MT3502 - Real Analysis

Kenneth Falconer

Martinmas Semester 2016-17

About the course

This module continues the study of analysis begun in MT2502 Analysis. The approach will be rigorous, with precise definitions of the concepts involved and careful proofs of important theorems which become the building blocks of more advanced work. Near the end of the course the language of metric spaces will be introduced to show how the ideas developed for real functions fit into a much broader framework

The course will have the following components:

1. Brief review of sets, numbers and functions
2. Countable and uncountable sets
3. Review of convergence and continuity; uniform continuity
4. Riemann integration
4. Power series
5. Convergence and continuity in normed and metric spaces

There are many books on analysis which cover most of this course and related topics, in particular the following (note that these texts are for further reading or getting a different perspective on the course).

John M. Howie, *Real Analysis*, Springer, 2001.

Robert G. Bartle & Donald R. Sherbert, *Introduction to Real Analysis*, 4th edition. (Wiley, 2011)

Kenneth Ross, *Elementary Analysis*, 2nd edition. (Springer, 2014).

David Brannan, *A First Course in Mathematical Analysis*. (Cambridge University Press, 2006)

D.J.H. Garling, *A Course in Mathematical Analysis, Vol.1*. (Cambridge University Press, 2014) (More advanced)

Ian Stewart and David Tall, *The Foundations of Mathematics*. 2nd Edition. (Oxford University Press, 2015) (For those interested in set theory, the nature of numbers, infinity, etc.)

• Writing mathematics clearly and carefully becomes increasingly important as you get further into the subject. Advanced mathematical arguments can have a complicated logical structure, and this structure must be clear to any reader (and even more importantly to the writer!).
• In particular, written mathematics should not just be a list of equations but should make clear their logical relationship to each other. The careful use of words to express such relationships is important. In particular, the precise use of words such as 'for all', 'for some', 'implies', 'if', 'only if' is crucial.
• Styles of mathematical writing vary considerably. You should develop your own style using words as well as symbols, to make your arguments as clear as you can.

1 Sets, functions and numbers - Revision

This section is just a reminder of the standard notation and terminology which we will be using.

1.1 Sets

Sets and relations between sets are fundamental to mathematics.

We will take a *set* to be a collection of objects which we term the *members* or *elements* of the set. Generally we use capital letters to denote sets and small letters to denote elements.

We write $a \in A$ to mean that an element a is a member of a set A and $a \notin A$ to mean it is not a member of A.

We sometimes write $\{\cdots\}$ for the set containing the elements $\cdots$.
Here are some standard sets that we will encounter:

$\mathbb{Z} = \{\ldots,-3,-2,-1,0,1,2,3,\ldots\}$ the *Integers*

$\mathbb{Q} = \{p/q$ such that $p,q \in \mathbb{Z}, q \neq 0\}$ the *Rational numbers*

$\mathbb{R} =$ the *Real numbers*

$\mathbb{C} =$ the *Complex numbers*

$\mathbb{N} = \mathbb{Z}^+ = \{1,2,3,\ldots\}$ the *Natural numbers* or *Positive integers*

$\mathbb{Q}^+ =$ the *Positive rationals*, etc.

$\varnothing$ the *Empty set* or *Null set*, i.e. the set with no elements.

We write

$$\{x : \text{ Property involving } x\} \quad \text{or} \quad \{x | \text{ Property involving } x\}$$

to mean "the set of x such that the Property involving x holds".

In this notation, the *closed* and *open intervals* of real numbers between a and b are, respectively,

$$[a,b] = \{x : a \leq x \leq b\} \quad \text{and} \quad (a,b) = \{x : a < x < b\}.$$

We say that A is a *subset* of B ($A \subseteq B$) if every element of A is an element of B. Note that $2 \in \mathbb{Z}$ but $\{2\} \subseteq \mathbb{Z}$, etc.

If $A \subseteq B$ and $A \neq B$ we say A is a *proper* subset of B (some books distinguish this by writing $A \subset B$).

1.1.1 Set operations

Here are some basic operations on sets:

Intersection $A \cap B = \{x : x \in A \text{ and } x \in B\}$;

Union $A \cup B = \{x : x \in A \text{ or } x \in B\}$;

Set difference $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$;

Complement $A^c = \{x : x \in U \text{ and } x \notin A\}$ where U is the universal set under consideration.

More generally, if we have a collection of sets A_i ($i \in I$) indexed by another set I, we can form their intersection and union:

$$\bigcap_{i \in I} A_i = \{x : x \in A_i \text{ for all } i \in I\};$$

$$\bigcup_{i \in I} A_i = \{x : \text{ there exists } i \in I \text{ such that } x \in A_i\}.$$

Power sets

The *power set* of A, written $\mathcal{P}(A)$, is the set of all subsets of A, that is $\mathcal{P}(A) = \{Y : Y \subseteq X\}$. For example, $\mathcal{P}(\{1,2\}) = \big\{\varnothing, \{1\}, \{2\}, \{1,2\}\big\}$.

Cartesian products

The *Cartesian product* or *direct product* $A \times B$ of two sets A and B is the set of all *ordered* pairs

$$A \times B = \{(a,b) : a \in A,\ b \in B\}.$$

Example 1.1.

1. If $A = \{1,2\}$ *and* $B = \{1,2,3\}$ *then*

$$A \times B = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3)\}.$$

2. $\mathbb{R} \times \mathbb{R} = \{(x,y) : x \in \mathbb{R}, y \in \mathbb{R}\} = \mathbb{R}^2 =$ the Euclidean plane.

1.2 Functions

Functions relate elements of different sets. A *function, mapping* or *transformation* f from a set A to a set B is a rule or formula that associates to each element $a \in A$ exactly one element $f(a) \in B$.

We write $f : A \to B$ to mean that f is a function from A to B. We call A the *domain* of f and B the *range* (or *codomain*) of f.

[Formally a function $f : A \to B$ is a subset $f \subseteq A \times B$ with the property that for each $a \in A$ there exists a unique element $b \in B$ such that $(a,b) \in f$. Note that some books, especially in algebra, write af instead of $f(a)$].

Two functions f and g are *equal* if and only if they have the same domain and the same range and $f(x) = g(x)$ for all x in the domain.

Example 1.2. *The following are functions:*

1. $f : \{a,b,c,d\} \to \{1,2,3\}$; $f(a) = 1$, $f(b) = 2$, $f(c) = 1$, $f(d) = 3$.
2. $f : \mathbb{Z} \to \mathbb{Z}$; $f(x) = x+1$ *for each* $x \in \mathbb{Z}$.
3. $f : \mathbb{Z} \to \mathbb{Z}$; $f(x) = 2x$ *for each* $x \in \mathbb{Z}$.
4. $f : \mathbb{Q} \to \mathbb{Q}$; $f(x) = 2x$ *for each* $x \in \mathbb{Q}$.
5. *For any given set* A *the mapping* $i_A : A \to A$ *given by* $f(x) = x$ *for each* $x \in A$ *is called the* identity function on A.

Example 1.3. *The operation of addition on the integers* $\mathbb{Z}$ *is a function* $+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$.

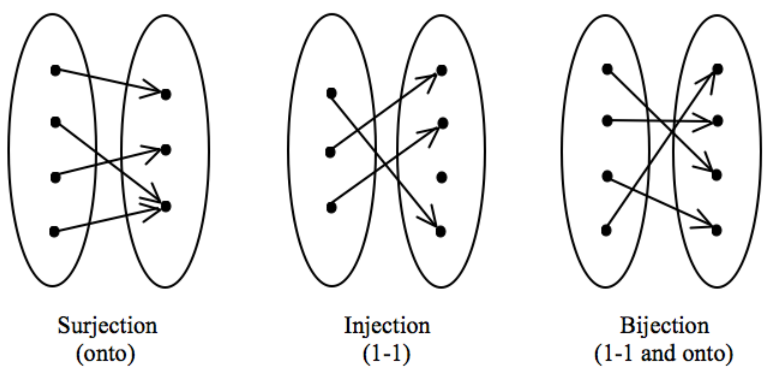
Example 1.4. *A sequence* $(a_1, a_2, a_3, \dots)$ *of elements from a set* A *is just a convenient way of defining the function* $f : \mathbb{N} \to A$ *given by* $f(n) = a_n$.

Types of function

A function $f : A \to B$ is *surjective* or *onto*, if, for every $b \in B$, there exists a (not necessarily unique) $a \in A$ such that $b = f(a)$.

A mapping $f : A \to B$ is *injective* or *one-to-one*, if distinct elements of A always map to distinct images in B, i.e. for $x, y \in A$, $x \neq y$, then $f(x) \neq f(y)$.

A mapping is *bijective* or a *one-to-one correspondence* if it is both injective and surjective.



In Example 1.2 above, 1, 2, 4 and 5 are surjective; 3 is not. For example, to prove function 2 is surjective: given any $y \in \mathbb{Z}$, let $x = y - 1$, then $f(x) = y - 1 + 1 = y$. To show that 3 is not surjective, note that there is no element $x \in \mathbb{Z}$ such that $f(x) = 1$.

Usually, a proof of injectivity uses the contrapositive of the condition just stated – one shows that:

$$f(x) = f(y) \implies x = y.$$

Functions 2, 3, 4 and 5 are injective; 1 is not. For example, function 3 is injective because

$$f(x) = f(y) \implies 2x = 2y \implies x = y.$$

Function 1 is not injective since the distinct elements a and c both map to 1.

Thus functions 2, 4 and 5 are bijections; 1 and 2 are not.

Example 1.5. *The following functions are all different:*

$$\begin{array}{lll} \sin: & \mathbb{R} \to \mathbb{R} & \textit{is neither injective or surjective} \\ \sin: & \mathbb{R} \to [-1,1] & \textit{is surjective} \\ \sin: & [-\frac{\pi}{2},\frac{\pi}{2}] \to \mathbb{R} & \textit{is injective} \\ \sin: & [-\frac{\pi}{2},\frac{\pi}{2}] \to [-1,1] & \textit{is bijective.} \end{array}$$

Composition and inverses of functions Given two functions $f : A \to B$ and $g : B \to C$, we often want to combine these functions to give us a new function from A to C whose effect on an element $a \in A$ is that of first applying the f to a, and then applying g to the result:

For two functions $f : A \to B$ and $g : B \to C$, their *composition* is the function $g \circ f : A \to C$ defined by $(g \circ f)(a) = g(f(a))$ for all $a \in A$.

Note that composition is associative; that is if $f : A \to B$, $g : B \to C$ and $h : C \to D$ then $f \circ (g \circ h)$ and $(f \circ g) \circ h$ are equal, as they both map $A \to D$ and

$$\big(f \circ (g \circ h)\big)(a) = f\big((g \circ h)(a)\big) = f\big(g(h(a))\big) = (f \circ g)\big((h)(a)\big) = \big((f \circ g) \circ h\big)(a)$$

for all $a \in A$.

Let $f : A \to B$. A function $g : B \to A$ is called the *inverse* of f if $g \circ f = i_A$ and $f \circ g = i_B$. If such an inverse exists we say that f is is *invertible*. We write f^{-1} for the inverse of f, so that $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$. (The inverse of a function is unique).

It turns out that we can exactly characterize the invertible mappings:

Theorem 1.6. *Let $f : A \to B$. Then f has an inverse if and only if it is a bijection.*

Proof. Suppose f has inverse f^{-1}. To show that f is injective, let $x, y \in A$ be such that $f(x) = f(y)$. Then $x = f^{-1}(f(x)) = f^{-1}(f(y)) = y$.

To show f is surjective, let $z \in B$ so that $f^{-1}(z) \in A$. Then $f(f^{-1}(z)) = i_B(z) = z$. So f is surjective and thus bijective.

Now suppose f is a bijection. We obtain an inverse by 'reversing the arrows'. The bijective property means that we can define $f^{-1} : B \to A$ by $f^{-1}(y) = \{\text{the unique } x : f(x) = y\}$. Then immediately $f(f^{-1}(y)) = y$ and $f^{-1}(f(x)) = x$. □

Example 1.7. $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ *is a bijection and* $\tan^{-1} : \mathbb{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$ *is its inverse.*

1.3 Numbers

Throughout this course we will assume the familiar and standard arithmetic properties of $\mathbb{Z}, \mathbb{Q}$ and, in particular for this course, $\mathbb{R}$.

Concisely, the reals $\mathbb{R}$ together with the operations of $+$ and $\times$ form a field, i.e. $\mathbb{R}$ is a commutative group under $+$, and $\mathbb{R} \setminus \{0\}$ is a commutative group under $\times$, and the distributive law holds.

$\mathbb{R}$ also satisfies the standard order properties, that is for all $a, b, c, d \in \mathbb{R}$ we have either $a \leq b$ or $b \leq a$, with both holding iff $a = b$, and

$a \leq b, b \leq c$ implies $a \leq c$
$a \leq b, c \leq d$ implies $a + c \leq b + d$
$0 \leq a \leq b, 0 \leq c \leq d$ implies $ac \leq bd$.

A key property of $\mathbb{R}$, that distinguishes it from $\mathbb{Q}$, is the *completeness property*: every Cauchy sequence of real numbers converges to a real number (there are various other equivalent definitions of completeness). We will come back to this in more detail later.

