + \delta$, contradicting the definition of s as the least upper bound of the set S. Thus $s = b$. But then $b \in S$, so that the function f is bounded on the interval $[a,b]$ as required. ∎

Theorem 1.19 *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous real-valued function defined on the interval $[a,b]$. Then there exist $u, v \in [a,b]$ with the property that $f(u) \leq f(x) \leq f(v)$ for all $x \in [a,b]$.*

Proof Let $C = \sup\{f(x) : a \leq x \leq b\}$. If there did not exist any $v \in [a,b]$ for which $f(v) = C$ then the function $x \mapsto 1/(C - f(x))$ would be a continuous function on the interval $[a,b]$ which was not bounded above on this interval, thus contradicting Theorem 1.18. Thus there must exist some $v \in [a,b]$ with the property that $f(v) = C$. A similar proof shows that there must exist some $u \in [a,b]$ with the property that $g(u) = c$, where $c = \inf\{f(x) : a \leq x \leq b\}$. But then $f(u) \leq f(x) \leq f(v)$ for all $x \in [a,b]$, as required. ∎

2 Convergence, Continuity and Open Sets in Euclidean Spaces

We denote by $\mathbb{R}^n$ the set consisting of all n-tuples $(x_1, x_2, \ldots, x_n)$ of real numbers. The set $\mathbb{R}^n$ represents n-dimensional *Euclidean space* (with respect to the standard Cartesian coordinate system). Let $\mathbf{x}$ and $\mathbf{y}$ be elements of $\mathbb{R}^n$, where

$$\mathbf{x} = (x_1, x_2, \ldots, x_n), \quad \mathbf{y} = (y_1, y_2, \ldots, y_n),$$

and let λ be a real number. We define

$$\begin{aligned}
\mathbf{x} + \mathbf{y} &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n), \\
\mathbf{x} - \mathbf{y} &= (x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n), \\
\lambda\mathbf{x} &= (\lambda x_1, \lambda x_2, \ldots, \lambda x_n), \\
\mathbf{x} \cdot \mathbf{y} &= x_1 y_1 + x_2 y_2 + \cdots + x_n y_n, \\
|\mathbf{x}| &= \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.
\end{aligned}$$

The quantity $\mathbf{x} \cdot \mathbf{y}$ is the *scalar product* (or *inner product*) of $\mathbf{x}$ and $\mathbf{y}$, and the quantity $|\mathbf{x}|$ is the *Euclidean norm* of $\mathbf{x}$. Note that $|\mathbf{x}|^2 = \mathbf{x} \cdot \mathbf{x}$. The *Euclidean distance* between two points $\mathbf{x}$ and $\mathbf{y}$ of $\mathbb{R}^n$ is defined to be the Euclidean norm $|\mathbf{y} - \mathbf{x}|$ of the vector $\mathbf{y} - \mathbf{x}$.

Lemma 2.1 (Schwarz' Inequality) *Let $\mathbf{x}$ and $\mathbf{y}$ be elements of $\mathbb{R}^n$. Then $|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|$.*

Proof We note that $|\lambda\mathbf{x} + \mu\mathbf{y}|^2 \geq 0$ for all real numbers λ and μ. But

$$|\lambda\mathbf{x} + \mu\mathbf{y}|^2 = (\lambda\mathbf{x} + \mu\mathbf{y}).(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda^2|\mathbf{x}|^2 + 2\lambda\mu\mathbf{x} \cdot \mathbf{y} + \mu^2|\mathbf{y}|^2.$$

Therefore $\lambda^2|\mathbf{x}|^2 + 2\lambda\mu\mathbf{x} \cdot \mathbf{y} + \mu^2|\mathbf{y}|^2 \geq 0$ for all real numbers λ and μ. In particular, suppose that $\lambda = |\mathbf{y}|^2$ and $\mu = -\mathbf{x} \cdot \mathbf{y}$. We conclude that

$$|\mathbf{y}|^4|\mathbf{x}|^2 - 2|\mathbf{y}|^2(\mathbf{x} \cdot \mathbf{y})^2 + (\mathbf{x} \cdot \mathbf{y})^2|\mathbf{y}|^2 \geq 0,$$

so that $\left(|\mathbf{x}|^2|\mathbf{y}|^2 - (\mathbf{x} \cdot \mathbf{y})^2\right)|\mathbf{y}|^2 \geq 0$. Thus if $\mathbf{y} \neq \mathbf{0}$ then $|\mathbf{y}| > 0$, and hence

$$|\mathbf{x}|^2|\mathbf{y}|^2 - (\mathbf{x} \cdot \mathbf{y})^2 \geq 0.$$

But this inequality is trivially satisfied when $\mathbf{y} = \mathbf{0}$. Thus $|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|$, as required. ∎

It follows easily from Schwarz' Inequality that $|\mathbf{x}+\mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. For

$$\begin{aligned} |\mathbf{x}+\mathbf{y}|^2 &= (\mathbf{x}+\mathbf{y}).(\mathbf{x}+\mathbf{y}) = |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2\mathbf{x}\cdot\mathbf{y} \\ &\leq |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2|\mathbf{x}||\mathbf{y}| = (|\mathbf{x}| + |\mathbf{y}|)^2. \end{aligned}$$

It follows that

$$|\mathbf{z}-\mathbf{x}| \leq |\mathbf{z}-\mathbf{y}| + |\mathbf{y}-\mathbf{x}|$$

for all points $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ of $\mathbb{R}^n$. This important inequality is known as the *Triangle Inequality.* It expresses the geometric fact the the length of any triangle in a Euclidean space is less than or equal to the sum of the lengths of the other two sides.

Definition A sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ of points in $\mathbb{R}^n$ is said to *converge* to a point $\mathbf{p}$ if and only if the following criterion is satisfied:—

> given any real number ε satisfying $\varepsilon > 0$ there exists some natural number N such that $|\mathbf{p} - \mathbf{x}_j| < \varepsilon$ whenever $j \geq N$.

We refer to $\mathbf{p}$ as the *limit* $\lim\limits_{j\to+\infty} \mathbf{x}_j$ of the sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$.

Lemma 2.2 *Let $\mathbf{p}$ be a point of $\mathbb{R}^n$, where $\mathbf{p} = (p_1, p_2, \ldots, p_n)$. Then a sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ of points in $\mathbb{R}^n$ converges to $\mathbf{p}$ if and only if the ith components of the elements of this sequence converge to p_i for $i = 1, 2, \ldots, n$.*

Proof Let x_{ji} and p_i denote the ith components of $\mathbf{x}_j$ and $\mathbf{p}$, where $\mathbf{p} = \lim\limits_{j\to+\infty} \mathbf{x}_j$. Then $|x_{ji} - p_i| \leq |\mathbf{x}_j - \mathbf{p}|$ for all j. It follows directly from the definition of convergence that if $\mathbf{x}_j \to \mathbf{p}$ as $j \to +\infty$ then $x_{ji} \to p_i$ as $j \to +\infty$.

Conversely suppose that, for each i, $x_{ji} \to p_i$ as $j \to +\infty$. Let $\varepsilon > 0$ be given. Then there exist natural numbers $N_1, N_2, \ldots, N_n$ such that $|x_{ji} - p_i| < \varepsilon/\sqrt{n}$ whenever $j \geq N_i$. Let N be the maximum of $N_1, N_2, \ldots, N_n$. If $j \geq N$ then

$$|\mathbf{x}_j - \mathbf{p}|^2 = \sum_{i=1}^{n} (x_{ji} - p_i)^2 < n(\varepsilon/\sqrt{n})^2 = \varepsilon^2,$$

so that $\mathbf{x}_j \to \mathbf{p}$ as $j \to +\infty$. ∎

Definition A sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ of points in $\mathbb{R}^n$ is said to be a *Cauchy sequence* if and only if the following criterion is satisfied:—

> given any real number ε satisfying $\varepsilon > 0$ there exists some natural number N such that $|\mathbf{x}_j - \mathbf{x}_k| < \varepsilon$ whenever $j \geq N$ and $k \geq N$.

Lemma 2.3 *A sequence of points in $\mathbb{R}^n$ is convergent if and only if it is a Cauchy sequence.*

Proof Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ be a sequence of points of $\mathbb{R}^n$ converging to some point $\mathbf{p}$. Let $\varepsilon > 0$ be given. Then there exists some natural number N such that $|\mathbf{x}_j - \mathbf{p}| < \frac{1}{2}\varepsilon$ whenever $j \geq N$. If $j \geq N$ and $k \geq N$ then

$$|\mathbf{x}_j - \mathbf{x}_k| \leq |\mathbf{x}_j - \mathbf{p}| + |\mathbf{p} - \mathbf{x}_k| < \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon,$$

by the Triangle Inequality. Thus every convergent sequence in $\mathbb{R}^n$ is a Cauchy sequence.

Now let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ be a Cauchy sequence in $\mathbb{R}^n$. Then the ith components of the elements of this sequence constitute a Cauchy sequence of real numbers. This Cauchy sequence must converge to some real number p_i, by Cauchy's Criterion for Convergence (Theorem 1.6). It follows from Lemma 2.2 that the Cauchy sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ converges to the point $\mathbf{p}$, where $\mathbf{p} = (p_1, p_2, \ldots, p_n)$. ∎

Definition Let X and Y be a subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. A function $f: X \to Y$ from X to Y is said to be *continuous* at a point $\mathbf{p}$ of X if and only if the following criterion is satisfied:—

> given any real number ε satisfying $\varepsilon > 0$ there exists some $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{p})| < \varepsilon$ for all points $\mathbf{x}$ of X satisfying $|\mathbf{x} - \mathbf{p}| < \delta$.

The function $f: X \to Y$ is said to be continuous on X if and only if it is continuous at every point $\mathbf{p}$ of X.

Lemma 2.4 *Let X, Y and Z be subsets of $\mathbb{R}^m$, $\mathbb{R}^n$ and $\mathbb{R}^k$ respectively, and let $f: X \to Y$ and $g: Y \to Z$ be functions satisfying $f(X) \subset Y$. Suppose that f is continuous at some point $\mathbf{p}$ of X and that g is continuous at $f(\mathbf{p})$. Then the composition function $g \circ f: X \to Z$ is continuous at $\mathbf{p}$.*

Proof Let $\varepsilon > 0$ be given. Then there exists some $\eta > 0$ such that $|g(\mathbf{y}) - g(f(\mathbf{p}))| < \varepsilon$ for all $\mathbf{y} \in Y$ satisfying $|\mathbf{y} - f(\mathbf{p})| < \eta$. But then there exists some $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{p})| < \eta$ for all $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$. It follows that $|g(f(\mathbf{x})) - g(f(\mathbf{p}))| < \varepsilon$ for all $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$, and thus $g \circ f$ is continuous at $\mathbf{p}$, as required. ∎

Lemma 2.5 *Let X and Y be a subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, and let $f: X \to Y$ be a continuous function from X to Y. Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ be a sequence of points of X which converges to some point $\mathbf{p}$ of X. Then the sequence $f(\mathbf{x}_1), f(\mathbf{x}_2), f(\mathbf{x}_3), \ldots$ converges to $f(\mathbf{p})$.*

Proof Let $\varepsilon > 0$ be given. Then there exists some $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{p})| < \varepsilon$ for all $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$, since the function f is continuous at $\mathbf{p}$. Also there exists some natural number N such that $|\mathbf{x}_j - \mathbf{p}| < \delta$ whenever $j \geq N$, since the sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ converges to $\mathbf{p}$. Thus if $j \geq N$ then $|f(\mathbf{x}_j) - f(\mathbf{p})| < \varepsilon$. Thus the sequence $f(\mathbf{x}_1), f(\mathbf{x}_2), f(\mathbf{x}_3), \ldots$ converges to $f(\mathbf{p})$, as required. ∎

Let X and Y be a subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, and let $f\colon X \to Y$ be a function from X to Y. Then

$$f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x}))$$

for all $\mathbf{x} \in X$, where $f_1, f_2, \ldots, f_n$ are functions from X to $\mathbb{R}$, referred to as the *components* of the function f.

Proposition 2.6 *Let X and Y be a subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. A function $f\colon X \to Y$ is continuous if and only if its components are continuous.*

Proof Note that the ith component f_i of f is given by $f_i = \pi_i \circ f$, where $\pi_i\colon \mathbb{R}^n \to \mathbb{R}$ is the continuous function which maps $(y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ onto its ith coordinate y_i. Now any composition of continuous functions is continuous, by Lemma 2.4. Thus if f is continuous, then so are the components of f.

Conversely suppose that the components of f are continuous at $\mathbf{p} \in X$. Let $\varepsilon > 0$ be given. Then there exist positive real numbers $\delta_1, \delta_2, \ldots, \delta_n$ such that $|f_i(\mathbf{x}) - f_i(\mathbf{p})| < \varepsilon/\sqrt{n}$ for $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta_i$. Let δ be the minimum of $\delta_1, \delta_2, \ldots, \delta_n$. If $\mathbf{x} \in X$ satisfies $|\mathbf{x} - \mathbf{p}| < \delta$ then

$$|f(\mathbf{x}) - f(\mathbf{p})|^2 = \sum_{i=1}^{n} |f_i(\mathbf{x}) - f_i(\mathbf{p})|^2 < \varepsilon^2,$$

and hence $|f(\mathbf{x}) - f(\mathbf{p})| < \varepsilon$. Thus the function f is continuous at $\mathbf{p}$, as required. ∎

Lemma 2.7 *The functions $s\colon \mathbb{R}^2 \to \mathbb{R}$ and $p\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $s(x, y) = x + y$ and $p(x, y) = xy$ are continuous.*

Proof Let $(u, v) \in \mathbb{R}^2$. We first show that $s\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous at (u, v). Let $\varepsilon > 0$ be given. Let $\delta = \frac{1}{2}\varepsilon$. If (x, y) is any point of $\mathbb{R}^2$ whose distance from (u, v) is less than δ then $|x - u| < \delta$ and $|y - v| < \delta$, and hence

$$|s(x, y) - s(u, v)| = |x + y - u - v| \leq |x - u| + |y - v| < 2\delta = \varepsilon.$$

This shows that $s\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous at (u, v).

Next we show that $p\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous at (u, v). Now

$$p(x, y) - p(u, v) = xy - uv = (x - u)(y - v) + u(y - v) + (x - u)v.$$

for all points (x, y) of $\mathbb{R}^2$. Thus if the distance from (x, y) to (u, v) is less than δ then $|x - u| < \delta$ and $|y - v| < \delta$, and hence $|p(x, y) - p(u, v)| < \delta^2 + (|u| + |v|)\delta$. Let $\varepsilon > 0$ is given. If $\delta > 0$ is chosen to be the minimum of 1 and $\varepsilon/(1 + |u| + |v|)$ then $\delta^2 + (|u| + |v|)\delta < (1 + |u| + |v|)\delta < \varepsilon$, and thus $|p(x, y) - p(u, v)| < \varepsilon$ for all points (x, y) of $\mathbb{R}^2$ whose distance from (u, v) is less than δ. This shows that $p\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous at (u, v). ∎

Proposition 2.8 *Let X be a subset of $\mathbb{R}^n$, and let $f\colon X \to \mathbb{R}$ and $g\colon X \to \mathbb{R}$ be continuous functions from X to $\mathbb{R}$. Then the functions $f + g$, $f - g$ and $f \cdot g$ are continuous. If in addition $g(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in X$ then the quotient function f/g is continuous.*

Proof Note that $f + g = s \circ h$ and $f \cdot g = p \circ h$, where $h\colon X \to \mathbb{R}^2$, $s\colon \mathbb{R}^2 \to \mathbb{R}$ and $p\colon \mathbb{R}^2 \to \mathbb{R}$ are given by $h(\mathbf{x}) = (f(\mathbf{x}), g(\mathbf{x}))$, $s(u, v) = u + v$ and $p(u, v) = uv$ for all $\mathbf{x} \in X$ and $u, v \in \mathbb{R}$. It follows from Proposition 2.6, Lemma 2.7 and Lemma 2.4 that $f + g$ and $f \cdot g$ are continuous, being compositions of continuous functions. Now $f - g = f + (-g)$, and both f and $-g$ are continuous. Therefore $f - g$ is continuous.

Now suppose that $g(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in X$. Note that $1/g = r \circ g$, where $r\colon \mathbb{R} \setminus \{0\} \to \mathbb{R}$ is the reciprocal function, defined by $r(t) = 1/t$. Now the reciprocal function r is continuous. Thus the function $1/g$ is a composition of continuous functions and is thus continuous. But then, using the fact that a product of continuous real-valued functions is continuous, we deduce that f/g is continuous. ∎

Example Consider the function $f\colon \mathbb{R}^2 \setminus \{(0, 0)\} \to \mathbb{R}^2$ defined by

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

The continuity of the components of the function f follows from straightforward applications of Proposition 2.8. It then follows from Proposition 2.6 that the function f is continuous on $\mathbb{R}^2 \setminus \{(0, 0)\}$.

Let X and Y be subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. A function $h\colon X \to Y$ from X to Y is said to be a *homeomorphism* if it is a bijection and both $h\colon X \to Y$ and its inverse $h^{-1}\colon Y \to X$ are continuous. If there exists a homeomorphism $h\colon X \to Y$ from X to Y then X and Y are said to be *homeomorphic*.

Example The interval $(-1, 1)$ and the real line $\mathbb{R}$ are homeomorphic. Indeed if $h\colon (-1, 1) \to \mathbb{R}$ is defined by $h(t) = \tan(\pi t/2)$, then h is a homeomorphism from $(-1, 1)$ to $\mathbb{R}$.

Example Let $B^n = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| < 1\}$, and let $\varphi\colon B^n \to \mathbb{R}^n$ be the function defined by

$$\varphi(\mathbf{x}) = \frac{1}{1 - |\mathbf{x}|^2}\mathbf{x}.$$

The function φ is a bijection from B^n to $\mathbb{R}^n$ whose inverse $\varphi^{-1}\colon \mathbb{R}^n \to B$ is given by

$$\varphi^{-1}(\mathbf{x}) = \begin{cases} \dfrac{-1 + \sqrt{1 + 4|\mathbf{x}|^2}}{2|\mathbf{x}|^2}\mathbf{x} & \text{if } \mathbf{x} \neq 0; \\ \mathbf{0} & \text{if } \mathbf{x} = 0. \end{cases}$$

We claim that $\varphi\colon B^n \to \mathbb{R}^n$ and $\varphi^{-1}\colon \mathbb{R}^n \to B$ are continuous. Now $\varphi^{-1}(\mathbf{x}) = f(|\mathbf{x}|)\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, where $f\colon \mathbb{R} \to \mathbb{R}$ is the continuous function defined by

$$f(t) = \begin{cases} \dfrac{-1 + \sqrt{1 + 4t^2}}{2t^2} & \text{if } t \neq 0; \\ 1 & \text{if } t = 0. \end{cases}$$

(The continuity of this function at 0 follows from a straightforward application of l'Hópital's rule.) Straighforward applications of Lemma 2.4, Proposition 2.6 and Proposition 2.8 show that φ and φ^{-1} are continuous. Thus $\varphi\colon B^n \to \mathbb{R}^n$ is a homeomorphism from B^n to $\mathbb{R}^n$. In particular, on setting $n = 1$, we conclude that the function $k\colon (-1, 1) \to \mathbb{R}$ defined by $k(t) = t/(1 - t^2)$ for all $t \in (-1, 1)$ is a homeomorphism from $(-1, 1)$ to $\mathbb{R}$.

Example Let S^n denote the n-sphere, defined by $S^n = \{\mathbf{x} \in \mathbb{R}^{n+1} : |\mathbf{x}| = 1\}$, and let $\mathbf{p}$ be the point of S^n with coordinates $(0, 0, \ldots, 0, 1)$. We show that $S^n \setminus \{\mathbf{p}\}$ is homeomorphic to $\mathbb{R}^n$. Define a function $h\colon S^n \setminus \{\mathbf{p}\} \to \mathbb{R}^n$ by

$$h(x_1, x_2, \ldots, x_n, x_{n+1}) = \left(\frac{x_1}{1 - x_{n+1}}, \frac{x_2}{1 - x_{n+1}}, \ldots, \frac{x_n}{1 - x_{n+1}}\right).$$

The function h is a bijection whose inverse $h^{-1}\colon \mathbb{R}^n \to S^n \setminus \{\mathbf{p}\}$ is given by

$$h^{-1}(y_1, y_2, \ldots, y_n) = \left(\frac{2y_1}{|\mathbf{y}|^2 + 1}, \frac{2y_2}{|\mathbf{y}|^2 + 1}, \ldots, \frac{2y_n}{|\mathbf{y}|^2 + 1}, \frac{|\mathbf{y}|^2 - 1}{|\mathbf{y}|^2 + 1}\right)$$

(where $|\mathbf{y}|^2 = y_1^2 + y_2^2 + \cdots + y_n^2$). A straightforward application of Propositions 2.6 and 2.8 shows that the functions h and h^{-1} are continuous. Thus $h\colon S^n \setminus \{\mathbf{p}\} \to \mathbb{R}^n$ is a homeomorphism. This homeomorphism represents *stereographic projection* from $S^n \setminus \{\mathbf{p}\}$ to $\mathbb{R}^n$.