2 Countable and uncountable sets

2.1 Size and similarity of sets

When do two sets have the same size? When is one set larger than another?

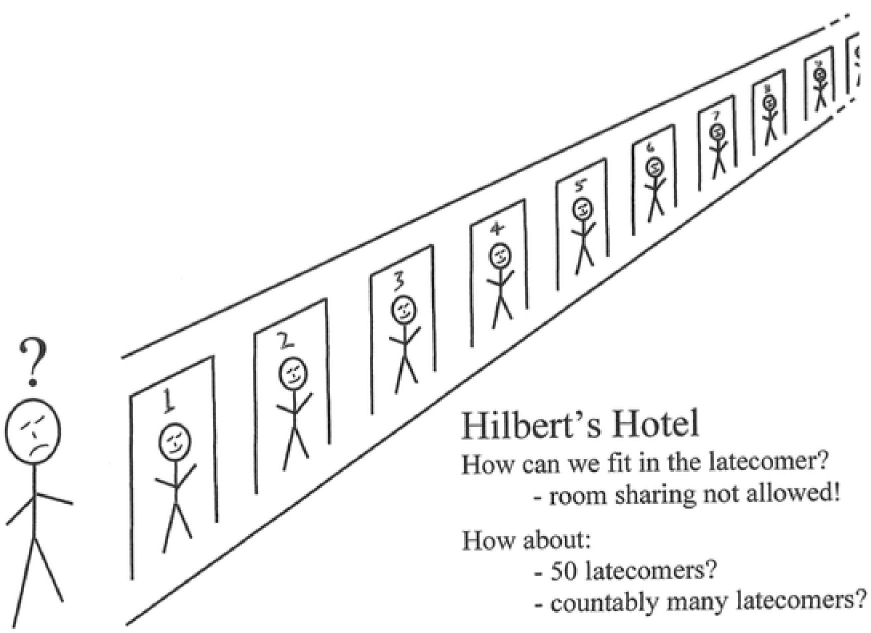
For finite sets, this is easy: two sets X and Y have the same size if they contain the same number of elements. The set X is larger than the set Y if X contains a greater number of elements. But, if you think about it, what is the 'number'? We determine the number of elements of a finite set by establishing a one-one correspondence between the set and another set for which we 'know' its number of elements (e.g. the 'set' of fingers on our hand). For example $\{1,2,3\}$ and $\{3,4,5\}$ have the same size because there is a one-one correspondence between them: $1 \leftrightarrow 3$, $2 \leftrightarrow 4$, $3 \leftrightarrow 5$. The set $\{1,2,3\}$ is smaller than the set $\{5,6,7,8\}$ because no such correspondence can be established, but there is an injection $1 \mapsto 5$, $2 \mapsto 6$, $3 \mapsto 7$.

What about infinite sets? For example, which set is larger: $\mathbb{N} = \{1,2,3,\ldots\}$ or $\mathbb{N} \setminus \{1\} = \{2,3,4,\ldots\}$? One may be tempted to say that $\mathbb{N}$ is larger as it contains all the elements of $\mathbb{N} \setminus \{1\}$ plus an extra element. But there is nevertheless a bijection or correspondence between the two sets of numbers:

$$\begin{array}{ccccccccc} \mathbb{N} & & 1 & 2 & 3 & 4 & 5 & 6 & \ldots \\ & & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ \mathbb{N} \setminus \{1\} & & 2 & 3 & 4 & 5 & 6 & 7 & \ldots \end{array}$$

– removing an element from an infinite set does not change its size from this point of view! But how about $\mathbb{N}$ and $2\mathbb{N} = \{2,4,6,\ldots\}$? This time we have removed 'lots' of elements: 1, 3, 5,.... Nonetheless, they still 'look the same'. More precisely, there is a bijection between them:

$$\begin{array}{cccccccc} \mathbb{N} & 1 & 2 & 3 & 4 & 5 & 6 & \ldots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ 2\mathbb{N} & 2 & 4 & 6 & 8 & 10 & 12 & \ldots \end{array}$$



This leads us to the definition of 'similarity' of sets.

Definition 2.1. *We say that a set A is* similar *to a set B if there is a bijection $f : A \to B$, in which case we write $A \simeq B$.*

Example 2.2. $\{1,2,3\} \simeq \{4,5,6\}$ *since $x \mapsto x+3$ is a bijection between the sets.*
$\{1,2,3,\dots\} \simeq \{4,5,6,\dots\}$ *since $x \mapsto x+3$ is a bijection between the sets.*

Proposition 2.3. *Similarity $\simeq$ is an equivalence relation; that is for all sets A, B and C:*

Reflexive: $A \simeq A$ (since the identity mapping $i_A : A \to A$, $i_A(x) = x$ is a bijection)

Symmetric: $A \simeq B \implies B \simeq A$ (since if $f : A \to B$ is a bijection then so is $f^{-1} : B \to A$).

Transitive: $A \simeq B$ *and* $B \simeq C \implies A \simeq C$ (since if $f : A \to B$ and $g : B \to C$ are bijections then so is their composition $g \circ f : A \to C$)

Because of the symmetry condition, we can simply say that two sets A and B are *similar* if there exists a bijection between A and B.

The 'size' of finite sets does not pose a problem.

Definition 2.4. *We say that a non-empty set A is* finite *if it is similar to $\{1,2,\dots,n\}$ for some $n \in \mathbb{N}$, in which case we say that A has* cardinality *$n \in \mathbb{N}$ and we write $|A| = n$. A non-empty set that is not finite is* infinite*.*

Proposition 2.5. *For A, B finite sets, $A \simeq B$ if and only if A and B have the same cardinality.*

Proof. If A and B both have cardinality n then $A \simeq \{1,2,\dots,n\}$ and $\{1,2,\dots,n\} \simeq B$ so it by transitivity $A \simeq B$.
If $A \simeq B$ and $|A| = n$ then $\{1,2,\dots,n\} \simeq A$ so by transitivity $\{1,2,\dots,n\} \simeq B$ so $|B| = n$. $\square$

The situation for infinite sets gets rather more complicated as some examples indicate.

Example 2.6.

1. $\mathbb{N} \simeq \mathbb{N} \setminus \{1\} \equiv \{2,3,4,\dots\}$ since the mapping $f : \mathbb{N} \to \mathbb{N} \setminus \{1\}$ given by $f(x) = x+1$ is a bijection.

2. $\mathbb{N} \simeq 5\mathbb{N} \equiv \{5,10,15,\dots\}$ since $f : \mathbb{N} \to 5\mathbb{N}$ given by $f(n) = 5n$ is a bijection; thus we have the following 1-1 correspondence:

$$\begin{array}{ccccccccc} \mathbb{N} & 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ 5\mathbb{N} & 5 & 10 & 15 & 20 & 25 & 30 & \dots \end{array}$$

3. $\mathbb{N} \simeq \mathbb{Z}$ since $f : \mathbb{N} \to \mathbb{Z}$ given by

$$f(n) = \begin{cases} n/2 & \text{if } n \text{ is even,} \\ -((n-1)/2) & \text{if } n \text{ is odd.} \end{cases}$$

is a bijection. Thus we have the pairing:

$$\begin{array}{cccccccc} \mathbb{N} & 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ \mathbb{Z} & 0 & 1 & -1 & 2 & -2 & 3 & \dots \end{array}$$

4. $\mathbb{R} \simeq \mathbb{R}^+$ since $\exp : \mathbb{R} \to \mathbb{R}^+$ is a bijection.

5. $(-\pi/2, \pi/2) \simeq \mathbb{R}$ since $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$ is a bijection.

Whilst these examples might seem reasonable, we will shortly encounter rather more surprising pairs of similar and non-similar sets. In particular, we will show that $\mathbb{N} \simeq \mathbb{Q}$ but $\mathbb{N} \not\simeq \mathbb{R}$. Thus the ‘size’ of $\mathbb{R}$ is ‘strictly larger’ than the ‘size’ of $\mathbb{N}$ and $\mathbb{Q}$, in other words the infinite sets $\mathbb{Q}$ and $\mathbb{R}$ are of different sizes.

2.2 Countable sets

Definition 2.7. *An infinite set A is* countable *if $\mathbb{N} \simeq A$, that is if there exists a bijection $f : \mathbb{N} \to A$. We can think of such a bijection as a* list *or* enumeration *of the elements of A:*

$$\begin{array}{cccccccc} \mathbb{N} & 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ A & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \dots \end{array}$$

where each element of A appears exactly once as an a_i.

Thus to show that an infinite set A is countable it is enough to show that we can *list*, *count*, or *enumerate* its elements as a sequence $(a_1, a_2, a_3, a_4, \dots)$ so that each element of A occurs somewhere in the sequence. (Bear in mind that such a list is just a way of specifying a bijection $\mathbb{N} \to A$).

Alternatively, since $\simeq$ is an equivalence relation, if we can show that $A \simeq B$ for a set B that is known to be countable, then A must be countable.

[Note that some books also consider finite sets to be countable.]

Example 2.8.

1. $\mathbb{Z}$ is countable since we may enumerate $\mathbb{Z}$ as $0, 1, -1, 2, -2, 3, -3, 4, \dots$.

2. $\mathbb{Q} \cap (0,1)$, i.e. the set of rational numbers between 0 and 1, is countable since it may be enumerated in the logical sequence:

$$\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \cancel{\frac{2}{4}}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \cancel{\frac{2}{6}}, \dots$$

where we delete any fraction that has already appeared in another form. (We will see later that $\mathbb{Q}$ itself is countable.)

3. The set $\mathbb{N} \times \mathbb{N}$ of all pairs of natural numbers is countable. To see this write $\mathbb{N} \times \mathbb{N}$ in an array as below (as coordinates of points in the plane with both coordinates natural numbers) and enumerate in the manner indicated by superscripts:

$$\begin{array}{cccccccc} \vdots & \nwarrow & \vdots & & \vdots & & \vdots & \\ ^{10}(4,1) & & ^{14}(4,2) & & ^{19}(4,3) & & ^{25}(4,4) & \dots \\ & \nwarrow & & \nwarrow & & \nwarrow & & \\ ^{6}(3,1) & & ^{9}(3,2) & & ^{13}(3,3) & & ^{18}(3,4) & \dots \\ & \nwarrow & & \nwarrow & & \nwarrow & & \\ ^{3}(2,1) & & ^{5}(2,2) & & ^{8}(2,3) & & ^{12}(2,4) & \dots \\ & \nwarrow & & \nwarrow & & \nwarrow & & \nwarrow \\ ^{1}(1,1) & & ^{2}(1,2) & & ^{4}(1,3) & & ^{7}(1,4) & \dots \end{array}$$

Thus we may enumerate $\mathbb{N} \times \mathbb{N}$ as

$$(1,1),(1,2),(2,1),(1,3),(2,2),(3,1),(1,4),(2,3),(3,2),(4,1),(1,5),\ldots.$$

More formally, the mapping $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$

$$f(a,b) = \left(\sum_{k=1}^{a+b-2} k\right) + a = \tfrac{1}{2}(a+b-2)(a+b-1) + a$$

is a bijection that gives the position of (a,b) in the list.

4. By writing pairs in an array in the same way as (3) we may show that if A and B are countable then the product $A \times B = \{(a,b) : a \in A, b \in B\}$ is countable. It follows that $\mathbb{Z}, \mathbb{Z} \times \mathbb{Z}, \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}, \ldots$ are all countable.

Sometimes it is easier to set up an injection or surjection than to find a bijection between $\mathbb{N}$ and a given set.

Proposition 2.9. *(i) Every subset of a countable set is countable or finite.*

(ii) If $f : A \to B$ is a surjection and A is countable then B is countable or finite.

(iii) If $f : A \to B$ is an injection and B is countable then A is countable or finite.

Proof. (i) If A is countable we may list its elements as $(a_1, a_2, a_3, \ldots)$. Any subset may be enumerated by deleting elements not in the subset, i.e. as $(a_{i_1}, a_{i_2}, a_{i_3}, \ldots)$ where $1 \le i_1 < i_2 < i_3 < \cdots$ and this will either terminate or give an enumeration of a countable set.

(ii) List A as $(a_1, a_2, a_3, \ldots)$. Then define $A' \subseteq A$ by $A' = \{a_i \in A$ such that $f(a_i) \neq f(a_j)$ for all $1 \le j < i\}$. Then $A' \simeq B$ and by (i) A' is countable or finite, so B is countable or finite.

(iii) We have $A \simeq f(A) \subseteq B$, so using (i) $f(A)$ and thus A is countable or finite. □

Part (iii) enables us to show certain sets are countable by 'coding' them by integers and using the uniqueness of factorisation of integers.

Example 2.10. *The map $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $f(a,b) = 2^a 3^b$ is an injection, since if $2^a 3^b = 2^{a'} 3^{b'}$ then $(a,b) = (a',b')$ by unique factorisation. By Proposition 2.9 (iii) $\mathbb{N} \times \mathbb{N}$ is countable. (This is an alternative to the direct proof above).*

The following property, often stated as 'A countable union of countable sets is countable', is useful in showing certain sets are countable.