Let X and Y be subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$, and let $h: X \to Y$ be a homeomorphism. A sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ of points of X converges in X to some point $\mathbf{p}$ if and only if the sequence $h(\mathbf{x}_1), h(\mathbf{x}_2), h(\mathbf{x}_3), \ldots$ converges in Y to $h(\mathbf{p})$. (This follows on applying Lemma 2.5 to $h: X \to Y$ and $h^{-1}: Y \to X$.) Also a function $f: Y \to Z$ mapping Y into some subset Z of $\mathbb{R}^p$ is continuous if and only if $f \circ h: X \to Z$ is continuous, and a function $g: W \to X$ mapping some subset W of $\mathbb{R}^k$ into X is continuous if and only if $h \circ g: W \to Y$ is continuous.

2.1 Open Sets in Euclidean Spaces

Let X be a subset of $\mathbb{R}^n$. Given a point $\mathbf{p}$ of X and a non-negative real number r, the *open ball* $B_X(\mathbf{p}, r)$ in X of *radius* r about $\mathbf{p}$ is defined to be the subset of X given by

$$B_X(\mathbf{p}, r) \equiv \{\mathbf{x} \in X : |\mathbf{x} - \mathbf{p}| < r\}.$$

(Thus $B_X(\mathbf{p}, r)$ is the set consisting of all points of X that lie within a sphere of radius r centred on the point $\mathbf{p}$.)

Definition Let X be a subset of $\mathbb{R}^n$. A subset V of X is said to be *open* in X if and only if, given any point $\mathbf{p}$ of V, there exists some $\delta > 0$ such that $B_X(\mathbf{p}, \delta) \subset V$.

By convention, we regard the empty set $\emptyset$ as being an open subset of X. (The criterion given above is satisfied vacuously in the case when V is the empty set.)

In particular, a subset V of $\mathbb{R}^n$ is said to be an *open set* (in $\mathbb{R}^n$) if and only if, given any point $\mathbf{p}$ of V, there exists some $\delta > 0$ such that $B(\mathbf{p}, \delta) \subset V$, where $B(\mathbf{p}, r) = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{p}| < r\}$.

Example Let $H = \{(x, y, z) \in \mathbb{R}^3 : z > c\}$, where c is some real number. Then H is an open set in $\mathbb{R}^3$. Indeed let $\mathbf{p}$ be a point of H. Then $\mathbf{p} = (u, v, w)$, where $w > c$. Let $\delta = w - c$. If the distance from a point (x, y, z) to the point (u, v, w) is less than δ then $|z - w| < \delta$, and hence $z > c$, so that $(x, y, z) \in H$. Thus $B(\mathbf{p}, \delta) \subset H$, and therefore H is an open set.

The previous example can be generalized. Given any integer i between 1 and n, and given any real number c_i, the sets

$$\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_i > c_i\}, \qquad \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_i < c_i\}$$

are open sets in $\mathbb{R}^n$.

Example Let U be an open set in $\mathbb{R}^n$. Then for any subset X of $\mathbb{R}^n$, the intersection $U \cap X$ is open in X. (This follows directly from the definitions.) Thus for example, let S^2 be the unit sphere in $\mathbb{R}^3$, given by

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$$

and let N be the subset of S^2 given by

$$N = \{(x, y, z) \in \mathbb{R}^n : x^2 + y^2 + z^2 = 1 \text{ and } z > 0\}.$$

Then N is open in S^2, since $N = H \cap S^2$, where H is the open set in $\mathbb{R}^3$ given by

$$H = \{(x, y, z) \in \mathbb{R}^3 : z > 0\}.$$

Note that N is not itself an open set in $\mathbb{R}^3$. Indeed the point $(0, 0, 1)$ belongs to N, but, for any $\delta > 0$, the open ball (in $\mathbb{R}^3$ of radius δ about $(0, 0, 1)$ contains points (x, y, z) for which $x^2 + y^2 + z^2 \neq 1$. Thus the open ball of radius δ about the point $(0, 0, 1)$ is not a subset of N.

Lemma 2.9 *Let X be a subset of $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. Then, for any positive real number r, the open ball $B_X(\mathbf{p}, r)$ in X of radius r about $\mathbf{p}$ is open in X.*

Proof Let $\mathbf{x}$ be an element of $B_X(\mathbf{p}, r)$. We must show that there exists some $\delta > 0$ such that $B_X(\mathbf{x}, \delta) \subset B_X(\mathbf{p}, r)$. Let $\delta = r - |\mathbf{x} - \mathbf{p}|$. Then $\delta > 0$, since $|\mathbf{x} - \mathbf{p}| < r$. Moreover if $\mathbf{y} \in B_X(\mathbf{x}, \delta)$ then

$$|\mathbf{y} - \mathbf{p}| \leq |\mathbf{y} - \mathbf{x}| + |\mathbf{x} - \mathbf{p}| < \delta + |\mathbf{x} - \mathbf{p}| = r,$$

by the Triangle Inequality, and hence $\mathbf{y} \in B_X(\mathbf{p}, r)$. Thus $B_X(\mathbf{x}, \delta) \subset B_X(\mathbf{p}, r)$. This shows that $B_X(\mathbf{p}, r)$ is an open set, as required. ∎

Lemma 2.10 *Let X be a subset of $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. Then, for any non-negative real number r, the set $\{\mathbf{x} \in X : |\mathbf{x} - \mathbf{p}| > r\}$ is an open set in X.*

Proof Let $\mathbf{x}$ be a point of X satisfying $|\mathbf{x} - \mathbf{p}| > r$, and let $\mathbf{y}$ be any point of X satisfying $|\mathbf{y} - \mathbf{x}| < \delta$, where $\delta = |\mathbf{x} - \mathbf{p}| - r$. Then

$$|\mathbf{x} - \mathbf{p}| \leq |\mathbf{x} - \mathbf{y}| + |\mathbf{y} - \mathbf{p}|,$$

by the Triangle Inequality, and therefore

$$|\mathbf{y} - \mathbf{p}| \geq |\mathbf{x} - \mathbf{p}| - |\mathbf{y} - \mathbf{x}| > |\mathbf{x} - \mathbf{p}| - \delta = r.$$

Thus $B_X(\mathbf{x}, \delta)$ is contained in the given set. The result follows. ∎

Proposition 2.11 *Let X be a subset of $\mathbb{R}^n$. The collection of open sets in X has the following properties:—*

(i) *the empty set $\emptyset$ and the whole set X are both open in X;*

(ii) *the union of any collection of open sets in X is itself open in X;*

(iii) *the intersection of any* finite *collection of open sets in X is itself open in X.*

Proof The empty set $\emptyset$ is an open set by convention. Moreover the definition of an open set is satisfied trivially by the whole set X. This proves (i).

Let $\mathcal{A}$ be any collection of open sets in X, and let U denote the union of all the open sets belonging to $\mathcal{A}$. We must show that U is itself open in X. Let $\mathbf{x} \in U$. Then $\mathbf{x} \in V$ for some set V belonging to the collection $\mathcal{A}$. It follows that there exists some $\delta > 0$ such that $B_X(\mathbf{x}, \delta) \subset V$. But $V \subset U$, and thus $B_X(\mathbf{x}, \delta) \subset U$. This shows that U is open in X. This proves (ii).

Finally let $V_1, V_2, V_3, \ldots, V_k$ be a *finite* collection of subsets of X that are open in X, and let V denote the intersection $V_1 \cap V_2 \cap \cdots \cap V_k$ of these sets. Let $\mathbf{x} \in V$. Now $\mathbf{x} \in V_j$ for $j = 1, 2, \ldots, k$, and therefore there exist strictly positive real numbers $\delta_1, \delta_2, \ldots, \delta_k$ such that $B_X(\mathbf{x}, \delta_j) \subset V_j$ for $j = 1, 2, \ldots, k$. Let δ be the minimum of $\delta_1, \delta_2, \ldots, \delta_k$. Then $\delta > 0$. (This is where we need the fact that we are dealing with a finite collection of sets.) Now $B_X(\mathbf{x}, \delta) \subset B_X(\mathbf{x}, \delta_j) \subset V_j$ for $j = 1, 2, \ldots, k$, and thus $B_X(\mathbf{x}, \delta) \subset V$. Thus the intersection V of the sets $V_1, V_2, \ldots, V_k$ is itself open in X. This proves (iii). ∎

Example The set $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 < 4 \text{ and } z > 1\}$ is an open set in $\mathbb{R}^3$, since it is the intersection of the open ball of radius 2 about the origin with the open set $\{(x, y, z) \in \mathbb{R}^3 : z > 1\}$.

Example The set $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 < 4 \text{ or } z > 1\}$ is an open set in $\mathbb{R}^3$, since it is the union of the open ball of radius 2 about the origin with the open set $\{(x, y, z) \in \mathbb{R}^3 : z > 1\}$.

Example The set

$$\{(x, y, z) \in \mathbb{R}^3 : (x - n)^2 + y^2 + z^2 < \tfrac{1}{4} \text{ for some } n \in \mathbb{Z}\}$$

is an open set in $\mathbb{R}^3$, since it is the union of the open balls of radius $\frac{1}{2}$ about the points $(n, 0, 0)$ for all integers n.

Example For each natural number k, let

$$V_k = \{(x, y, z) \in \mathbb{R}^3 : k^2(x^2 + y^2 + z^2) < 1\}.$$

Now each set V_k is an open ball of radius $1/k$ about the origin, and is therefore an open set in $\mathbb{R}^3$. However the intersection of the sets V_k for all natural numbers k is the set $\{(0, 0, 0)\}$, and thus the intersection of the sets V_k for all natural numbers k is not itself an open set in $\mathbb{R}^3$. This example demonstrates that infinite intersections of open sets need not be open.

Lemma 2.12 *A sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ of points in $\mathbb{R}^n$ converges to a point $\mathbf{p}$ if and only if, given any open set U which contains $\mathbf{p}$, there exists some natural number N such that $\mathbf{x}_j \in U$ for all j satisfying $j \geq N$.*

Proof Suppose that the sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ has the property that, given any open set U which contains $\mathbf{p}$, there exists some natural number N such that $\mathbf{x}_j \in U$ whenever $j \geq N$. Let $\varepsilon > 0$ be given. The open ball $B(\mathbf{p}, \varepsilon)$ of radius ε about $\mathbf{p}$ is an open set by Lemma 2.9. Therefore there exists some natural number N such that $\mathbf{x}_j \in B(\mathbf{p}, \varepsilon)$ whenever $j \geq N$. Thus $|\mathbf{x}_j - \mathbf{p}| < \varepsilon$ whenever $j \geq N$. This shows that the sequence converges to $\mathbf{p}$.

Conversely, suppose that the sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ converges to $\mathbf{p}$. Let U be an open set which contains $\mathbf{p}$. Then there exists some $\varepsilon > 0$ such that the open ball $B(\mathbf{p}, \varepsilon)$ of radius ε about $\mathbf{p}$ is a subset of U. Thus there exists some $\varepsilon > 0$ such that U contains all points $\mathbf{x}$ of X that satisfy $|\mathbf{x} - \mathbf{p}| < \varepsilon$. But there exists some natural number N with the property that $|\mathbf{x}_j - \mathbf{p}| < \varepsilon$ whenever $j \geq N$, since the sequence converges to $\mathbf{p}$. Therefore $\mathbf{x}_j \in U$ whenever $j \geq N$, as required. ∎

2.2 Closed Sets in Euclidean Spaces

Let X be a subset of $\mathbb{R}^n$. A subset F of X is said to be *closed* in X if and only if its complement $X \setminus F$ in X is open in X. (Recall that $X \setminus F = \{\mathbf{x} \in X : \mathbf{x} \notin F\}$.)

Example The sets $\{(x, y, z) \in \mathbb{R}^3 : z \geq c\}$, $\{(x, y, z) \in \mathbb{R}^3 : z \leq c\}$, and $\{(x, y, z) \in \mathbb{R}^3 : z = c\}$ are closed sets in $\mathbb{R}^3$ for each real number c, since the complements of these sets are open in $\mathbb{R}^3$.

Example Let X be a subset of $\mathbb{R}^n$, and let $\mathbf{x}_0$ be a point of X. Then the sets $\{\mathbf{x} \in X : |\mathbf{x} - \mathbf{x}_0| \leq r\}$ and $\{\mathbf{x} \in X : |\mathbf{x} - \mathbf{x}_0| \geq r\}$ are closed for each non-negative real number r. In particular, the set $\{\mathbf{x}_0\}$ consisting of the single point $\mathbf{x}_0$ is a closed set in X. (These results follow immediately using Lemma 2.9 and Lemma 2.10 and the definition of closed sets.)

Let $\mathcal{A}$ be some collection of subsets of a set X. Then

$$X \setminus \bigcup_{S \in \mathcal{A}} S = \bigcap_{S \in \mathcal{A}} (X \setminus S), \qquad X \setminus \bigcap_{S \in \mathcal{A}} S = \bigcup_{S \in \mathcal{A}} (X \setminus S)$$

(i.e., the complement of the union of some collection of subsets of X is the intersection of the complements of those sets, and the complement of the intersection of some collection of subsets of X is the union of the complements of those sets). The following result therefore follows directly from Proposition 2.11.

Proposition 2.13 *Let X be a subset of $\mathbb{R}^n$. The collection of closed sets in X has the following properties:—*

(i) *the empty set $\emptyset$ and the whole set X are both closed in X;*

(ii) *the intersection of any collection of closed sets in X is itself closed in X;*

(iii) *the union of any* finite *collection of closed sets in X is itself closed in X.*

Lemma 2.14 *Let X be a subset of $\mathbb{R}^n$, and let F be a subset of X which is closed in X. Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ be a sequence of points of F which converges to a point $\mathbf{p}$ of X. Then $\mathbf{p} \in F$.*

Proof The complement $X \setminus F$ of F in X is open, since F is closed. Suppose that $\mathbf{p}$ were a point belonging to $X \setminus F$. It would then follow from Lemma 2.12 that $\mathbf{x}_j \in X \setminus F$ for all values of j greater than some positive integer N, contradicting the fact that $\mathbf{x}_j \in F$ for all j. This contradiction shows that $\mathbf{p}$ must belong to F, as required. ∎

2.3 Continuous Functions and Open and Closed Sets

Let X and Y be subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$, and let $f\colon X \to Y$ be a function from X to Y. We recall that the function f is continuous at a point $\mathbf{p}$ of X if, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that $|f(\mathbf{u}) - f(\mathbf{p})| < \varepsilon$ for all points $\mathbf{u}$ of X satisfying $|\mathbf{u} - \mathbf{p}| < \delta$. Thus the function $f\colon X \to Y$ is continuous at $\mathbf{p}$ if and only if, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that the function f maps $B_X(\mathbf{p}, \delta)$ into $B_Y(f(\mathbf{p}), \varepsilon)$ (where $B_X(\mathbf{p}, \delta)$ and $B_Y(f(\mathbf{p}), \varepsilon)$ denote the open balls in X and Y of radius δ and ε about $\mathbf{p}$ and $f(\mathbf{p})$ respectively).

Given any function $f\colon X \to Y$, we denote by $f^{-1}(V)$ the *preimage* of a subset V of Y under the map f, defined by $f^{-1}(V) = \{\mathbf{x} \in X : f(\mathbf{x}) \in V\}$.

Proposition 2.15 *Let X and Y be subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$, and let $f: X \to Y$ be a function from X to Y. The function f is continuous if and only if $f^{-1}(V)$ is open in X for every open subset V of Y.*

Proof Suppose that $f: X \to Y$ is continuous. Let V be an open set in Y. We must show that $f^{-1}(V)$ is open in X. Let $\mathbf{p} \in f^{-1}(V)$. Then $f(\mathbf{p}) \in V$. But V is open, hence there exists some $\varepsilon > 0$ with the property that $B_Y(f(\mathbf{p}), \varepsilon) \subset V$. But f is continuous at $\mathbf{p}$. Therefore there exists some $\delta > 0$ such that f maps $B_X(\mathbf{p}, \delta)$ into $B_Y(f(\mathbf{p}), \varepsilon)$ (see the remarks above). Thus $f(\mathbf{x}) \in V$ for all $\mathbf{x} \in B_X(\mathbf{p}, \delta)$, showing that $B_X(\mathbf{p}, \delta) \subset f^{-1}(V)$. This shows that $f^{-1}(V)$ is open in X for every open set V in Y.

Conversely suppose that $f: X \to Y$ is a function with the property that $f^{-1}(V)$ is open in X for every open set V in Y. Let $\mathbf{p} \in X$. We must show that f is continuous at $\mathbf{p}$. Let $\varepsilon > 0$ be given. Then $B_Y(f(\mathbf{p}), \varepsilon)$ is an open set in Y, by Lemma 2.9, hence $f^{-1}\left(B_Y(f(\mathbf{p}), \varepsilon)\right)$ is an open set in X which contains $\mathbf{p}$. It follows that there exists some $\delta > 0$ such that $B_X(\mathbf{p}, \delta) \subset f^{-1}\left(B_Y(f(\mathbf{p}), \varepsilon)\right)$. Thus, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that f maps $B_X(\mathbf{p}, \delta)$ into $B_Y(f(\mathbf{p}), \varepsilon)$. We conclude that f is continuous at $\mathbf{p}$, as required. ∎

Let X be a subset of $\mathbb{R}^n$, let $f: X \to \mathbb{R}$ be continuous, and let c be some real number. Then the sets $\{\mathbf{x} \in X : f(\mathbf{x}) > c\}$ and $\{\mathbf{x} \in X : f(\mathbf{x}) < c\}$ are open in X, and, given real numbers a and b satisfying $a < b$, the set $\{\mathbf{x} \in X : a < f(\mathbf{x}) < b\}$ is open in X.

Let X and Y be subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$. Any homeomorphism $h: X \to Y$ induces a one-to-one correspondence between the open sets of X and the open sets of Y: a subset V of Y is open in Y if and only if $h^{-1}(V)$ is open in X. This result follows immediately on applying Proposition 2.15 to $h: X \to Y$ and its inverse $h^{-1}: Y \to X$.

2.4 Continuous Functions on Closed Bounded Sets

We shall prove that continuous functions are bounded on closed bounded sets in $\mathbb{R}^n$. First we prove an n-dimensional generalization of the Bolzano-Weierstrass Theorem.

A sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ of points in $\mathbb{R}^n$ is said to be *bounded* if there exists some constant K such that $|\mathbf{x}_j| \leq K$ for all j.