Theorem 2.11 (Cantor's Theorem)**.** *Let $A_1, A_2, \ldots$ be a countable family of countable sets. Then $\bigcup_{i=1}^{\infty} A_i$ is countable.*

Proof. For each $i = 1,2,3\ldots$ let $A_i = \{a_{i1}, a_{i2}, a_{i3}, \ldots\}$. We can enumerate $\bigcup_{i=1}^{\infty} A_i$ as $(a_{11}, a_{12}, a_{21}, a_{13}, a_{22}, a_{31}, a_{14}, \ldots)$ and then delete any duplicates. Alternatively, define an injection $f : \bigcup_{i=1}^{\infty} A_i \to \mathbb{N}$ by $f(a_{ij}) = 2^i 3^j$ so that countability follows from Proposition 2.9 (iii). □

Example 2.12. *(i) The set $\mathbb{Q}$ of all rationals is countable.*

(ii) The set of all algebraic numbers (i.e. solutions of polynomial equations with integer coefficients) is countable.

Proof. (i) For $n = 1,2,3,\ldots$ let $A_n = \{r/n : r \in \mathbb{Z}\}$. Then A_n is countable, so by Theorem 2.11 $\mathbb{Q} = \bigcup_{n=1}^{\infty} A_n$ is countable.

(ii) For $n = 1,2,3,\ldots$ let A_n be the set of all zeros of polynomials with integer coefficients of degree $\le n$. There are countably many such polynomials (since the list of coefficients are in bijective correspondence with the $(n+1)$-fold product $\mathbb{Z} \times \cdots \times \mathbb{Z}$), and each such polynomial has at most n distinct roots, so each A_n is countable. But the set of algebraic numbers is $\bigcup_{n=1}^{\infty} A_n$ which is countable by Theorem 2.11. □

2.3 Some uncountable sets

An infinite set that is not countable is called *uncountable*. We start with Cantor's classical 'diagonal argument' that demonstrates that the real numbers are uncountable.

Proposition 2.13. $\mathbb{R}\cap(0,1)$ *is uncountable and so* $\mathbb{R}$ *is uncountable.*

Proof. Suppose, for a contradiction, that $\mathbb{R}\cap(0,1)$ is countable, so that we may enumerate its elements as a list $(a_1,a_2,a_3,a_4,\ldots)$ which must contain every number in $(0,1)$. We may express these numbers in decimal form

$$\begin{array}{rcl} a_1 & = & 0\,.\,\underline{a_{11}}\,a_{12}\,a_{13}\,a_{14}\ldots \\ a_2 & = & 0\,.\,a_{21}\,\underline{a_{22}}\,a_{23}\,a_{24}\ldots \\ a_3 & = & 0\,.\,a_{31}\,a_{32}\,\underline{a_{33}}\,a_{34}\ldots \\ a_4 & = & 0\,.\,a_{41}\,a_{42}\,a_{43}\,\underline{a_{44}}\ldots \\ \vdots & & \quad\vdots \end{array}$$

so that a_{ij} is the jth decimal digit of a_i. (If a_i is a number with two decimal expansions, take the one ending in a string of 0s rather than that ending in a string of 9s.)

Define

$$b = 0.b_1\,b_2\,b_3\,b_4\ldots \text{ where } b_i = \begin{cases} 5 & \text{if} \quad a_{ii}\neq 5 \\ 7 & \text{if} \quad a_{ii}=5 \end{cases}.$$

Then $b\neq a_i$ for all i, since $b_i\neq a_{ii}$, that is b differs from a_i in the ith decimal place. Thus b is not in the list, which contradicts the assumption that the list contains all real numbers in $(0,1)$. $\square$

Recall that $\mathcal{P}(A)$ denotes the power set of A, that is the set of all subsets of A. The following result, which is really just a variant of the previous one, intuitively says that the power set of a set A has cardinality strictly larger than that of A itself.

Theorem 2.14. *Let A be a non-empty set. Then $\mathcal{P}(A)\not\simeq A$.*

Proof. Suppose, for a contradiction, that there exists a bijection $f : A\to\mathcal{P}(A)$. Let $B=\{x\in A \text{ such that } x\notin f(x)\}$. Since f is a bijection, $B=f(a)$ for some $a\in A$.

From the definition of B, $a\in B$ iff $a\notin f(a)=B$, a contradiction. Thus there that there can be no bijection from A to B. $\square$

It follows, at least intuitively, that, given any infinite set there is a strictly larger one, so there are infinitely many 'different sizes' of infinity.

Example 2.15. *(i) The set of all subsets of* $\mathbb{N}$ *is uncountable.*

(ii) However, the set of all finite *subsets of* $\mathbb{N}$ *is countable.*

Proof. (i) This follows from Theorem 2.14.

(ii) For each n let A_n be the set of all subsets of $\{1,2,\ldots,n\}$. Then A_n is finite; indeed $|A_n|=2^n$. The set of all finite subsets of $\mathbb{N}$ is just $\bigcup_{n=1}^{\infty}A_n$ so is countable by Theorem 2.11. $\square$

2.4 Cardinality of infinite sets - a very brief discussion

This section is non-examinable.

We return to thinking of cardinality as representing the size of sets – can the idea of $|A|$ denoting the number of elements in a finite set A be extended in a meaningful manner to infinite sets? We recall and extend the Definition 2.1 to allow comparison as well as equality of cardinalities – we start to think of $|A|$ as the size of A even if A is infinite.

Definition 2.16. *As before, two sets A and B are* similar *if there is a bijection $f : A \to B$, in which case we now write $|A| = |B|$ and say that they have the* same cardinality.

We say that the set A has cardinality less than or equal to *B if there is an injection $f : A \to B$, and we write this as $|A| \leq |B|$. We say that the set A has* cardinality strictly less than *B if $|A| \leq |B|$ and $|A| \neq |B|$ in which case we write $|A| < |B|$*

Example 2.17. $|\mathbb{Z}| < |\mathbb{R}|$, *since $f : \mathbb{Z} \to \mathbb{R}$ given by $f(n) = n$ is an injection and $\mathbb{Z} \not\sim \mathbb{R}$.*

Theorem 2.18. *Let A, B, C be sets. Then*

(i) $|A| = |A|$,

(ii) *If* $|A| = |B|$ *then* $|B| = |A|$,

(iii) *If* $|A| = |B|$ *and* $|B| = |C|$ *then* $|A| = |C|$,

(iv) $|A| \leq |A|$,

(v) *If* $|A| \leq |B|$ *and* $|B| \leq |C|$ *then* $|A| \leq |C|$.

Proof. Parts (i)-(iii) are just a restatement of Proposition 2.1 concerning properties of bijections.

Part (iv) follows since the identity map is an injection, and (v) follows since if $f : A \to B$ and $g : B \to C$ are injections then $g \circ f : A \to C$ is an injection. □

Despite the notation '$=$' and '$\leq$' with its intuitive connotations, one thing is missing from the above list of properties, namely that if $|A| \leq |B|$ and $|B| \leq |A|$ then $|A| = |B|$. Without this 'antisymmetry' property one might have both $|A| < |B|$ and $|B| < |A|$ and cardinality would be a rather limited notion. Of course, this can be proved, but it is a serious theorem known as the Schroeder–Bernstein Theorem.

Theorem 2.19 (Schroeder–Bernstein)**.** *Given two sets A and B, if there exist injections $f : A \to B$ and $g : B \to A$ then there exists a bijection $h : A \to B$. More succinctly:*

$$|A| \leq |B| \text{ and } |B| \leq |A| \implies |A| = |B|.$$

Proof. **Included for information only.**

If $a \in A$ is such that $f(a) = b$, call a the *parent* of b. Similarly, $b \in B$ is the parent of $c \in A$ if $g(b) = c$. (Notice that the injectivity of f and g means that an element can have at most one parent.)

Let $z \in A \cup B$. An *ancestral chain* for z is a sequence $z_0, z_1, \ldots$ such that $z_0 = z$ and z_{i+1} is the parent of z_i for each i. (An ancestral chain may be of finite or infinite length.) If there is no infinite ancestral chain for z, then the *depth* of z is the index of the last element in the unique longest ancestral chain for z; otherwise z has *infinite depth*. (Observe that z may have depth 0.)

Let A_e, B_e be the subsets of A, B consisting of even-depth elements; A_o, B_o be their subsets consisting of odd-depth elements; and A_∞, B_∞ be their subsets consisting of infinite-depth elements.

Notice that f maps A_e to B_o, A_o to B_e, and A_∞ to B_∞. Similarly, g maps B_e to A_o, B_o to A_e, and B_∞ to A_∞.

Observe that elements of $A_o \cup B_o \cup A_\infty \cup B_\infty$ always have parents; this may not be true for elements of $A_e \cup B_e$, since an element of this set may have depth 0.

Define $h : A \to B$ by

$$h(a) = \begin{cases} f(a) & \text{if } a \in A_e \cup A_\infty, \\ g^{-1}(a) & \text{if } a \in A_o. \end{cases}$$

This mapping is defined everywhere since $g^{-1}(a)$ exists for all $a \in A_o$ and is unique by the injectivity of g.

Suppose $a_1, a_2 \in A$ are such that $h(a_1) = h(a_2)$. If $h(a_1) = h(a_2)$ lies in B_e, then $a_1, a_2 \in A_o$. So $g^{-1}(a_1) = h(a_1) = h(a_2) = g^{-1}(a_2)$. So $a_1 = g(g^{-1}(a_1)) = g(g^{-1}(a_2)) = a_2$. If $h(a_1) = h(a_2)$ lies in $B_o \cup B_\infty$, then $a_1, a_2 \in A_e \cup A_\infty$. So $f(a_1) = h(a_1) = h(a_2) = f(a_2)$. But f is injective, so $a_1 = a_2$. Thus h is injective.

Choose $b \in B$. If $b \in B_e$, let $a = g(b)$. Then $h(a) = g^{-1}(g(b)) = b$. If $b \in B_o \cup B_\infty$, then b has a parent $a \in A$ with $f(a) = b$. In fact, a must lie in $A_e \cup A_\infty$, so $h(a) = f(a) = b$. Thus h is surjective.

Therefore h is a bijection from A to B and thus $|A| = |B|$. □

This fact that $\leq$ satisfies all the properties of a total order allows us to extend some of our earlier notions to infinite cardinalities.

Corollary 2.20. *(i) Let A be a non-empty set. Then $|A| < |\mathcal{P}(A)|$.*

(ii) There are infinitely many different infinite cardinalities.

(iii) There is no largest cardinality.

(iv) There is no set containing all sets.

Proof. (i) We showed in Theorem 2.14 that $|A| \neq |\mathcal{P}(A)|$, but clearly $|A| \leq |\mathcal{P}(A)|$ since $a \mapsto \{a\}$ is an injection. Then $|A| < |\mathcal{P}(A)|$.

(ii) We may define a sequence by $A_1 = \mathbb{N}$ and $A_n = \mathcal{P}(A_{n-1})$ for $n \geq 2$. By (i) $|A_1| < |A_2| < |A_3| < \cdots$.

(iii) This follows from (i).

(iv) If there was such a set it would have to have cardinality strictly greater than its own, by (i). □

So far we have not given a meaning to $|A|$ outside the context of '=' and '$\leq$'. Motivated by finite sets, we can think of $|A|$ as an 'infinite number' or *cardinal* representing the cardinality of A, and so of any set B such that $A \simeq B$.

Example 2.21. *We write $\aleph_0$ ('aleph nought') for the cardinality of $\mathbb{N}$ so a set is countable if $|A| = \aleph_0$. Thus $|\mathbb{N}| = |\mathbb{Q}| = \aleph_0$.*

We write $\mathfrak{c}$ for the cardinality of $\mathbb{R}$ or 'cardinality of the continuum' . Thus $|\mathbb{R}| = |\mathbb{C}| = \mathfrak{c}$ and $\aleph_0 < \mathfrak{c}$.

It is possible to define an arithmetic on cardinals:

Definition 2.22. *Let A and B be disjoint infinite sets. The sum and product of cardinals $|A|$ and $|B|$ is defined by*

$$|A| + |B| = |A \cup B|, \; |A| \cdot |B| = |A \times B|.$$

It may be checked that these operations are well-defined, i.e. independent of which sets of given cardinality are chosen. The basic arithmetic for infinite cardinals is very simple:

Theorem 2.23. *For any two infinite sets A and B*

$$|A| + |B| = |A| \cdot |B| = \max(|A|, |B|).$$

Typically, if you are given specific A and B this is reasonably easy to prove. However, proving the statement in general is quite difficult, and is beyond our scope here.

Example 2.24. $|\mathbb{C}| = |\mathbb{R} \times \mathbb{R}| = \mathfrak{c} \times \mathfrak{c} = \mathfrak{c}$.

The Continuum Hypothesis

Does there exist a cardinal $\mathfrak{a}$ such that $\aleph_0 < \mathfrak{a} < \mathfrak{c}$, or to put it another way, is there a set X such that $|\mathbb{N}| < |X| < |\mathbb{R}|$? That is, is there an uncountable set that is 'smaller' than the reals?

The conjecture, originally made by Cantor, that no such set exists is called the *Continuum Hypothesis*.

Gödel showed that one cannot *disprove* the Continuum Hypothesis using standard mathematical logic, and later Cohen showed that one cannot *prove* the Continuum Hypothesis using standard mathematical logic.