Theorem 2.16 *Every bounded sequence of points in $\mathbb{R}^n$ has a convergent subsequence.*

Proof Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ be a bounded sequence in $\mathbb{R}^n$. The first components of the elements of the sequence constitute a bounded sequence of real numbers. It follows from the Bolzano-Weierstrass Theorem (Theorem 1.4) that there exists a subsequence of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ with the property that the first components of the elements of the subsequence converge to some real number p_1. We can then extract from this subsequence a further subsequence with the property that the second components of the elements of the new subsequence converge to some real number p_2. By proceeding in this fashion (i.e., replacing subsequences of the original sequence with subsequences of themselves) one can extract a subsequence $\mathbf{x}_{n_1}, \mathbf{x}_{n_2}, \mathbf{x}_{n_3}, \ldots$ of the original sequence with the property that the ith components of $\mathbf{x}_{n_j}$ converge to some real number p_i as $j \to +\infty$. It then follows from Lemma 2.2 that the sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ converges to $\mathbf{p}$, where $\mathbf{p} = (p_1, p_2, \ldots, p_n)$. ∎

Theorem 2.17 *Let X be a closed bounded set in $\mathbb{R}^n$, and let $f: X \to \mathbb{R}$ be a continuous function on X. Then there exists some constant K such that $|f(\mathbf{x})| \leq K$ for all $\mathbf{x} \in X$.*

Proof Suppose that the function were not bounded on the set X. Then there would exist a sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ of points of X such that $|f(\mathbf{x}_j)| > j$ for $j = 1, 2, 3, \ldots$. This sequence would be bounded, since X is bounded, and would therefore possess a convergent subsequence $\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \mathbf{x}_{j_3}, \ldots$, by Theorem 2.16). Moreover the limit $\mathbf{p}$ of this subsequence would belong to X by Lemma 2.14, since X is closed. But then $f(\mathbf{x}_{j_k}) \to f(\mathbf{p})$ as $k \to +\infty$, by Lemma 2.5. But this leads to a contradiction, since $|f(\mathbf{x}_{j_k})| > j_k$ for all k, so that $|f(\mathbf{x}_{j_k})|$ increases without limit as $k \to +\infty$, whereas any convergent sequence of real numbers is bounded. This contradiction shows that the function f is bounded on the set X, as required. ∎

Corollary 2.18 *Let X be a closed bounded set in $\mathbb{R}^n$, and let $f: X \to \mathbb{R}$ be a continuous function on X. Then there exist points $\mathbf{u}$ and $\mathbf{v}$ in X with the property that $f(\mathbf{u}) \leq f(\mathbf{x}) \leq f(\mathbf{v})$ for all $\mathbf{x} \in X$.*

Proof Let $C = \sup\{f(\mathbf{x}) : \mathbf{x} \in X\}$. If there did not exist any $\mathbf{v} \in X$ for which $f(\mathbf{v}) = C$ then the function $\mathbf{x} \mapsto 1/(C - f(\mathbf{x}))$ would be a continuous function on the set X which was not bounded above on this set, thus contradicting Theorem 2.17. Thus there must exist some $\mathbf{v} \in X$ with the property that $f(\mathbf{v}) = C$. A similar proof shows that there must exist some $\mathbf{u} \in X$ with the property that $g(\mathbf{u}) = c$, where $c = \inf\{f(\mathbf{x}) : \mathbf{x} \in X\}$. But then $f(\mathbf{u}) \leq f(\mathbf{x}) \leq f(\mathbf{v})$ for all $\mathbf{x} \in X$, as required. ∎

3 Metric Spaces

Definition A *metric space* (X, d) consists of a set X together with a *distance function* $d: X \times X \to [0, +\infty)$ on X satisfying the following axioms:

(i) $d(x, y) \geq 0$ for all $x, y \in X$,

(ii) $d(x, y) = d(y, x)$ for all $x, y \in X$,

(iii) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$,

(iv) $d(x, y) = 0$ if and only if $x = y$.

The quantity $d(x, y)$ should be thought of as measuring the *distance* between the points x and y. The inequality $d(x, z) \leq d(x, y) + d(y, z)$ is referred to as the *Triangle Inequality.* The elements of a metric space are usually referred to as *points* of that metric space.

Note that if X is a metric space with distance function d and if A is a subset of X then the restriction $d|A \times A$ of d to pairs of points of A defines a distance function on A satisfying the axioms for a metric space.

The set $\mathbb{R}$ of real numbers becomes a metric space with distance function d given by $d(x, y) = |x - y|$ for all $x, y \in \mathbb{R}$. Similarly the set $\mathbb{C}$ of complex numbers becomes a metric space with distance function d given by $d(z, w) = |z - w|$ for all $z, w \in \mathbb{C}$, and n-dimensional Euclidean space $\mathbb{R}^n$ is a metric space with with respect to the *Euclidean distance function* d, given by

$$d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Any subset X of $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{R}^n$ may be regarded as a metric space whose distance function is the restriction to X of the distance function on $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{R}^n$ defined above.

Example The *n-sphere* S^n is defined to be the subset of $(n+1)$-dimensional Euclidean space $\mathbb{R}^{n+1}$ consisting of all elements $\mathbf{x}$ of $\mathbb{R}^{n+1}$ for which $|\mathbf{x}| = 1$. Thus

$$S^n = \{(x_1, x_2, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : x_1^2 + x_2^2 + \cdots + x_{n+1}^2 = 1\}.$$

(Note that S^2 is the standard (2-dimensional) unit sphere in 3-dimensional Euclidean space.) The *chordal distance* between two points $\mathbf{x}$ and $\mathbf{y}$ of S^n is defined to be the length $|\mathbf{x} - \mathbf{y}|$ of the line segment joining $\mathbf{x}$ and $\mathbf{y}$. The n-sphere S^n is a metric space with respect to the chordal distance function.

Example Let $C([a, b])$ denote the set of all continuous real-valued functions on the closed interval $[a, b]$, where a and b are real numbers satisfying $a < b$. Then $C([a, b])$ is a metric space with respect to the distance function d, where $d(f, g) = \sup_{t \in [a,b]} |f(t) - g(t)|$ for all continuous functions f and g from X to $\mathbb{R}$. (Note that, for all $f, g \in C([a, b])$, $f - g$ is a continuous function on $[a, b]$ and is therefore bounded on $[a, b]$. Therefore the distance function d is well-defined.)

3.1 Convergence and Continuity in Metric Spaces

Definition Let X be a metric space with distance function d. A sequence $x_1, x_2, x_3, \ldots$ of points in X is said to *converge* to a point p in X if and only if the following criterion is satisfied:—

- given any real number ε satisfying $\varepsilon > 0$, there exists some natural number N such that $d(x_n, p) < \varepsilon$ whenever $n \geq N$.

We refer to p as the *limit* $\lim\limits_{n \to +\infty} x_n$ of the sequence $x_1, x_2, x_3, \ldots$.

Note that this definition of convergence generalizes to arbitrary metric spaces the standard definition of convergence for sequences of real or complex numbers.

If a sequence of points in a metric space is convergent then the limit of that sequence is unique. Indeed let $x_1, x_2, x_3, \ldots$ be a sequence of points in a metric space (X, d) which converges to points p and p' of X. We show that $p = p'$. Now, given any $\varepsilon > 0$, there exist natural numbers N_1 and N_2 such that $d(x_n, p) < \varepsilon$ whenever $n \geq N_1$ and $d(x_n, p') < \varepsilon$ whenever $n \geq N_2$. On choosing n so that $n \geq N_1$ and $n \geq N_2$ we see that

$$0 \leq d(p, p') \leq d(p, x_n) + d(x_n, p') < 2\varepsilon$$

by a straightforward application of the metric space axioms (i)–(iii). Thus $0 \leq d(p, p') < 2\varepsilon$ for every $\varepsilon > 0$, and hence $d(p, p') = 0$, so that $p = p'$ by Axiom (iv).

Lemma 3.1 *Let (X, d) be a metric space, and let $x_1, x_2, x_3, \ldots$ be a sequence of points of X which converges to some point p of X. Then, for any point y of X, $d(x_n, y) \to d(p, y)$ as $n \to +\infty$.*

Proof Let $\varepsilon > 0$ be given. We must show that there exists some natural number N such that $|d(x_n, y) - d(p, y)| < \varepsilon$ whenever $n \geq N$. However N can be chosen such that $d(x_n, p) < \varepsilon$ whenever $n \geq N$. But

$$d(x_n, y) \leq d(x_n, p) + d(p, y), \qquad d(p, y) \leq d(p, x_n) + d(x_n, y)$$

for all n, hence

$$-d(x_n, p) \leq d(x_n, y) - d(p, y) \leq d(x_n, p)$$

for all n, and hence $|d(x_n, y) - d(p, y)| < \varepsilon$ whenever $n \geq N$, as required. ∎

Definition Let X and Y be metric spaces with distance functions d_X and d_Y respectively. A function $f\colon X \to Y$ from X to Y is said to be *continuous* at a point x of X if and only if the following criterion is satisfied:—

- given any real number ε satisfying $\varepsilon > 0$ there exists some $\delta > 0$ such that $d_Y(f(x), f(x')) < \varepsilon$ for all points x' of X satisfying $d_X(x, x') < \delta$.

The function $f\colon X \to Y$ is said to be continuous on X if and only if it is continuous at x for every point x of X.

Note that this definition of continuity for functions between metric spaces generalizes the definition of continuity for functions of a real or complex variable.

Lemma 3.2 *Let X, Y and Z be metric spaces, and let $f\colon X \to Y$ and $g\colon Y \to Z$ be continuous functions. Then the composition function $g \circ f\colon X \to Z$ is continuous.*

Proof We denote by d_X, d_Y and d_Z the distance functions on X, Y and Z respectively. Let x be any point of X. We show that $g \circ f$ is continuous at x. Let $\varepsilon > 0$ be given. Now the function g is continuous at $f(x)$. Hence there exists some $\eta > 0$ such that $d_Z(g(y), g(f(x))) < \varepsilon$ for all $y \in Y$ satisfying $d_Y(y, f(x)) < \eta$. But then there exists some $\delta > 0$ such that $d_Y(f(x'), f(x)) < \eta$ for all $x' \in X$ satisfying $d_X(x', x) < \delta$. Thus $d_Z(g(f(x')), g(f(x))) < \varepsilon$ for all $x' \in X$ satisfying $d_X(x', x) < \delta$, showing that $g \circ f$ is continuous at x, as required. ∎

Lemma 3.3 *Let $f\colon X \to Y$ be a continuous function between metric spaces X and Y, and let $x_1, x_2, x_3, \ldots$ be a sequence of points in X which converges to some point p of X. Then the sequence $f(x_1), f(x_2), f(x_3), \ldots$ converges to $f(p)$.*

Proof We denote by d_X and d_Y the distance functions on X and Y respectively. Let $\varepsilon > 0$ be given. We must show that there exists some natural number N such that $d_Y(f(x_n), f(p)) < \varepsilon$ whenever $n \geq N$. However there exists some $\delta > 0$ such that $d_Y(f(x'), f(p)) < \varepsilon$ for all $x' \in X$ satisfying $d_X(x', p) < \delta$, since the function f is continuous at p. Also there exists some natural number N such that $d_X(x_n, p) < \delta$ whenever $n \geq N$, since the sequence $x_1, x_2, x_3, \ldots$ converges to p. Thus if $n \geq N$ then $d_Y(f(x_n), f(p)) < \varepsilon$, as required. ∎

3.2 Open Sets in Metric Spaces

Definition Let (X, d) be a metric space. Given a point x of X and $r \geq 0$, the *open ball* $B_X(x, r)$ of *radius* r about x in X is defined by

$$B_X(x, r) \equiv \{x' \in X : d(x', x) < r\}.$$

Definition Let (X, d) be a metric space. A subset V of X is said to be an *open set* if and only if the following condition is satisfied:

- given any point v of V there exists some $\delta > 0$ such that $B_X(v, \delta) \subset V$.