Therefore the Continuum Hypothesis is independent of the usual axioms of mathematics and one can choose whether to assume that the Continuum Hypothesis is true or that its negation is true.

3 Review of convergence and continuity; uniform continuity

'Mathematical analysis' usually refers to the rigorous treatment of those areas of mathematics that involve some sort of limit, including continuity, differentiation, integration, etc., and the remainder of this course concerns central topics in analysis. We need to be careful to work from precise definitions and develop properties and theorems from those definitions, otherwise it is easy to draw false conclusions. Sketch diagrams, such as graphs, are often very helpful in getting an idea of what is happening, but this intuition then has to be converted into precise argument.

This chapter essentially reviews material that you will have met in MT2502 that we will need in this course. For detailed proofs and many examples, see the MT2502 notes or any book on basic analysis. There is one 'new' topic in this chapter, namely uniform continuity.

3.1 Bounds of sets of real numbers

Many of the difficulties that are encountered with real numbers stem from the the fact that a set of real numbers need not contain a maximum or minimum element. To get around this, we introduce infima and suprema.

Definition 3.1. (Bounds, suprema and infima) *Let X be a non-empty set of real numbers. Then s is an* upper bound *for X if $x \leq s$ for all $x \in X$. Similarly s is an* lower bound *for X if $x \geq s$ for all $x \in X$. A set that has an upper (resp. lower) bound is termed* bounded above (resp. below) *A set of numbers that has both a lower and upper bound is called* bounded. (*Thus A is bounded if there exists a number M such that $|x| \leq M$ for all $x \in X$.*)

For X a non-empty set of real numbers, a number s is called the supremum *or* least upper bound *of X, written* $\sup X$*, if*

(a) $x \leq s$ for all $x \in X$ (i.e. s is an upper bound) and

(b) for all $t < s$, there exists $x \in X$ with $x > t$ (i.e. no number less than s is an upper bound).

Similarly a number s is called the infimum *or* greatest lower bound *of X, written* $\inf X$*, if*

(a) $x \geq s$ for all $x \in X$ (i.e. s is a lower bound) and

(b) for all $t > s$, there exists $x \in X$ with $x < t$ (i.e. no number greater than s is a lower bound).

We think of $\sup X$ and $\inf X$ as the 'generalised maximum and minimum' of the set X, but it is important to realise that $\sup X$ and $\inf X$ need not be members of X.

Examples

1. The closed interval $[2,7]$ has a maximum element of 7, which is the supremum, and a minimum element of 2, which is the infimum.

2. However, the open interval $(2,7)$ does not contain a maximum element. Nevertheless, 7 is the supremum of the set, since it is an upper bound and no number less than 7 is an upper bound. Similarly, 2 is the infimum of the set. Thus $\sup(2,7) = 7$ and $\inf(2,7) = 2$.

3. The set $X = \{0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots\} = \{(n-1)/n : n \in \mathbb{N}\}$ has $\sup X = 1, \inf X = \min X = 0$, and X contains its infimum, but not its supremum.

Axiom of completeness

The *axiom of completeness* states: every non-empty set of real numbers X that is bounded above has a supremum (i.e. a least upper bound).

The axiom of completeness is what distinguishes the real numbers from the rationals, and it cannot be proved from other (non-equivalent) statements. It is equivalent to asserting the existence of real numbers since we can essentially regard any real number as the supremum of a set of rationals.

For example, let

$$X = \{x \in \mathbb{Q}^+ : x^2 < 2\} \supseteq \{1, 1.4, 1.41, 1.414., 1.4142, \ldots\}.$$

Clearly X is bounded above (by 10, say) so by the axiom has a supremum. It is easy to check that $(\sup X)^2 = 2$, so the axiom guarantees the existence of $\sqrt{2}$ as a real number, but $\sup X \notin \mathbb{Q}$ as there is no rational which when squared gives 2. .

3.2 Convergence of sequences

Formally, a sequence (of real numbers) is a function $f : \mathbb{N} \to \mathbb{R}$, but we invariably write $x_n = f(n)$ and think of a sequence as an infinite list of numbers $(x_1, x_2, x_3, \ldots)$ which we may abbreviate as $(x_n)_{n=1}^{\infty}$ or $(x_n)_n$ or just (x_n) if the context is clear.

Usually the most important thing about a sequence is how the terms x_n behave when n is large, that is the limiting behaviour as $n \to \infty$ and in particular whether or not the sequence converges.

Definition 3.2. (Convergence) *The sequence* x_n converges *to* x *or has* limit x*, written* $x_n \to x$ *or* $\lim x_n = x$*, if for all* $\varepsilon > 0$ *there exists* $N \in \mathbb{N}$ *such that* $|x_n - x| < \varepsilon$ *for all* $n \geq N$*. Such a sequence is called* convergent*; a sequence that is not convergent is* divergent.

To show that a sequence is convergent, we must show that for every number $\varepsilon > 0$ we can find an N as stated. Note that we do not have to find the least such N.

Examples 1. The sequence $\left(\frac{2n+1}{3n+1}\right)$ converges to $\frac{2}{3}$.

Proof. Let $\varepsilon > 0$ be given. If $n > 1/(9\varepsilon)$ then

$$\left|\frac{2n+1}{3n+1} - \frac{2}{3}\right| = \frac{1}{9n+3} < \frac{1}{9n} < \varepsilon,$$

so $\frac{2n+1}{3n+1} \to \frac{2}{3}$. (We have taken '$N$' $= \lceil 1/9\varepsilon \rceil$.)

2. The sequence $\big((-1)^n\big)$ is divergent.

Proof. Suppose $(-1)^n \to x$. Take $\varepsilon = \frac{1}{2}$. If $x \geq 0$ then for all odd n, $|(-1)^n - x| = |-1-x| > \frac{1}{2}$, and if $x \leq 0$ then for all even n, $|(-1)^n - x| = |1-x| > \frac{1}{2}$. In either case there is no N such that $|(-1)^n - x| < \frac{1}{2}$ for all $n \geq N$, so the sequence is divergent.

3. (Standard sequences: fixed powers of integers) For all $\alpha > 0$, $1/n^{\alpha} \to 0$.

4. (Standard sequences: powers of a fixed number) For $c \in \mathbb{R}$

$$\begin{array}{lll} c^n \to 0 & \text{if} & |c| < 1 \\ c^n \to 1 & \text{if} & c = 1 \\ c^n \text{ is divergent} & \text{if} & |c| > 1 \text{ or } c = -1 \end{array}$$

The following standard results on convergence of sequences are recorded here for reference. Full details of proofs, etc, may be found in the MT2503 notes or in books on analysis.

Lemma 3.3. (Boundedness) *Every convergent sequence is bounded.*

Proposition 3.4. (Arithmetic properties) *If $(x_n), (y_n)$ are sequences with $x_n \to x$, $y_n \to y$ and $\lambda \in \mathbb{R}$, then:*

(a) $x_n + y_n \to x + y$
(b) $\lambda x_n \to \lambda x$
(c) $x_n - y_n \to x - y$
(d) $x_n y_n \to xy$
(e) $x_n / y_n \to x/y$ *provided* $y_n, y \neq 0$.

Sample proof. (d) By Lemma 3.3 (y_n) is bounded, say $|y_n| \leq M$ for all n. Using the triangle inequality,

$$\begin{aligned} |x_n y_n - xy| = |(x_n - x)y_n + x(y_n - y)| &\leq |x_n - x||y_n| + |x||y_n - y| \\ &\leq |x_n - x|M + (|x|+1)|y_n - y|. \quad (*) \end{aligned}$$

Given $\varepsilon > 0$, chose N_1 such that $|x_n - x| < \frac{\varepsilon}{2M}$ if $n \geq N_1$ and chose N_2 such that $|y_n - y| < \frac{\varepsilon}{2(|x|+1)}$ if $n \geq N_2$. By $(*)$, if $n \geq \max\{N_1, N_2\}$ then $|x_n y_n - xy| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$. Thus $x_n y_n \to xy$.

Lemma 3.5. (Subsequences) *Let $x_n \to x$. If y_n is a subsequence of x_n (i.e. a sequence obtained by deleting terms) then $y_n \to x$.*

The following key properties of sequences are in fact equivalent to the axiom of completeness. They also guarantee that certain sequences converge without the need to find the limit.

Proposition 3.6. (Increasing sequences) *Every increasing sequence bounded above is convergent. Similarly, every decreasing sequence bounded below is convergent.*

Proof. Let (x_n) be an increasing sequence bounded above by some M, so that $x_n \leq x_{n+1}$ and $x_n \leq M$ for all n. By the axiom of completeness the set of terms $\{x_n : n \in \mathbb{N}\}$ has a supremum x.

Let $\varepsilon > 0$. By definition of supremum, there is some N such that $x_N > x - \varepsilon$. So for all $n \geq N$

$$x - \varepsilon < x_N \leq x_n \leq x < x + \varepsilon.$$

Thus $x_n \to x$, i.e. the sequence is convergent. □

Theorem 3.7. (Convergent subsequences) *Every bounded sequence has a convergent subsequence.*

Definition 3.8. (Cauchy sequences) (x_n) *is a* Cauchy sequence *if given $\varepsilon > 0$ there exists N such that $|x_n - x_m| < \varepsilon$ for all $n, m \geq N$.*

Theorem 3.9. (General principle of convergence.) *A sequence of real numbers is convergent if and only it is a Cauchy sequence.*

Again, this can guarantee convergence without the need to find the limit. For example, the sequence given by

$$x_1 = 1, \;\; x_{n+1} = x_n + \frac{\sin n}{2^n}$$

is easily seen to be Cauchy and so is convergent.

3.3 Continuous functions

Throughout this section we will consider functions $f : A \to \mathbb{R}$ where $A \subseteq \mathbb{R}$. Usually A will be either an interval $[a,b]$ or $\mathbb{R}$ and these are the cases you should keep in mind.

Definition 3.10. (Continuous function) *A function $f : A \to \mathbb{R}$ is* continuous at $c \in A$ *if for all $\varepsilon > 0$ there exists $\delta > 0$ such that if $x \in A$ and $|x-c| < \delta$ then $|f(x) - f(c)| < \varepsilon$.*
A function f is continuous on *the set A if f is continuous at all $x \in A$.*

To remember and understand this definition, think of the picture!
Example $f : [0,1] \to \mathbb{R}$ given by $f(x) = x^2$ is continuous.

Proof. Let $c \in [0,1]$. Let $\varepsilon > 0$ be given. Suppose $x \in [0,1]$ satisfies $|x-c| < \frac{1}{2}\varepsilon$ (thus $\frac{1}{2}\varepsilon$ is our 'δ'). Then

$$|f(x) - f(c)| = |x^2 - c^2| = |x+c||x-c| \leq 2|x-c| < 2 \times \tfrac{1}{2}\varepsilon = \varepsilon.$$

Thus f is continuous at c for every $c \in [0,1]$, so is continuous on $[0,1]$. □

The following characterisation of continuity in terms of sequences can be very useful.

Proposition 3.11. (Sequence characterisation of continuity) *The function $f : A \to \mathbb{R}$ is continuous at $c \in A$ if and only if $f(x_n) \to f(c)$ for all sequences (x_n) with $x_n \in A$ and $x_n \to c$.*

Example $f : [0,1] \to \mathbb{R}$ given by $f(x) = x^2$ is continuous.

Proof. Let $c \in [0,1]$. Suppose $x_n \to c$. Then by Proposition 3.4(d) $f(x_n) = x_n^2 \to c^2 = f(c)$. Thus f is continuous at c by the sequence definition, so f is continuous on $[0,1]$. □

The standard properties of continuity may be proved from the definitions, in particular combining continuous functions in natural ways yields continuous functions.

Proposition 3.12. (Arithmetic properties) *Let $f, g : A \to \mathbb{R}$ be continuous at $c \in A$ (resp. on A), and $\lambda \in \mathbb{R}$. Then the following are continuous at c (resp. on A):*

(a) $f+g$	*(b)* λf
(c) $f-g$	*(d)* fg
(e) f/g (provided $g(x) \neq 0$ *for* $x \in A$)	*(f)* $\lvert f \rvert$
(g) $\min\{f,g\}$	*(h)* $\max\{f,g\}$.

Sample proof. (c) (using an ε-δ argument) Let f, g be continuous at c. So given $\varepsilon > 0$ there exist $\delta_1 > 0$ such that $|f(x) - f(c)| < \frac{1}{2}\varepsilon$ if $|x-c| < \delta_1$, and $\delta_2 > 0$ such that $|g(x) - g(c)| < \frac{1}{2}\varepsilon$ if $|x-c| < \delta_2$. Then if $|x-c| < \min\{\delta_1, \delta_2\}$,

$$\begin{aligned} |(f(x) - g(x)) - (f(c) - g(c))| &= |(f(x) - f(c)) - (g(x) - g(c))| \\ &\leq |f(x) - f(c)| + |g(x) - g(c)| < \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon. \end{aligned}$$

Thus $f - g$ is continuous at c.

Sample proof. (e) (using a sequence argument) Let $x_n \to c$. Since f, g are continuous at c, $f(x_n) \to f(c)$ and $g(x_n) \to g(c) \neq 0$. Applying Proposition 3.4(e), $f(x_n)/g(x_n) \to f(c)/g(c)$, so f/g is continuous at c.

Sample proof. (g) Note that $\min\{f,g\} = \frac{1}{2}\big((f+g) - |f-g|\big)$ and use parts (a), (c), (f). □

Corollary 3.13. (Polynomials) *Every polynomial p is continuous on $\mathbb{R}$. (A polynomial is a function of the form $p(x) = a_n x^n + \cdots + a_1 x + a_0$ where $a_i \in \mathbb{R}$.)*

Proof. The constant functions $f(x) = a$ and the identity $g(x) = x$ are clearly continuous on $\mathbb{R}$. Every polynomial may be formed from these two functions using combinations of $+, -, \times$, so, using Proposition 3.12 repeatedly, is continuous on $\mathbb{R}$. $\square$

Proposition 3.14. (Composition) *Let $g : \mathbb{R} \to \mathbb{R}$ be continuous at x and $f : \mathbb{R} \to \mathbb{R}$ be continuous at $g(x)$. Then the composition $f \circ g : \mathbb{R} \to \mathbb{R}$ is continuous at x (where $(f \circ g)(x) = f(g(x))$).*

Proof. Exercise - use sequences to get a very short proof! $\square$

There are two very important theorems concerning continuous functions on *closed* intervals: the Extreme Value Theorem and the Intermediate Value Theorem. We state them here; there are various proofs which can be found in books or in the MT2503 notes, but all of them depend crucially on the completeness of the real numbers (or some equivalent property) along with the definition of continuity.

Theorem 3.15. (Extreme value theorem) *Let $f : [a,b] \to \mathbb{R}$ be continuous. Then f is bounded and attains its bounds, i.e. there exists M such that $|f(x)| \le M$ for all $x \in [a,b]$, and there exist $c_1, c_2 \in [a,b]$ such that $f(c_1) = \inf\{f(x) : x \in [a,b]\}$ and $f(c_2) = \sup\{f(x) : x \in [a,b]\}$.*

Theorem 3.16. (Intermediate value theorem) *Let $f, g : [a,b] \to \mathbb{R}$ be continuous. Let $k \in \mathbb{R}$ satisfy $f(a) < k < f(b)$ (or $f(b) < k < f(a)$). Then there exists $c \in (a,b)$ with $f(c) = k$.*

3.4 Uniform continuity

Uniform continuity is a very useful (and natural) notion that allows us to simplify certain arguments. There is one main result that we will need, namely that 'a continuous function on a closed interval is uniformly continuous'; in many ways this is in the spirit of the extreme value and intermediate value theorems mentioned above.

Consider the following proofs that certain functions are continuous on $\mathbb{R}$.

Examples 1. Let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = 3\sin x$. Let $c \in \mathbb{R}$. Given $\varepsilon > 0$, if $|x - c| < \frac{1}{3}\varepsilon$ then

$$|f(x) - f(c)| = |3\sin x - 3\sin c| = 3|\sin x - \sin c| \le 3|x - c| < 3 \times \tfrac{1}{3}\varepsilon = \varepsilon.$$

Thus f is continuous at c for all $c \in \mathbb{R}$, so is continuous on $\mathbb{R}$.

Note here that the same 'δ'= $\frac{1}{3}\varepsilon$ *works for all $c \in \mathbb{R}$.*

2. Let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = x^2$. Let $c \in \mathbb{R}$. Given $\varepsilon > 0$, if $|x - c| < \min\{1, \varepsilon/(2|c|+1)\}$ then

$$|f(x) - f(c)| = |x^2 - c^2| = |x + c||x - c| \le (|x| + |c|)|x - c| \le \big((|c|+1) + |c|\big)|x - c| < \varepsilon. \quad (*)$$

Thus f is continuous at c, so is continuous on $\mathbb{R}$.

Note here that the larger c is, the smaller this 'δ' = $\min\{1, \varepsilon/(2|c|+1)\}$ *is. For a given ε we cannot find a δ that works for all c. From $(*)$, we need $|x - c| < \delta$ to imply that $|f(x) - f(c)| = |x + c||x - c| < \varepsilon$, so $|x - c|$ has to be very small if c is large and x is close to c making $|x + c|$ large.*

Thus the situation in Example 1 is nicer than that in Example 2 in the sense that, in the definition of continuity, for any given ε we can find a δ that works for *all* $c \in \mathbb{R}$ simultaneously. In mathematics we use the word *uniform* to refer to such situations when a number does not depend on some underlying parameter (c in this case).

This leads to the following definition.

Definition 3.17. (Uniform continuity) *Let $A \subseteq \mathbb{R}$. A function $f : A \to \mathbb{R}$ is* uniformly continuous *on A if for all $\varepsilon > 0$ there exists $\delta > 0$ such that $\underline{\text{for all } c \in A}$ and all $x \in A$ with $|x - c| < \delta$ we have $|f(x) - f(c)| < \varepsilon$.*

This should be compared with Definition 3.10 relating to continuity on a set which may be stated in the following way. Note that changing the order of the conditions significantly changes the meaning.

Alternative Definition 3.10. (Continuity on a set) *Let $A \subseteq \mathbb{R}$. A function $f : A \to \mathbb{R}$ is* continuous on A *if $\underline{\text{for all } c \in A}$ and for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in A$ with $|x - c| < \delta$ we have $|f(x) - f(c)| < \varepsilon$.*

Thus the function in Example 1 is uniformly continuous on $\mathbb{R}$ but the function in Example 2 is continuous on $\mathbb{R}$ but not uniformly continuous.

Of course, every function that is uniformly continuous function on a set A is necessarily continuous on A. However, the converse is not in general true, as Example 2 shows. However, the converse *is* true in the very important case when $A = [a, b]$ is a bounded closed interval.

Theorem 3.18. (Uniform continuity) *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then f is uniformly continuous on $[a, b]$.*

Proof. This may be proved in various ways that all depend (directly or indirectly) on the completeness axiom. Here we give a proof using the fact that every bounded sequence has a convergent subsequence.

Assume, seeking a contradiction, that $f : [a, b] \to \mathbb{R}$ is continuous but *not* uniformly continuous. Then there is some $\varepsilon > 0$ such that for every $\delta > 0$ we can find $c, x \in [a, b]$ such that $|x - c| < \delta$ and $|f(x) - f(c)| \geq \varepsilon$.

Thus, for each $n \in \mathbb{N}$, taking $\delta = \frac{1}{n}$, choose $c_n, x_n \in [a, b]$ such that

$$|x_n - c_n| < \tfrac{1}{n} \quad \text{and} \quad |f(x_n) - f(c_n)| \geq \varepsilon.$$

By Proposition 3.7 (c_n) has a convergent subsequence (c'_n), with $c'_n \to c$, say; let (x'_n) be the corresponding subsequence of (x_n). Then

$$x'_n = (x'_n - c'_n) + c'_n \to 0 + c.$$

Since f is continuous at c, using the sequence definition of continuity,

$$0 < \varepsilon \leq |f(x'_n) - f(c'_n)| \to |f(c) - f(c)| = 0,$$

which is the contradiction sought. □

Examples 1. Every polynomial is uniformly continuous on every bounded closed interval, since polynomials are continuous.

2. Let $f : (0, 1) \to \mathbb{R}$ be $f(x) = 1/x$. Then f is not uniformly continuous on $(0, 1)$ (of course it is continuous on $(0, 1)$). To see this, take $\varepsilon = 1$ and some $\delta > 0$. Then $|f(c + \delta) - f(c)| = \left|\frac{1}{c+\delta} - \frac{1}{c}\right| = \frac{\delta}{c(c+\delta)}$ which will be greater than 1 if we take c close enough to 0 to ensure that $\frac{\delta}{c(c+\delta)} > 1$.

This example shows that Theorem 3.18 is not in general true for open intervals (a, b).

4 Riemann integration

Integration is often first encountered as a means of finding the area under a curve. The difficulty, of course, is that, whilst we can easily find the area of a rectangle by multiplying its side-lengths, for an irregular region finding or even defining the area may be less obvious. A similar problem arises with volumes, averages and other quantities associated with varying functions. We think of an integral as some sort of limit of approximating sums, but this needs to be made precise.

When one first meets integration one learns that to integrate a function one has to find an 'anti-derivative'. It is really rather remarkable that two apparently different concepts are related in this way. A precise definition of derivative was given in MT2502 and in this chapter we will define the integral carefully using the definition due to Bernhard Riemann in 1854. We will then see how various properties of integration follow from this definition, including the 'fundamental theorem of calculus' which relates integration and differentiation.

4.1 Dissections and lower and upper sums

The underlying idea in defining an integral is to approximate a function by 'step functions' for which the areas under the graphs are easily calculated. We start with the notion of a dissection of an interval, which we use to define lower and upper sums that approximate the integral.

Definition 4.1. (Dissection) *A* dissection *or* partition $\mathcal{D}$ *of a closed interval* $[a,b]$ *is a set of numbers*

$$a = a_0 < a_1 < a_2 < \cdots < a_{n-1} < a_n = b.$$

The numbers $\{a_i : i = 0,1,\ldots,n\}$ *are called the* points of dissection *of* $[a,b]$. *The maximum length of the intervals formed by the dissection* $\max_{1\leq i\leq n}\{a_i - a_{i-1}\}$ *is sometimes called the* mesh *or* norm *of the dissection.*

Definition 4.2. (Lower and upper sums) *Let* $f : [a,b] \to \mathbb{R}$ *be a bounded function. We define the* lower *and* upper sums *of* f *with respect to the dissection* $\mathcal{D} = \{a = a_0 < a_1 < \cdots < a_n = b\}$ *by*

$$\underline{S}(\mathcal{D}) = \sum_{i=1}^{n}(a_i - a_{i-1})\inf\{f(x) : a_{i-1} \leq x \leq a_i\}, \tag{4.1}$$

$$\overline{S}(\mathcal{D}) = \sum_{i=1}^{n}(a_i - a_{i-1})\sup\{f(x) : a_{i-1} \leq x \leq a_i\}. \tag{4.2}$$

We sometimes write $\underline{S}_f(\mathcal{D})$ or $\overline{S}_f(\mathcal{D})$ instead of $\underline{S}(\mathcal{D})$ or $\overline{S}(\mathcal{D})$ if we wish to emphasise the function, but when just one function is under consideration we omit the subscript.

Note that $\underline{S}(\mathcal{D})$ and $\overline{S}(\mathcal{D})$ are the areas under the graphs of lower and upper approximations to f by piecewise constant functions. Note also that $\underline{S}(\mathcal{D})$, respectively $\overline{S}(\mathcal{D})$, are finite numbers if and only if f is bounded below, respectively above.

Clearly $\underline{S}(\mathcal{D}) \leq \overline{S}(\mathcal{D})$ for every dissection $\mathcal{D}$.

The next lemma and its corollary note how the lower and upper sums change when extra points are added to a dissection.

Lemma 4.3. *Let* $\mathcal{D}$ *be a dissection of* $[a,b]$ *and let* $\mathcal{D}'$ *be a dissection obtained from* $\mathcal{D}$ *by inserting a new point of dissection* a'. *Then*

$$\underline{S}(\mathcal{D}') \geq \underline{S}(\mathcal{D}) \quad \textit{and} \quad \overline{S}(\mathcal{D}') \leq \overline{S}(\mathcal{D}).$$

Proof. Let $\mathcal{D}$ be the dissection $a = a_0 < a_1 < a_2 < \cdots < a_{n-1} < a_n = b$ so $\mathcal{D}'$ is the dissection $a = a_0 < \cdots < a_k < a' < a_{k+1} < \cdots < a_n = b$ for some $0 \leq k \leq n-1$.

From the definition (4.1) the only difference between the lower sums $\underline{S}(\mathcal{D}')$ and $\underline{S}(\mathcal{D})$ results from the contributions between a_k and a_{k+1}. Thus

$$\begin{aligned}
\underline{S}(\mathcal{D}') - \underline{S}(\mathcal{D}) &= \left[(a' - a_k) \inf_{a_k \leq x \leq a'} f(x) + (a_{k+1} - a') \inf_{a' \leq x \leq a_{k+1}} f(x)\right] - (a_{k+1} - a_k) \inf_{a_k \leq x \leq a_{k+1}} f(x) \\
&\geq \left[(a' - a_k) \inf_{a_k \leq x \leq a_{k+1}} f(x) + (a_{k+1} - a') \inf_{a_k \leq x \leq a_{k+1}} f(x)\right] - (a_{k+1} - a_k) \inf_{a_k \leq x \leq a_{k+1}} f(x) \\
&= 0.
\end{aligned}$$

Similarly $\overline{S}(\mathcal{D}') - \overline{S}(\mathcal{D}) \leq 0$. □

Definition 4.4. (Refinement) *We say that a dissection $\mathcal{D}'$ is a* refinement *of a dissection $\mathcal{D}$ if every point in $\mathcal{D}$ is a point of $\mathcal{D}'$.*

Corollary 4.5. *Let $\mathcal{D}'$ be a refinement of $\mathcal{D}$. Then*

$$\underline{S}(\mathcal{D}') \geq \underline{S}(\mathcal{D}) \quad \textit{and} \quad \overline{S}(\mathcal{D}') \leq \overline{S}(\mathcal{D}).$$

Proof. This follows from Lemma 4.3 by repeatedly adding single points to the dissection. □

It is not surprising that upper sums are always greater than or equal to lower sums.