By convention, we regard the empty set $\emptyset$ as being an open subset of X. (The criterion given above is satisfied vacuously in this case.)

Lemma 3.4 *Let X be a metric space with distance function d, and let x_0 be a point of X. Then, for any $r > 0$, the open ball $B_X(x_0, r)$ of radius r about x_0 is an open set in X.*

Proof Let $x \in B_X(x_0, r)$. We must show that there exists some $\delta > 0$ such that $B_X(x, \delta) \subset B_X(x_0, r)$. Now $d(x, x_0) < r$, and hence $\delta > 0$, where $\delta = r - d(x, x_0)$. Moreover if $x' \in B_X(x, \delta)$ then

$$d(x', x_0) \leq d(x', x) + d(x, x_0) < \delta + d(x, x_0) = r,$$

by the Triangle Inequality, hence $x' \in B_X(x_0, r)$. Thus $B_X(x, \delta) \subset B_X(x_0, r)$, showing that $B_X(x_0, r)$ is an open set, as required. ∎

Lemma 3.5 *Let X be a metric space with distance function d, and let x_0 be a point of X. Then, for any $r \geq 0$, the set $\{x \in X : d(x, x_0) > r\}$ is an open set in X.*

Proof Let x be a point of X satisfying $d(x, x_0) > r$, and let x' be any point of X satisfying $d(x', x) < \delta$, where $\delta = d(x, x_0) - r$. Then

$$d(x, x_0) \leq d(x, x') + d(x', x_0),$$

by the Triangle Inequality, and therefore

$$d(x', x_0) \geq d(x, x_0) - d(x, x') > d(x, x_0) - \delta = r.$$

Thus $B_X(x, \delta) \subset \{x' \in X : d(x', x_0) > r\}$, as required. ∎

Proposition 3.6 *Let X be a metric space. The collection of open sets in X has the following properties:—*

(i) *the empty set* $\emptyset$ *and the whole set* X *are both open sets;*

(ii) *the union of any collection of open sets is itself an open set;*

(iii) *the intersection of any* finite *collection of open sets is itself an open set.*

Proof The empty set $\emptyset$ is an open set by convention. Moreover the definition of an open set is satisfied trivially by the whole set X. Thus (i) is satisfied.

Let $\mathcal{A}$ be any collection of open sets in X, and let U denote the union of all the open sets belonging to $\mathcal{A}$. We must show that U is itself an open set. Let $x \in U$. Then $x \in V$ for some open set V belonging to the collection $\mathcal{A}$. Therefore there exists some $\delta > 0$ such that $B_X(x, \delta) \subset V$. But $V \subset U$, and thus $B_X(x, \delta) \subset U$. This shows that U is open. Thus (ii) is satisfied.

Finally let $V_1, V_2, V_3, \ldots, V_k$ be a *finite* collection of open sets in X, and let $V = V_1 \cap V_2 \cap \cdots \cap V_k$. Let $x \in V$. Now $x \in V_j$ for all j, and therefore there exist strictly positive real numbers $\delta_1, \delta_2, \ldots, \delta_k$ such that $B_X(x, \delta_j) \subset V_j$ for $j = 1, 2, \ldots, k$. Let δ be the minimum of $\delta_1, \delta_2, \ldots, \delta_k$. Then $\delta > 0$. (This is where we need the fact that we are dealing with a finite collection of open sets.) Moreover $B_X(x, \delta) \subset B_X(x, \delta_j) \subset V_j$ for $j = 1, 2, \ldots, k$, and thus $B_X(x, \delta) \subset V$. This shows that the intersection V of the open sets $V_1, V_2, \ldots, V_k$ is itself open. Thus (iii) is satisfied. ∎

Remark For each natural number n, let V_n denote the open set in the complex plane $\mathbb{C}$ defined by

$$V_n = \{z \in \mathbb{C} : |z| < 1/n\}.$$

The intersection of all of these sets (as n ranges over the set of natural numbers) consists of the set $\{0\}$, and this set is not an open subset of the complex plane. This demonstrates that an intersection of an infinite number of open sets in a metric space is not necessarily an open set.

Lemma 3.7 *Let X be a metric space. A sequence $x_1, x_2, x_3, \ldots$ of points in X converges to a point p if and only if, given any open set U which contains p, there exists some natural number N such that $x_j \in U$ for all $j \geq N$.*

Proof Let $x_1, x_2, x_3, \ldots$ be a sequence satisfying the given criterion, and let $\varepsilon > 0$ be given. The open ball $B_X(p, \varepsilon)$ of radius ε about p is an open set (see Lemma 3.4). Therefore there exists some natural number N such that, if $j \geq N$, then $x_j \in B_X(p, \varepsilon)$, and thus $d(x_j, p) < \varepsilon$. Hence the sequence (x_j) converges to p.

Conversely, suppose that the sequence (x_j) converges to p. Let U be an open set which contains p. Then there exists some $\varepsilon > 0$ such that $B_X(p, \varepsilon) \subset U$. But $x_j \to p$ as $j \to +\infty$, and therefore there exists some natural number N such that $d(x_j, p) < \varepsilon$ for all $j \geq N$. If $j \geq N$ then $x_j \in B_X(p, \varepsilon)$ and thus $x_j \in U$, as required. ∎

Definition Let (X, d) be a metric space, and let x be a point of X. A subset N of X is said to be a *neighbourhood* of x (in X) if and only if there exists some $\delta > 0$ such that $B_X(x, \delta) \subset N$, where $B_X(x, \delta)$ is the open ball of radius δ about x.

It follows directly from the relevant definitions that a subset V of a metric space X is an open set if and only if V is a neighbourhood of v for all $v \in V$.

3.3 Closed Sets in a Metric Space

A subset F of a metric space X is said to be a *closed set* in X if and only if its complement $X \setminus F$ is open. (Recall that the *complement* $X \setminus F$ of F in X is, by definition, the set of all points of the metric space X that do not belong to F.) The following result follows immediately from Lemma 3.4 and Lemma 3.5.

Lemma 3.8 *Let X be a metric space with distance function d, and let $x_0 \in X$. Given any $r \geq 0$, the sets*

$$\{x \in X : d(x, x_0) \leq r\}, \qquad \{x \in X : d(x, x_0) \geq r\}$$

are closed. In particular, the set $\{x_0\}$ consisting of the single point x_0 is a closed set in X.

Let $\mathcal{A}$ be some collection of subsets of a set X. Then

$$X \setminus \bigcup_{S \in \mathcal{A}} S = \bigcap_{S \in \mathcal{A}} (X \setminus S), \qquad X \setminus \bigcap_{S \in \mathcal{A}} S = \bigcup_{S \in \mathcal{A}} (X \setminus S)$$

(i.e., the complement of the union of some collection of subsets of X is the intersection of the complements of those sets, and the complement of the intersection of some collection of subsets of X is the union of the complements of those sets, so that the operation of taking complements converts unions into intersections and intersections into unions). The following result therefore follows directly from Proposition 3.6.

Proposition 3.9 *Let X be a metric space. The collection of closed sets in X has the following properties:—*

(i) *the empty set* $\emptyset$ *and the whole set* X *are both closed sets;*

(ii) *the intersection of any collection of closed sets in* X *is itself a closed set;*

(iii) *the union of any* finite *collection of closed sets in* X *is itself a closed set.*

Lemma 3.10 *Let* F *be a closed set in a metric space* X *and let* $(x_j : j \in \mathbb{N})$ *be a sequence of points of* F. *Suppose that* $x_j \to p$ *as* $j \to +\infty$. *Then* p *also belongs to* F.

Proof Suppose that the limit p of the sequence were to belong to the complement $X \setminus F$ of the closed set F. Now $X \setminus F$ is open, and thus it would follow from Lemma 3.7 that there would exist some natural number N such that $x_j \in X \setminus F$ for all $j \geq N$, contradicting the fact that $x_j \in F$ for all j. This contradiction shows that p must belong to F, as required. ∎

Definition Let A be a subset of a metric space X. The *closure* $\overline{A}$ of A is the intersection of all closed subsets of X containing A.

Let A be a subset of the metric space X. Note that the closure $\overline{A}$ of A is itself a closed set in X, since the intersection of any collection of closed subsets of X is itself a closed subset of X (see Proposition 3.9). Moreover if F is any closed subset of X, and if $A \subset F$, then $\overline{A} \subset F$. Thus the closure $\overline{A}$ of A is the smallest closed subset of X containing A.

Lemma 3.11 *Let* X *be a metric space with distance function* d, *let* A *be a subset of* X, *and let* x *be a point of* X. *Then* x *belongs to the closure* $\overline{A}$ *of* A *if and only if, given any* $\varepsilon > 0$, *there exists some point* a *of* A *such that* $d(x, a) < \varepsilon$.

Proof Let x be a point of X with the property that, given any $\varepsilon > 0$, there exists some $a \in A$ satisfying $d(x, a) < \varepsilon$. Let F be any closed subset of X containing A. If x did not belong to F then there would exist some $\varepsilon > 0$ with the property that $B_X(x, \varepsilon) \cap F = \emptyset$, where $B_X(x, \varepsilon)$ denotes the open ball of radius ε about x. But this would contradict the fact that $B_X(x, \varepsilon) \cap A$ is non-empty for all $\varepsilon > 0$. Thus the point x belongs to every closed subset F of X that contains A, and therefore $x \in \overline{A}$, by definition of the closure $\overline{A}$ of A.