Proposition 4.6. *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be* any *dissections of $[a,b]$. Then*

$$\underline{S}(\mathcal{D}_1) \leq \overline{S}(\mathcal{D}_2).$$

Proof. Let $\mathcal{D}_1 \cup \mathcal{D}_2$ be the dissection consisting of the union of the points of dissection in $\mathcal{D}_1$ and $\mathcal{D}_2$. Then $\mathcal{D}_1 \cup \mathcal{D}_2$ is a refinement of *both* $\mathcal{D}_1$ and $\mathcal{D}_2$, so by Corollary 4.5

$$\underline{S}(\mathcal{D}_1) \leq \underline{S}(\mathcal{D}_1 \cup \mathcal{D}_2) \leq \overline{S}(\mathcal{D}_1 \cup \mathcal{D}_2) \leq \overline{S}(\mathcal{D}_2). \qquad \square$$

4.2 Definition of the integral

From Proposition 4.6,

$$\sup_{\mathcal{D}} \underline{S}(\mathcal{D}) \;\leq\; \inf_{\mathcal{D}} \overline{S}(\mathcal{D}). \tag{4.3}$$

It is easy to find functions for which there is strict inequality in (4.3). However, for very many functions we get equality and we take the common value of $\sup_{\mathcal{D}} \underline{S}(\mathcal{D})$ and $\inf_{\mathcal{D}} \overline{S}(\mathcal{D})$ to be the integral of the function.

Definition 4.7. (Integrable, integral) *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. If $\sup_{\mathcal{D}} \underline{S}_f(\mathcal{D}) = \inf_{\mathcal{D}} \overline{S}_f(\mathcal{D})$ we say that f is* Riemann integrable*, or that the* integral of f exists*, and we call the common value the* integral of f*. We denote this by the usual symbol, thus*

$$\int_a^b f(x)dx \;=\; \sup_{\mathcal{D}} \underline{S}_f(\mathcal{D}) \;=\; \inf_{\mathcal{D}} \overline{S}_f(\mathcal{D}). \tag{4.4}$$

To avoid a clutter of symbols we may just write $\int_a^b f$ or even $\int f$ when the interval of definition is clear,

If $b < a$ it is convenient to define $\int_a^b f(x)dx = -\int_b^a f(x)dx$.

It is often easier to work with a sequence of dissections rather than considering the inf or sup over all dissections.

Corollary 4.8. *A bounded function $f : [a,b] \to \mathbb{R}$ is integrable with $\int_a^b f = l$ if and only if there is a sequence of dissections $\mathcal{D}_n$ such that $\underline{S}(\mathcal{D}_n) \to l$ and $\overline{S}(\mathcal{D}_n) \to l$.*

In particular, f is integrable if and only if for all $\varepsilon > 0$ there is a dissection $\mathcal{D}$ such that $\overline{S}(\mathcal{D}) - \underline{S}(\mathcal{D}) < \varepsilon$.

Proof. '$\Longrightarrow$' If f is integrable then for each $n \in \mathbb{N}$ we may find dissections:

$\mathcal{D}'_n$ with $\underline{S}(\mathcal{D}'_n) \geq \sup_{\mathcal{D}} \underline{S}(\mathcal{D}) - \frac{1}{n} = l - \frac{1}{n}$

and $\mathcal{D}''_n$ with $\overline{S}(\mathcal{D}''_n) \leq \inf_{\mathcal{D}} \overline{S}(\mathcal{D}) + \frac{1}{n} = l + \frac{1}{n}$.

Let $\mathcal{D}_n = \mathcal{D}'_n \cup \mathcal{D}''_n$. Then

$$l - \tfrac{1}{n} \leq \underline{S}(\mathcal{D}'_n) \leq \underline{S}(\mathcal{D}_n) \leq \overline{S}(\mathcal{D}_n) \leq \overline{S}(\mathcal{D}''_n) \leq l + \tfrac{1}{n},$$

so $\underline{S}(\mathcal{D}_n),\ \overline{S}(\mathcal{D}_n) \to l$.

'$\Longleftarrow$' Since $\quad \underline{S}(\mathcal{D}_n) \leq \sup_{\mathcal{D}} \underline{S}(\mathcal{D}) \leq \inf_{\mathcal{D}} \overline{S}(\mathcal{D}) \leq \overline{S}(\mathcal{D}_n),$

if $\underline{S}(\mathcal{D}_n) \to l$ and $\overline{S}(\mathcal{D}_n) \to l$ then $\sup_{\mathcal{D}} \underline{S}(\mathcal{D}) = \inf_{\mathcal{D}} \overline{S}(\mathcal{D}) = l$, so f is integrable. $\square$

Using the definition of integral we can now integrate basic functions.

Example 4.9. *The function $f : [0,1] \to \mathbb{R}$ given by $f(x) = x^2$ is integrable with $\int_0^1 x^2 dx = \frac{1}{3}$.*

Calculation. For each $n \in \mathbb{N}$, let $\mathcal{D}_n$ be the dissection of $[0,1]$ into n equal parts, $0 < \frac{1}{n} < \frac{2}{n} < \cdots < \frac{n-1}{n} < 1$. Then

$$\begin{aligned}
\underline{S}(\mathcal{D}_n) &= \sum_{i=1}^{n} \frac{1}{n}\left(\frac{i-1}{n}\right)^2 \\
&= \frac{1}{n^3}\sum_{i=1}^{n}(i-1)^2 \\
&= \frac{1}{n^3} \times \frac{1}{6}(n-1)n(2n-1) \\
&= \frac{1}{6}\left(1-\frac{1}{n}\right)\left(2-\frac{1}{n}\right) \\
&\to \frac{1}{3} \qquad \text{as } n \to \infty.
\end{aligned}
\qquad\qquad
\begin{aligned}
\overline{S}(\mathcal{D}_n) &= \sum_{i=1}^{n} \frac{1}{n}\left(\frac{i}{n}\right)^2 \\
&= \frac{1}{n^3}\sum_{i=1}^{n} i^2 \\
&= \frac{1}{n^3} \times \frac{1}{6}n(n+1)(2n+1) \\
&= \frac{1}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right) \\
&\to \frac{1}{3} \qquad \text{as } n \to \infty.
\end{aligned}$$

[*Recall the sum of squares formula*: $\sum_{i=1}^n i^2 = \frac{1}{6}n(n+1)(2n+1)$.]

We conclude that $\int_0^1 x^2 dx$ exists and $\int_0^1 x^2 dx = \frac{1}{3}$. $\square$

Example 4.10. *Define $f : [0,1] \to \mathbb{R}$ by $f(x) = \begin{cases} 1 & (x \in \mathbb{Q}) \\ 0 & (x \notin \mathbb{Q}) \end{cases}$. Then f is not integrable.*

Calculation. Let $\mathcal{D}$ be *any* dissection $0 = a_0 < a_1 < \cdots < a_n = 1$. Since every interval contains both rational and irrational numbers, $\underline{S}(\mathcal{D}) = \sum_{i=1}^n (a_i - a_{i-1}) \times 0 = 0$ and $\overline{S}(\mathcal{D}) = \sum_{i=1}^n (a_i - a_{i-1}) \times 1 = a_n - a_0 = 1$. Thus $\sup_{\mathcal{D}} \underline{S}(\mathcal{D}) = 0$ and $\inf_{\mathcal{D}} \overline{S}(\mathcal{D}) = 1$ so f is not integrable. $\square$

Of course, many functions are integrable, in particular all continuous functions and all monotonic functions.

Proposition 4.11. (Continuous functions) *Let $f : [a,b] \to \mathbb{R}$ be continuous. Then f is integrable.*

Proof. By Theorem 3.18 f is *uniformly* continuous on $[a,b]$. Thus, given $\varepsilon > 0$ there exists $\delta > 0$ such that if $|x-y| < \delta$ then $|f(x)-f(y)| < \frac{\varepsilon}{b-a}$. Let $\mathcal{D}$ be any dissection of $[a,b]$ with intervals all $< \delta$ in length. Then:

$$\begin{aligned}\overline{S}(\mathcal{D}) - \underline{S}(\mathcal{D}) &= \sum_{i=1}^{n}(a_i - a_{i-1})\Big[\sup\{f(x) : a_{i-1} \le x \le a_i\} - \inf\{f(x) : a_{i-1} \le x \le a_i\}\Big] \\ &\le \sum_{i=1}^{n}(a_i - a_{i-1})\Big(\frac{\varepsilon}{b-a}\Big) \;=\; (b-a)\frac{\varepsilon}{b-a} \;=\; \varepsilon.\end{aligned}$$

This is true for all $\varepsilon > 0$ so $\sup_{\mathcal{D}} \underline{S}(\mathcal{D}) \;=\; \inf_{\mathcal{D}} \overline{S}(\mathcal{D})$ and f is integrable. □

Recall that a function $f : [a,b] \to \mathbb{R}$ is *increasing* if $f(x) \le f(y)$ for all $a \le x \le y \le b$ and is *decreasing* if $f(x) \ge f(y)$ for all $a \le x \le y \le b$. We call a function that is either increasing or decreasing *monotonic*.

Proposition 4.12. (Monotonic functions) *Let $f : [a,b] \to \mathbb{R}$ be monotonic. Then f is integrable.*

Proof. Assume that f is increasing; the proof for decreasing functions is similar. We may also assume that $f(a) < f(b)$ (otherwise f is constant and trivially integrable). Let $\varepsilon > 0$ and let $\mathcal{D}$ be a dissection of $[a,b]$ such that $a_i - a_{i-1} \le \frac{\varepsilon}{f(b)-f(a)}$ for all $i = 1, \ldots, n$. Then:

$$\begin{aligned}\overline{S}(\mathcal{D}) - \underline{S}(\mathcal{D}) &\le \sum_{i=1}^{n}(a_i - a_{i-1})f(a_i) \;-\; \sum_{i=1}^{n}(a_i - a_{i-1})f(a_{i-1}) \\ &= \sum_{i=1}^{n}(a_i - a_{i-1})\big[f(a_i) - f(a_{i-1})\big] \\ &\le \frac{\varepsilon}{f(b)-f(a)}\;\sum_{i=1}^{n}\big[f(a_i) - f(a_{i-1})\big] \;=\; \varepsilon.\end{aligned}$$

This is true for all $\varepsilon > 0$ so $\sup_{\mathcal{D}} \underline{S}(\mathcal{D}) \;=\; \inf_{\mathcal{D}} \overline{S}(\mathcal{D})$ and f is integrable. □

4.3 Properties of the integral

There are many familiar properties of the integral which need to be derived from the definition. We list these and indicate proofs. As before, we assume that our functions are bounded.

1. *If $f : [a,b] \to \mathbb{R}$ and $a < c < b$ then $\int_a^b f = \int_a^c f + \int_c^b f$, provided the integrals exist.*

Proof. Let $\mathcal{D}_n', \mathcal{D}_n''$ be dissections of $[a,c], [c,b]$ with

$$\underline{S}(\mathcal{D}_n') \to \textstyle\int_a^c f \;\text{ and }\; \overline{S}(\mathcal{D}_n') \to \int_a^c f,$$
$$\underline{S}(\mathcal{D}_n'') \to \textstyle\int_c^b f \;\text{ and }\; \overline{S}(\mathcal{D}_n'') \to \int_c^b f.$$

Let $\mathcal{D}_n$ be the dissection of $[a,b]$ given by the points of $\mathcal{D}_n', \mathcal{D}_n''$ together with the point c. Then

$$\underline{S}(\mathcal{D}_n) = \underline{S}(\mathcal{D}_n') + \underline{S}(\mathcal{D}_n'') \to \textstyle\int_a^c f + \int_c^b f,$$
$$\overline{S}(\mathcal{D}_n) = \overline{S}(\mathcal{D}_n') + \overline{S}(\mathcal{D}_n'') \to \textstyle\int_a^c f + \int_c^b f.$$

2. *Let f be integrable and $\lambda \in \mathbb{R}$. Then $\int_a^b \lambda f = \lambda \int_a^b f$.*

Proof. Follows since for every dissection $\underline{S}_{\lambda f}(\mathcal{D}) = \lambda \underline{S}_f(\mathcal{D})$ ($\lambda \geq 0$) and $\underline{S}_{\lambda f}(\mathcal{D}) = \lambda \overline{S}_f(\mathcal{D})$ ($\lambda < 0$) and similarly for the upper sums.

3. *Let $f, g : [a,b] \to \mathbb{R}$ be integrable. Then $f+g$ is integrable and $\int_a^b (f+g) = \int_a^b f + \int_a^b g$.*

Proof. Take dissections $\mathcal{D}_n', \mathcal{D}_n''$ with

$$\underline{S}_f(\mathcal{D}_n'), \overline{S}_f(\mathcal{D}_n') \to \int_a^b f \text{ and } \underline{S}_g(\mathcal{D}_n''), \overline{S}_g(\mathcal{D}_n'') \to \int_a^b g.$$

Let $\mathcal{D}_n = \mathcal{D}_n' \cup \mathcal{D}_n''$. Then

$$\underline{S}_f(\mathcal{D}_n), \overline{S}_f(\mathcal{D}_n) \to \int_a^b f \text{ and } \underline{S}_g(\mathcal{D}_n), \overline{S}_g(\mathcal{D}_n) \to \int_a^b g.$$

Note that

$$\sup\big[(f+g)(x) : a_{i-1} \leq x \leq a_i\big] \leq \sup\big[f(x) : a_{i-1} \leq x \leq a_i\big] + \sup\big[g(x) : a_{i-1} \leq x \leq a_i\big],$$

with the reverse inequality for 'inf', so that

$$\underline{S}_f(\mathcal{D}_n) + \underline{S}_g(\mathcal{D}_n) \leq \underline{S}_{f+g}(\mathcal{D}_n) \leq \overline{S}_{f+g}(\mathcal{D}_n) \leq \overline{S}_f(\mathcal{D}_n) + \overline{S}_g(\mathcal{D}_n),$$

from which $\underline{S}_{f+g}(\mathcal{D}_n), \overline{S}_{f+g}(\mathcal{D}_n) \to \int_a^b f + \int_a^b g$.

Note that (2) and (3) show that integration is linear, i.e. $\int_a^b (\lambda f + \mu g) = \lambda \int_a^b f + \mu \int_a^b g$.

4. *Let $f, g : [a,b] \to \mathbb{R}$ be integrable. Then fg is integrable.*

Proof. First we show that f^2 is integrable. Since f is integrable, it is bounded, so $|f(x)| \leq M$ for some M, for all $x \in [a,b]$. Then for all x, y,

$$|f(x)^2 - f(y)^2| = |f(x)+f(y)||f(x)-f(y)| \leq \big(|f(x)|+|f(y)|\big)|f(x)-f(y)| \leq 2M|f(x)-f(y)|.$$

In particular, for any interval I,

$$\sup_{x\in I}\big\{f(x)^2\big\} - \inf_{x\in I}\big\{f(x)^2\big\} \leq 2M\Big[\sup_{x\in I}\big\{f(x)\big\} - \inf_{x\in I}\big\{f(x)\big\}\Big].$$

Thus for every dissection $\mathcal{D}$

$$\overline{S}_{f^2}(\mathcal{D}) - \underline{S}_{f^2}(\mathcal{D}) \leq 2M\big[\overline{S}_f(\mathcal{D}) - \underline{S}_f(\mathcal{D})\big].$$

Given $\varepsilon > 0$ we can find a dissection such that $\overline{S}_f(\mathcal{D}) - \underline{S}_f(\mathcal{D}) < \varepsilon/2M$, so $\overline{S}_{f^2}(\mathcal{D}) - \underline{S}_{f^2}(\mathcal{D}) < \varepsilon$, and it follows by Corollary 4.8 that f is integrable.

The conclusion for general products now follows using (2) and (3), since $fg = \frac{1}{4}\big[(f+g)^2 - (f-g)^2\big]$.

5. *Let $f : [a,b] \to \mathbb{R}$ be integrable. Then $|f|$ is integrable and*

$$\Big|\int_a^b f\Big| \leq \int_a^b |f|.$$

Proof. Note that for any interval I,

$$\sup_{x\in I}\big\{|f(x)|\big\} - \inf_{x\in I}\big\{|f(x)|\big\} \leq \sup_{x\in I}\big\{f(x)\big\} - \inf_{x\in I}\big\{f(x)\big\},$$

so for every dissection $\mathcal{D}$

$$\overline{S}_{|f|}(\mathcal{D}) - \underline{S}_{|f|}(\mathcal{D}) \leq \overline{S}_f(\mathcal{D}) - \underline{S}_f(\mathcal{D}).$$

Thus if f is integrable so is $|f|$ by Corollary 4.8. Also $\underline{S}_f(\mathcal{D}) \leq \underline{S}_{|f|}(\mathcal{D})$ for all dissections $\mathcal{D}$, so taking suprema over dissections gives the inequality.

6. *Let $f : [a,b] \to \mathbb{R}$ be integrable with $M_1 \leq f(x) \leq M_2$ for all $x \in [a,b]$. Then*

$$M_1(b-a) \leq \int_a^b f \leq M_2(b-a).$$

In particular, if $f \geq 0$ then $\int_a^b f \geq 0$.

Proof. This follows since for any dissection $\mathcal{D}$, $M_1(b-a) \leq \underline{S}(\mathcal{D}) \leq \overline{S}(\mathcal{D}) \leq M_2(b-a)$.

7. (Cauchy-Schwartz inequality) *Let $f, g : [a,b] \to \mathbb{R}$ be integrable.Then*

$$\Big(\int_a^b |fg|\Big)^2 \leq \int_a^b f^2 \int_a^b g^2.$$

Proof. Note that for all $\lambda \in \mathbb{R}$

$$0 \leq \int_a^b (\lambda|f| + |g|)^2 = \lambda^2 \int_a^b f^2 + 2\lambda \int_a^b |fg| + \int_a^b g^2.$$

The inequality now follows from the discriminant condition '$b^2 \leq 4ac$' for the quadratic expression to be non-negative for all λ.

4.4 Integration and differentiation

Given the different ways in which they are defined, it is remarkable that integration and differentiation are 'inverse' operations. This is made precise in the 'fundamental theorem of calculus' which is conveniently stated in two parts.

Note that in (4.5) where the upper limit of integration is a variable x, we often just write $\int f$ to denote this function of x in the usual way.

Theorem 4.13. (Fundamental theorem of calculus – Part 1) *Let $f : [a,b] \to \mathbb{R}$ be integrable. Define*

$$F(x) = \int_a^x f(t)dt \qquad (a \leq x \leq b). \tag{4.5}$$

Then F is continuous on $[a,b]$. Moreover, F is differentiable at all x at which f is continuous, with

$$F'(x) = f(x). \tag{4.6}$$

Proof. Since every integrable function is bounded, $|f(x)| \leq M$ for some $M > 0$. Using the properties (1), (5), (6) of the integral, if $x, x+h \in [a,b]$,

$$\big|F(x+h) - F(x)\big| = \Big|\int_a^{x+h} f - \int_a^x f\Big| = \Big|\int_x^{x+h} f\Big| \leq \Big|\int_x^{x+h} |f|\Big| \leq Mh.$$

Thus given $\varepsilon > 0$, if $|h| < \varepsilon/M$ then $\big|F(x+h) - F(x)\big| < M \times (\varepsilon/M) = \varepsilon$, so F is continuous at x and so is continuous on $[a,b]$.

Now let x be a point at which f is continuous. Given $\varepsilon > 0$, there exists $\delta > 0$ such that if $|y-x| < \delta$ then $f(x)-\varepsilon < f(y) < f(x)+\varepsilon$. By property (6), if $0 < h < \delta$,

$$h(f(x)-\varepsilon) \le \int_x^{x+h} f, \ \int_{x-h}^{x} f \le h(f(x)+\varepsilon),$$

that is, if $0 < |h| < \delta$,

$$f(x)-\varepsilon \le \frac{F(x+h)-F(x)}{h} \ = \ \frac{1}{h}\int_x^{x+h} f \ \le f(x)+\varepsilon.$$

Thus $\lim_{h\to 0} \dfrac{F(x+h)-F(x)}{h} = f(x)$, so F is differentiable at x with derivative f. □

Theorem 4.14. (Fundamental theorem of calculus – Part 2) *Let $f : [a,b] \to \mathbb{R}$ be differentiable with derivative f' continuous on $[a,b]$. Then*

$$\int_a^x f' \ = \ f(x)-f(a) \ \equiv \ \big[f\big]_a^x. \tag{4.7}$$

Proof. Let $\phi(x) = \int_a^x f'$. By Theorem 4.13, since f' is continuous, ϕ is differentiable with $\phi'(x) = f'(x)$ for all $x \in [a,b]$. Hence $\phi'(x) - f'(x) \equiv 0$, so by the mean value theorem $\phi(x) - f(x)$ is constant on $[a,b]$. In particular,

$$\phi(x) - f(x) \ = \ \phi(a) - f(a) \ = \ -f(a),$$

that is

$$\int_a^x f' \ = \ \phi(x) \ = \ f(x)-f(a) \qquad (x \in [a,b]). \qquad \square$$

It follows from Theorem 4.13 that to integrate a function f it is enough to find a function $F(x)$ with derivative $f(x)$ and add (or subtract) a constant so that $F(a) = 0$. Of course, we conventionally write $\int f$ for the *indefinite integral* F which is a function as opposed to the *definite integral* $\int_a^b f$ which is a number. Thus

$$\int_1^x (4x^3 - 6x^2 + 2x - 1)dx \ = \ x^4 - 2x^3 + x^2 - x + 1,$$

by checking the derivative of the RHS and getting the constant term by setting $x = 1$.

The standard rules for integration follow from the Fundamental Theorem of Calculus

Proposition 4.15. (Integration by substitution) *Let $g : [a,b] \to \mathbb{R}$ be differentiable with g' continuous, and let $f : g([a,b]) \to \mathbb{R}$ be continuous. Then*

$$\int_{g(a)}^{g(b)} f(x)dx \ = \ \int_a^b f\big(g(y)\big)g'(y)dy. \tag{4.8}$$

Proof. Write $F(x) = \int_c^x f$ for some c, so $F'(x) = f(x)$ by (4.6). Then

$$\begin{aligned}
\int_a^b f\big(g(y)\big)g'(y)dy \ &= \ \int_a^b F'\big(g(y)\big)g'(y)dy \\
&= \ \int_a^b (F\circ g)'(y)dy && \text{(chain rule)} \\
&= \ F\big(g(b)\big) - F\big(g(a)\big) && \text{(by(4.7))} \\
&= \ \int_c^{g(b)} f - \int_c^{g(a)} f && \text{(definition of } F) \\
&= \ \int_{g(a)}^{g(b)} f. && \square
\end{aligned}$$

Example

$$\int_0^{\sqrt{\pi}} \sin(y^2)\ y\ dy = \int_0^{\sqrt{\pi}} \tfrac{1}{2}\sin(y^2)\ 2y\ dy = \int_0^{\pi} \tfrac{1}{2}\sin x\ dx = \big[-\tfrac{1}{2}\cos x\big]_0^{\pi} = 1,$$

where we have taken $f(x) = \frac{1}{2}\sin x$, $g(y) = y^2$, $a = 0$ and $b = \sqrt{\pi}$ in (4.8).

Proposition 4.16. (Integration by parts) *Let $f, g : [a,b] \to \mathbb{R}$ be differentiable with f', g' continuous. Then*

$$\int_a^b fg' \ = \ \big[fg\big]_a^b \ - \ \int_a^b f'g.$$

Proof. From the product rule for differentiation, $(fg)' = fg' + f'g$. By (4.7)

$$\big[fg\big]_a^b \ = \ \int_a^b (fg)' \ = \ \int_a^b fg' + \int_a^b f'g. \quad \square$$

4.5 Improper integrals

So far we have assumed that we are integrating over a bounded interval $[a,b]$ and that the integrand f is bounded. We can remove these restrictions by defining integrals as the limit of integrals of bounded functions over bounded integrals.

Definition 4.17. (Improper integral - unbounded domain) *Let $f : [a,\infty) \to \mathbb{R}$ be a bounded function. We define*

$$\int_a^{\infty} f(x)\ dx = \lim_{X\to\infty} \int_a^X f(x)\ dx,$$

provided that the right-hand integrals exist for all X and the limit exists.

Example

$$\int_1^{\infty} \frac{dx}{x^2} \ = \ \lim_{X\to\infty} \int_1^X \frac{dx}{x^2} \ = \ \lim_{X\to\infty} \Big[-\frac{1}{x}\Big]_1^X \ = \ \lim_{X\to\infty} \Big[1 - \frac{1}{X}\Big] \ = \ 1. \qquad \square$$

Domains of the form $(-\infty, b]$ or $(-\infty,\infty)$ may be treated in a similar way.

Definition 4.18. (Improper integral - unbounded function) *Let $f : [a,b] \to \mathbb{R}$ be bounded below but not above. We define*

$$\int_a^b f(x)\ dx = \lim_{M\to\infty} \int_a^b f_M(x)\ dx,$$

where $f_M(x) = \min\{f(x), M\}$, provided that the right-hand integral exists for all M and the limit exists.

Example To find $\displaystyle\int_0^1 \frac{dx}{\sqrt{x}}$ let $f_M(x) = \begin{cases} M & (0 \le x \le 1/M^2) \\ 1/\sqrt{x} & (1/M^2 \le x \le 1) \end{cases}$ for $M > 1$. Then

$$\int_0^1 f_M(x)dx = \int_0^{1/M^2} M dx + \int_{1/M^2}^1 \frac{dx}{\sqrt{x}} \ = \ \frac{1}{M^2} \times M + \Big[2x^{1/2}\Big]_{1/M^2}^1 \ = \ \frac{1}{M} + \Big[2 - \frac{2}{M}\Big] \to 2. \quad \square$$

5 Sequences and series of functions and power series

The ultimate aim of this chapter is to determine when we can take a convergent power series and differentiate or integrate it term by term to get a series summing to the derivative or integral of the original series sum. First, we consider sequences of functions which will eventually play the role of partial sums of series.