Conversely let $x \in \overline{A}$, and let $\varepsilon > 0$ be given. Let F be the complement $X \setminus B_X(x, \varepsilon)$ of $B_X(x, \varepsilon)$. Then F is a closed subset of X, and the point x does not belong to F. If $B_X(x, \varepsilon) \cap A = \emptyset$ then A would be contained in F, and hence $x \in F$, which is impossible. Therefore there exists $a \in A$ satisfying $d(x, a) < \varepsilon$, as required. ∎

3.4 Continuous Functions and Open and Closed Sets

Let X and Y be metric spaces, and let $f\colon X \to Y$ be a function from X to Y. We recall that the function f is continuous at a point x of X if and only if, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that $d_Y(f(x'), f(x)) < \varepsilon$ for all points x' of X satisfying $d_X(x', x) < \delta$, where d_X and d_Y denote the distance functions on X and Y respectively. Expressed in terms of open balls, this means that the function $f\colon X \to Y$ is continuous at x if and only if, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that f maps $B_X(x, \delta)$ into $B_Y(f(x), \varepsilon)$ (where $B_X(x, \delta)$ and $B_Y(f(x), \varepsilon)$ denote the open balls of radius δ and ε about x and $f(x)$ respectively).

Let $f\colon X \to Y$ be a function from a set X to a set Y. Given any subset V of Y, we denote by $f^{-1}(V)$ the *preimage* of V under the map f, defined by

$$f^{-1}(V) = \{x \in X : f(x) \in V\}.$$

Proposition 3.12 *Let X and Y be metric spaces, and let $f\colon X \to Y$ be a function from X to Y. The function f is continuous if and only if $f^{-1}(V)$ is an open set in X for every open set V of Y.*

Proof Suppose that $f\colon X \to Y$ is continuous. Let V be an open set in Y. We must show that $f^{-1}(V)$ is open in X. Let x be a point belonging to $f^{-1}(V)$. We must show that there exists some $\delta > 0$ with the property that $B_X(x, \delta) \subset f^{-1}(V)$. Now $f(x)$ belongs to V. But V is open, hence there exists some $\varepsilon > 0$ with the property that $B_Y(f(x), \varepsilon) \subset V$. But f is continuous at x. Therefore there exists some $\delta > 0$ such that f maps the open ball $B_X(x, \delta)$ into $B_Y(f(x), \varepsilon)$ (see the remarks above). Thus $f(x') \in V$ for all $x' \in B_X(x, \delta)$, showing that $B_X(x, \delta) \subset f^{-1}(V)$. We have thus shown that if $f\colon X \to Y$ is continuous then $f^{-1}(V)$ is open in X for every open set V in Y.

Conversely suppose that $f\colon X \to Y$ has the property that $f^{-1}(V)$ is open in X for every open set V in Y. Let x be any point of X. We must show that f is continuous at x. Let $\varepsilon > 0$ be given. The open ball $B_Y(f(x), \varepsilon)$ is an open set in Y, by Lemma 3.4, hence $f^{-1}\left(B_Y(f(x), \varepsilon)\right)$ is an open set in X which contains x. It follows that there exists some $\delta > 0$ such that $B_X(x, \delta) \subset f^{-1}\left(B_Y(f(x), \varepsilon)\right)$. We have thus shown that, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that f maps the open ball $B_X(x, \delta)$ into $B_Y(f(x), \varepsilon)$. We conclude that f is continuous at x, as required. ∎

Let $f\colon X \to Y$ be a function between metric spaces X and Y. Then the preimage $f^{-1}(Y \setminus G)$ of the complement $Y \setminus G$ of any subset G of Y is equal to the complement $X \setminus f^{-1}(G)$ of the preimage $f^{-1}(G)$ of G. Indeed

$$x \in f^{-1}(Y \setminus G) \iff f(x) \in Y \setminus G \iff f(x) \notin G \iff x \notin f^{-1}(G).$$

Also a subset of a metric space is closed if and only if its complement is open. The following result therefore follows directly from Proposition 3.12.

Corollary 3.13 *Let X and Y be metric spaces, and let $f\colon X \to Y$ be a function from X to Y. The function f is continuous if and only if $f^{-1}(G)$ is a closed set in X for every closed set G in Y.*

Let $f\colon X \to Y$ be a continuous function from a metric space X to a metric space Y. Then, for any point y of Y, the set $\{x \in X : f(x) = y\}$ is a closed subset of X. This follows from Corollary 3.13, together with the fact that the set $\{y\}$ consisting of the single point y is a closed subset of the metric space Y.

Let X be a metric space, and let $f\colon X \to \mathbb{R}$ be a continuous function from X to $\mathbb{R}$. Then, given any real number c, the sets

$$\{x \in X : f(x) > c\}, \qquad \{x \in X : f(x) < c\}$$

are open subsets of X, and the sets

$$\{x \in X : f(x) \geq c\}, \qquad \{x \in X : f(x) \leq c\}, \qquad \{x \in X : f(x) = c\}$$

are closed subsets of X. Also, given real numbers a and b satisfying $a < b$, the set

$$\{x \in X : a < f(x) < b\}$$

is an open subset of X, and the set

$$\{x \in X : a \leq f(x) \leq b\}$$

is a closed subset of X.

Similar results hold for continuous functions $f\colon X \to \mathbb{C}$ from X to $\mathbb{C}$. Thus, for example,

$$\{x \in X : |f(x)| < R\}, \qquad \{x \in X : |f(x)| > R\}$$

are open subsets of X and

$$\{x \in X : |f(x)| \leq R\}, \qquad \{x \in X : |f(x)| \geq R\}, \qquad \{x \in X : |f(x)| = R\}$$

are closed subsets of X, for any non-negative real number R.

3.5 Homeomorphisms

Let X and Y be metric spaces. A function $h: X \to Y$ from X to Y is said to be a *homeomorphism* if it is a bijection and both $h: X \to Y$ and its inverse $h^{-1}: Y \to X$ are continuous. If there exists a homeomorphism $h: X \to Y$ from a metric space X to a metric space Y, then the metric spaces X and Y are said to be *homeomorphic*.

The following result follows directly on applying Proposition 3.12 to $h: X \to Y$ and to $h^{-1}: Y \to X$.

Lemma 3.14 *Any homeomorphism $h: X \to Y$ between metric spaces X and Y induces a one-to-one correspondence between the open sets of X and the open sets of Y: a subset V of Y is open in Y if and only if $h^{-1}(V)$ is open in X.*

Let X and Y be metric spaces, and let $h: X \to Y$ be a homeomorphism. A sequence $x_1, x_2, x_3, \ldots$ of points in X is convergent in X if and only if the corresponding sequence $h(x_1), h(x_2), h(x_3), \ldots$ is convergent in Y. (This follows directly on applying Lemma 3.3 to $h: X \to Y$ and its inverse $h^{-1}: Y \to X$.) Let Z and W be metric spaces. A function $f: Z \to X$ is continuous if and only if $h \circ f: Z \to Y$ is continuous, and a function $g: Y \to W$ is continuous if and only if $g \circ h: X \to W$ is continuous.

3.6 Complete Metric Spaces

Definition Let X be a metric space with distance function d. A sequence $x_1, x_2, x_3, \ldots$ of points of X is said to be a *Cauchy sequence* in X if and only if, given any $\varepsilon > 0$, there exists some natural number N such that $d(x_j, x_k) < \varepsilon$ for all j and k satisfying $j \geq N$ and $k \geq N$.

Every convergent sequence in a metric space is a Cauchy sequence. Indeed let X be a metric space with distance function d, and let $x_1, x_2, x_3, \ldots$ be a sequence of points in X which converges to some point p of X. Given any $\varepsilon > 0$, there exists some natural number N such that $d(x_n, p) < \varepsilon/2$ whenever $n \geq N$. But then it follows from the Triangle Inequality that

$$d(x_j, x_k) \leq d(x_j, p) + d(p, x_k) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

whenever $j \geq N$ and $k \geq N$.

Definition A metric space (X, d) is said to be *complete* if every Cauchy sequence in X converges to some point of X.

The spaces $\mathbb{R}$ and $\mathbb{C}$ are complete metric spaces with respect to the distance function given by $d(z, w) = |z - w|$. Indeed this result is *Cauchy's Criterion for Convergence*. However the space $\mathbb{Q}$ of rational numbers (with distance function $d(q, r) = |q - r|$) is not complete. Indeed one can construct an infinite sequence $q_1, q_2, q_3, \ldots$ of rational numbers which converges (in $\mathbb{R}$) to $\sqrt{2}$. Such a sequence of rational numbers is a Cauchy sequence in both $\mathbb{R}$ and $\mathbb{Q}$. However this Cauchy sequence does not converge to an point of the metric space $\mathbb{Q}$ (since $\sqrt{2}$ is an irrational number). Thus the metric space $\mathbb{Q}$ is not complete.

Lemma 3.15 *Let X be a complete metric space, and let A be a subset of X. Then A is complete if and only if A is closed in X.*

Proof Suppose that A is closed in X. Let $a_1, a_2, a_3, \ldots$ be a Cauchy sequence in A. This Cauchy sequence must converge to some point p of X, since X is complete. But the limit of every sequence of points of A must belong to A, since A is closed (see Lemma 3.10). In particular $p \in A$. We deduce that A is complete.

Conversely, suppose that A is complete. Suppose that A were not closed. Then the complement $X \setminus A$ of A would not be open, and therefore there would exist a point p of $X \setminus A$ with the property that $B_X(p, \delta) \cap A$ is non-empty for all $\delta > 0$, where $B_X(p, \delta)$ denotes the open ball in X of radius δ centred at p. We could then find a sequence $a_1, a_2, a_3, \ldots$ of points of A satisfying $d(a_j, p) < 1/j$ for all natural numbers j. This sequence would be a Cauchy sequence in A which did not converge to a point of A, contradicting the completeness of A. Thus A must be closed, as required. ∎

Theorem 3.16 *The metric space $\mathbb{R}^n$ (with the Euclidean distance function) is a complete metric space.*

Proof Let $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \ldots$ be a Cauchy sequence in $\mathbb{R}^n$. Then for each integer m between 1 and n, the sequence $(\mathbf{p}_1)_m, (\mathbf{p}_2)_m, (\mathbf{p}_3)_m, \ldots$ is a Cauchy sequence of real numbers, where $(\mathbf{p}_j)_m$ denotes the mth component of $\mathbf{p}_j$. But every Cauchy sequence of real numbers is convergent (Cauchy's criterion for convergence). Let $q_m = \lim_{j \to +\infty} (\mathbf{p}_j)_m$ for $m = 1, 2, \ldots, n$, and let $\mathbf{q} = (q_1, q_2, \ldots, q_n)$. We claim that $\mathbf{p}_j \to \mathbf{q}$ as $j \to +\infty$.

Let $\varepsilon > 0$ be given. Then there exist natural numbers $N_1, N_2, \ldots, N_n$ such that $|(\mathbf{p}_j)_m - q_m| < \varepsilon/\sqrt{n}$ whenever $j \geq N_m$ (where $m = 1, 2, \ldots, n$). Let N be the maximum of $N_1, N_2, \ldots, N_n$. If $j \geq N$ then

$$|\mathbf{p}_j - \mathbf{q}|^2 = \sum_{m=1}^{n} ((\mathbf{p}_j)_m - q_m)^2 < \varepsilon^2.$$

Thus $\mathbf{p}_j \to \mathbf{q}$ as $j \to +\infty$. Thus every Cauchy sequence in $\mathbb{R}^n$ is convergent, as required. ∎

The following result follows directly from Lemma 3.15 and Theorem 3.16.

Corollary 3.17 *A subset X of $\mathbb{R}^n$ is complete if and only if it is closed.*

Example The n-sphere S^n (with the chordal distance function given by $d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|$) is a complete metric space, where

$$S^n = \{(x_1, x_2, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : x_1^2 + x_2^2 + \cdots + x_{n+1}^2 = 1\}.$$

3.7 The Completion of a Metric Space

We describe below a construction whereby any metric space can be embedded in a complete metric space.

Lemma 3.18 *Let X be a metric space with distance function d, let (x_j) and (y_j) be Cauchy sequences of points in X, and let $d_j = d(x_j, y_j)$ for all natural numbers j. Then (d_j) is a Cauchy sequence of real numbers.*

Proof It follows from the Triangle Inequality that

$$d_j \leq d(x_j, x_k) + d_k + d(y_k, y_j)$$

and thus $d_j - d_k \leq d(x_j, x_k) + d(y_j, y_k)$ for all integers j and k. Similarly $d_k - d_j \leq d(x_j, x_k) + d(y_j, y_k)$. It follows that

$$|d_j - d_k| \leq d(x_j, x_k) + d(y_j, y_k)$$

for all integers j and k.

Let $\varepsilon > 0$ be given. Then there exists some natural number N such that $d(x_j, x_k) < \frac{1}{2}\varepsilon$ and $d(y_j, y_k) < \frac{1}{2}\varepsilon$ whenever $j \geq N$ and $k \geq N$, since the sequences (x_j) and (y_j) are Cauchy sequences in X. But then $|d_j - d_k| < \varepsilon$ whenever $j \geq N$ and $k \geq N$. Thus the sequence (d_j) is a Cauchy sequence of real numbers, as required. ∎

Let X be a metric space with distance function d. It follows from Cauchy's Criterion for Convergence and Lemma 3.18 that $\lim\limits_{j \to +\infty} d(x_j, y_j)$ exists for all Cauchy sequences (x_j) and (y_j) in X.

Lemma 3.19 *Let X be a metric space with distance function d, and let (x_j), (y_j) and (z_j) be Cauchy sequences of points in X. Then*

$$0 \leq \lim_{j \to +\infty} d(x_j, z_j) \leq \lim_{j \to +\infty} d(x_j, y_j) + \lim_{j \to +\infty} d(y_j, z_j).$$

Proof This follows immediately on taking limits of both sides of the Triangle Inequality. ∎

Lemma 3.20 *Let X be a metric space with distance function d, and let (x_j), (y_j) and (z_j) be Cauchy sequences of points in X. Suppose that*

$$\lim_{j\to+\infty} d(x_j, y_j) = 0 \text{ and } \lim_{j\to+\infty} d(y_j, z_j) = 0.$$

Then $\lim_{j\to+\infty} d(x_j, z_j) = 0$.

Proof This is an immediate consequence of Lemma 3.19. ∎

Lemma 3.21 *Let X be a metric space with distance function d, and let (x_j), (x'_j), (y_j) and (y'_j) be Cauchy sequences of points in X. Suppose that $\lim_{j\to+\infty} d(x_j, x'_j) = 0$ and $\lim_{j\to+\infty} d(y_j, y'_j) = 0$. Then $\lim_{j\to+\infty} d(x_j, y_j) = \lim_{j\to+\infty} d(x'_j, y'_j)$.*

Proof It follows from Lemma 3.19 that

$$\begin{aligned}
\lim_{j\to+\infty} d(x_j, y_j) &\leq \lim_{j\to+\infty} d(x_j, x'_j) + \lim_{j\to+\infty} d(x'_j, y'_j) + \lim_{j\to+\infty} d(y'_j, y_j) \\
&= \lim_{j\to+\infty} d(x'_j, y'_j).
\end{aligned}$$

Similarly $\lim_{j\to+\infty} d(x'_j, y'_j) \leq \lim_{j\to+\infty} d(x_j, y_j)$. It follows that $\lim_{j\to+\infty} d(x_j, y_j) = \lim_{j\to+\infty} d(x'_j, y'_j)$, as required. ∎

Let X be a metric space with distance function d. Then there is an equivalence relation on the set of Cauchy sequences of points in X, where two Cauchy sequences (x_j) and (x'_j) in X are equivalent if and only if $\lim_{j\to+\infty} d(x_j, x'_j) = 0$. Let $\tilde{X}$ denote the set of equivalence classes of Cauchy sequences in X with respect to this equivalence relation. Let $\tilde{x}$ and $\tilde{y}$ be elements of $\tilde{X}$, and let (x_j) and (y_j) be Cauchy sequences belonging to the equivalence classes represented by $\tilde{x}$ and $\tilde{y}$. We define

$$d(\tilde{x}, \tilde{y}) = \lim_{j\to+\infty} d(x_j, y_j).$$

It follows from Lemma 3.21 that the value $d(\tilde{x}, \tilde{y})$ does not depend on the choice of Cauchy sequences (x_j) and (y_j) representing $\tilde{x}$ and $\tilde{y}$. We obtain in this way a distance function on the set $\tilde{X}$. This distance function satisfies the Triangle Inequality (Lemma 3.19) and the other metric space axioms.

Therefore $\tilde{X}$ with this distance function is a metric space. We refer to the space $\tilde{X}$ as the *completion* of the metric space X.

We can regard the metric space X as being embedded in its completion $\tilde{X}$, where a point x of X is represented in $\tilde{X}$ by the equivalence class of the constant sequence $x, x, x, \ldots$.

Example The completion of the space $\mathbb{Q}$ of rational numbers is the space $\mathbb{R}$ of real numbers.

Theorem 3.22 *The completion $\tilde{X}$ of a metric space X is a complete metric space.*

Proof Let $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \ldots$ be a Cauchy sequence in the completion $\tilde{X}$ of X. For each positive integer m let $x_{m,1}, x_{m,2}, x_{m,3}, \ldots$ be a Cauchy sequence in X belonging to the equivalence class that represents the element $\tilde{x}_m$ of $\tilde{X}$. Then, for each positive integer m there exists a natural number $N(m)$ such that $d(x_{m,j}, x_{m,k}) < 1/m$ whenever $j \geq N(m)$ and $k \geq N(m)$. Let $y_m = x_{m,N(m)}$. We claim that the sequence $y_1, y_2, y_3, \ldots$ is a Cauchy sequence in X, and that the element $\tilde{y}$ of $\tilde{X}$ corresponding to this Cauchy sequence is the limit in $\tilde{X}$ of the sequence $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \ldots$.

Let $\varepsilon > 0$ be given. Then there exists some positive integer M such that $M > 3/\varepsilon$ and $d(\tilde{x}_p, \tilde{x}_q) < \frac{1}{3}\varepsilon$ whenever $p \geq M$ and $q \geq M$. It follows from the definition of the distance function on $\tilde{X}$ that if $p \geq M$ and $q \geq M$ then $d(x_{p,k}, x_{q,k}) < \frac{1}{3}\varepsilon$ for all sufficiently large natural numbers k. If $p \geq M$ and $k \geq N(p)$ then

$$d(y_p, x_{p,k}) = d(x_{p,N(p)}, x_{p,k}) < 1/p \leq 1/M < \tfrac{1}{3}\varepsilon$$

It follows that if $p \geq M$ and $q \geq M$, and if k is sufficiently large, then $d(y_p, x_{p,k}) < \frac{1}{3}\varepsilon$, $d(y_q, x_{q,k}) < \frac{1}{3}\varepsilon$, and $d(x_{p,k}, x_{q,k}) < \frac{1}{3}\varepsilon$, and hence $d(y_p, y_q) < \varepsilon$. We conclude that the sequence $y_1, y_2, y_3, \ldots$ of points of X is indeed a Cauchy sequence.

Let $\tilde{y}$ be the element of $\tilde{X}$ which is represented by the Cauchy sequence $y_1, y_2, y_3, \ldots$ of points of X, and, for each natural number m, let $\tilde{y}_m$ be the element of $\tilde{X}$ represented by the constant sequence $y_m, y_m, y_m, \ldots$ in X. Now

$$d(\tilde{y}, \tilde{y_m}) = \lim_{p \to +\infty} d(y_p, y_m),$$

and therefore $d(\tilde{y}, \tilde{y_m}) \to 0$ as $m \to +\infty$. Also

$$d(\tilde{y}_m, \tilde{x}_m) = \lim_{j \to +\infty} d(x_{m,N(m)}, x_{m,j}) \leq \frac{1}{m}$$

and hence $d(\tilde{y}_m, \tilde{x}_m) \to 0$ as $m \to +\infty$. It follows from this that $d(\tilde{y}, \tilde{x}_m) \to 0$ as $m \to +\infty$, and therefore the Cauchy sequence $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \ldots$ in $\tilde{X}$ converges to the point $\tilde{y}$ of $\tilde{X}$. We conclude that $\tilde{X}$ is a complete metric space, since we have shown that every Cauchy sequence in $\tilde{X}$ is convergent. ∎

Remark In a paper published in 1872, Cantor gave a construction of the real number system in which real numbers are represented as Cauchy sequences of rational numbers. The real numbers represented by two Cauchy sequences of rational numbers are equal if and only if the difference of the Cauchy sequences converges to zero. Thus the construction of the completion of a metric space, described above, generalizes Cantor's construction of the system of real numbers from the system of rational numbers.