5.1 Convergence of sequences of functions

Consider a sequence of functions $f_n : [a,b] \to \mathbb{R}$ (or $\mathbb{R} \to \mathbb{R}$). What does it mean to say that (f_n) converges to a function f? There are many possible ways of defining this, the most obvious definition just requires convergence of the sequence of numbers $f_n(x)$ at every x in the domain.

Definition 5.1. (Pointwise convergence) *Let $f_n, f : [a,b] \to \mathbb{R}$ (or $\mathbb{R} \to \mathbb{R}$) . We say that $f_n \to f$* pointwise *if for all $x \in [a,b]$ (or $\mathbb{R}$), $f_n(x) \to f(x)$ (as a sequence of real numbers).*

Examples.

1. $\dfrac{x^2}{n+x^2} \to 0$ pointwise on $\mathbb{R}$, since, given $\varepsilon > 0$, $\left|\dfrac{x^2}{n+x^2} - 0\right| \leq \dfrac{x^2}{n} < \varepsilon$ if $n > x^2/\varepsilon$.

However, the following examples show that pointwise convergence does not behave very well. (Drawing graphs may help you see what is happening.)

2. Let $f_n : \mathbb{R} \to \mathbb{R}$ be $f_n(x) = \begin{cases} 0 & (x \leq 0) \\ nx & (0 \leq x \leq 1/n) \\ 1 & (1/n \leq x) \end{cases}$.

   Then $f_n \to f$ pointwise where $f(x) = \begin{cases} 0 & \text{if} \quad (x \leq 0) \\ 1 & \text{if} \quad (0 < x) \end{cases}$. Note that f is not continuous even though every f_n is continuous.

3. Let $f_n : \mathbb{R} \to \mathbb{R}$ be $f_n(x) = \dfrac{nx^2}{1+nx^2}$.

   Then $f_n(0) = 0 \to 0$, but $f_n(x) \to 1$ if $x \neq 0$, since given $\varepsilon > 0$, if $n > 1/(\varepsilon x^2)$,

   $$|f_n(x) - 1| = \left|\frac{nx^2}{1+nx^2} - 1\right| = \frac{1}{1+nx^2} < \frac{1}{nx^2} < \varepsilon.$$

   Again, f is not continuous even though every f_n is continuous.

4. Let $f_n : [0,1] \to \mathbb{R}$ be $f_n(x) = \begin{cases} n^2 x & (0 \leq x \leq 1/n) \\ n^2(2/n - x) & (1/n \leq x \leq 2/n) \\ 0 & (2/n \leq x \leq 1) \end{cases}$.

   Then $f_n(x) \to 0$ for all $x \in [0,1]$. But for all n

   $$1 = \int_0^1 f_n \not\to \int_0^1 0 = 0.$$

5. Let $f_n : \mathbb{R} \to \mathbb{R}$ be $f_n(x) = \frac{1}{n}\sin nx$. Then $|f_n(x)| \leq \frac{1}{n} \to 0$ so $f_n \to 0$ pointwise, but $f_n'(x) = \cos nx \not\to 0 = \frac{d}{dx}0$.

These examples show that continuity, integration and differentiation do not behave well on taking pointwise limits of a sequence of functions. To overcome such problems we introduce a stronger form of convergence, which requires the same 'N' to apply at all points x simultaneously in the definition of convergence.

Definition 5.2. (Uniform convergence) *Let $f_n : [a,b] \to \mathbb{R}$ (or $\mathbb{R} \to \mathbb{R}$). We say that $f_n \to f$* uniformly *if given $\varepsilon > 0$ there exists N such that for all $n \geq N$ and all $x \in [a,b]$ (or $\mathbb{R}$), we have $|f_n(x) - f(x)| < \varepsilon$.*

Thus for pointwise convergence, 'N' depends on x, whilst for uniform convergence the same N is required for all x.

Note that if f_n converges to f uniformly then $f_n \to f$ pointwise, but the converse is not true. In the above examples, the sequnce in (5) converges uniformly, but those in (1), (2), (3), (4) do not.

Example. Let $f_n : [0,5] \to \mathbb{R}$ be $f_n(x) = \dfrac{4n+3\sin x}{2n+x^2}$. Then $f_n(x) \to 2$ uniformly on $[0,5]$. For given $\varepsilon > 0$,

$$\left|\frac{4n+3\sin x}{2n+x^2} - 2\right| = \left|\frac{3\sin x - 2x^2}{2n+x^2}\right| \leq \frac{|3\sin x| + 2|x^2|}{2n} \leq \frac{53}{2n} < \varepsilon$$

if $n > 53/2\varepsilon$; this is independent of $x \in [0,1]$, so convergence is uniform.

The first advantage of uniform convergence is that it behaves well with respect to continuity. The following result is often stated succinctly as 'The uniform limit of a sequence of continuous functions is continuous'.

Theorem 5.3. (Continuity and uniform convergence) *Let $f, f_n : [a,b] \to \mathbb{R}$ with $f_n \to f$ uniformly. If f_n is continuous for all n, then f is continuous.*

Proof. Let $c \in [a,b]$. Given $\varepsilon > 0$ we may choose $N \in \mathbb{N}$ such that $|f_N(x) - f(x)| < \frac{1}{3}\varepsilon$ for *all* $x \in [a,b]$ by uniform continuity. Since f_N is continuous at c there exists $\delta > 0$ such that $|f_N(x) - f_N(c)| < \frac{1}{3}\varepsilon$ if $|x - c| < \delta$. Thus, if $|x - c| < \delta$,

$$\begin{aligned}
|f(x) - f(c)| &= \big|f(x) - f_N(x) + f_N(x) - f_N(c) + f_N(c) - f(c)\big| \\
&\leq \big|f(x) - f_N(x)\big| + \big|f_N(x) - f_N(c)\big| + \big|f_N(c) - f(c))\big| \quad \text{(triangle inequality)} \\
&< \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon \;=\; \varepsilon.
\end{aligned}$$

Hence f is continuous at c. $\square$

Uniform convergence also behaves well with respect to integration.

Theorem 5.4. (Integration and uniform convergence) *Let $f_n : [a,b] \to \mathbb{R}$ be integrable, and $f_n \to f$ uniformly. Then f is integrable, and $\displaystyle\int_a^b f_n \to \int_a^b f$.*

Proof. Given $\varepsilon > 0$, by uniform convergence choose N such that $|f_n(x) - f(x)| < \frac{1}{3(b-a)}\varepsilon$ for all $x \in [a,b]$ and $n \geq N$. Take a dissection $\mathcal{D}$ such that $\overline{S}_{f_N}(\mathcal{D}) - \underline{S}_{f_N}(\mathcal{D}) < \frac{1}{3}\varepsilon$.
Since $f_N(x) > f(x) - \frac{1}{3(b-a)}\varepsilon$,

$$\overline{S}_{f_N}(\mathcal{D}) > \overline{S}_f(\mathcal{D}) - (b-a)\frac{1}{3(b-a)}\varepsilon = \overline{S}_f(\mathcal{D}) - \tfrac{1}{3}\varepsilon.$$

Similarly

$$\underline{S}_{f_N}(\mathcal{D}) < \underline{S}_f(\mathcal{D}) + \tfrac{1}{3}\varepsilon.$$

Thus

$$\overline{S}_f(\mathcal{D}) - \underline{S}_f(\mathcal{D}) \leq \overline{S}_{f_N}(\mathcal{D}) + \tfrac{1}{3}\varepsilon - \underline{S}_{f_N}(\mathcal{D}) + \tfrac{1}{3}\varepsilon < \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon = \varepsilon.$$

We can find such a dissection for all $\varepsilon > 0$, so f is integrable.

Moreover, for all $n \geq N$,

$$f_n(x) - \frac{1}{3(b-a)}\varepsilon \leq f(x) \leq f_n(x) + \frac{1}{3(b-a)}\varepsilon,$$

so integrating

$$\int_a^b f_n - \tfrac{1}{3}\varepsilon \leq \int_a^b f \leq \int_a^b f_n + \tfrac{1}{3}\varepsilon$$

for all $n \geq N$. Thus $\int_a^b f_n \to \int_a^b f$. $\square$

Example. From the uniform convergence in the previous example,

$$\int_0^5 \frac{4n + 3\sin x}{2n + x^2} dx \to \int_0^5 2dx = 10.$$

We need to be a little careful differentiating sequences of functions and taking the limit - here we need uniform convergence of the derivative rather than just of the function.

Corollary 5.5. (Differentiation and uniform convergence) *Let $f_n : [a,b] \to \mathbb{R}$ be differentiable with f_n' continuous for all n. Suppose $f_n \to f$ pointwise and f_n' converges uniformly on $[a,b]$. Then $f : [a,b] \to \mathbb{R}$ is differentiable and $f'(x) = \lim_{n\to\infty} f_n'(x)$ for all $x \in [a,b]$.*

Proof. Let $f_n' \to g$ uniformly on $[a,b]$. Then

$$\begin{aligned}
\int_a^x g &= \lim_{n\to\infty} \int_a^x f_n' && \text{(as } f_n' \to g \text{ uniformly)} \\
&= \lim_{n\to\infty} \left[f_n(x) - f_n(a)\right] && \text{(FTC-2)} \\
&= \lim_{n\to\infty} f_n(x) - \lim_{n\to\infty} f_n(a) \\
&= f(x) - f(a).
\end{aligned}$$

As g is continuous (as the uniform limit of continuous functions) applying FTC-1 gives $f'(x) = g(x) = \lim_{n\to\infty} f_n'(x)$. $\square$

5.2 Convergence of series of functions

Like sequences of real numbers, sequences of functions often occur in the form of partial sums of series, in particular of power series. We can formulate the results of the previous section in terms of series of functions.

Definition 5.6. (Uniform convergence of series) *Let $f_n : [a,b] \to \mathbb{R}$ (or $\mathbb{R} \to \mathbb{R}$) and define the partial sums $S_N(x) = \sum_{n=1}^{N} f_n(x)$. If S_N tends to a function S uniformly on E, we say that the series $\sum_{n=1}^{\infty} f_n$* converges uniformly *on $[a,b]$ (or $\mathbb{R}$) with sum S.*

We can read off properties of integration and differentiation of series from the corresponding results for sequences of functions in the last section.

Corollary 5.7. *Let $S(x) = \sum_{n=1}^{\infty} f_n(x)$ be a uniformly convergent series on $[a,b]$.*

(1) If f_n is continuous for all n then the sum S is continuous,

(2) if f_n is integrable for all n then the sum S is integrable with $\int_a^b S(x)dx = \sum_{n=1}^{\infty} \int_a^b f_n(x)dx$,

(3) if f_n is differentiable with f_n' continuous for all n and $\sum_{n=1}^{\infty} f_n'(x)$ is uniformly convergent on $[a,b]$, then the sum S is differentiable with $S'(x) = \sum_{n=1}^{\infty} f_n'(x)$.

Proof. (1) and (2) follow from the definiton of uniform convergence of a series by applying Theorems 5.3 and 5.4 to the partial sums $S_N(x) = \sum_{n=1}^{N} f_n(x)$ which are uniformly convergent to $S(x)$.

For (3), we have $S_N(x) \to S(x)$ pointwise and $S_N'(x) = \sum_{n=1}^{N} f_n'(x)$ continuous and uniformly convergent on $[a,b]$. By Corollary 5.5 $S(x)$ is differentiable with $S'(x) = \lim_{N\to\infty} S_N'(x) = \lim_{N\to\infty} \sum_{n=1}^{N} f_n'(x) = \sum_{n=1}^{\infty} f_n'(x)$. □

The Weierstrass M-test gives a very useful way of checking uniform convergence of series. Note that here and elsewhere, if $a_n \geq 0$ for all n, we follow the usual convention of writing $\sum_{n=1}^{\infty} a_n < \infty$ to mean that the series of non-negative terms is convergent.

Proposition 5.8. (Weierstrass M-test) *Let $f_n : [a,b] \to \mathbb{R}$ (or $\mathbb{R} \to \mathbb{R}$). Suppose that there is a sequence of non-negative numbers $(M_n)_{n\in\mathbb{N}}$ such that $\sum_{n=1}^{\infty} M_n < \infty$ and $|f_n(x)| \leq M_n$ for all $n \in \mathbb{N}$ and $x \in [a,b]$ (or $\mathbb{R}$). Then the series $\displaystyle\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on $[a,b]$ (or $\mathbb{R}$).*

Proof. First note that for each $x \in [a,b]$ the series $\sum_{n=1}^{\infty} |f_n(x)|$ is convergent by the comparison test (comparing with $\sum_{n=1}^{\infty} M_n$), so we may define the sum $S(x) = \sum_{n=1}^{\infty} f_n(x)$ for each x.

Since $\sum_1^{\infty} M_n$ is convergent, given $\varepsilon > 0$ we may find N such that $\sum_{N+1}^{\infty} M_n < \varepsilon$. Then for all $x \in [a,b]$,

$$\left|S(x) - \sum_{1}^{N} f_n(x)\right| = \Big|\sum_{N+1}^{\infty} f_n(x)\Big| \;\leq\; \sum_{N+1}^{\infty} |f_n(x)| \;\leq\; \sum_{N+1}^{\infty} M_n < \varepsilon.$$

Hence the partial sums $\sum_{n=1}^{N} f_n(x)$ converge uniformly to $S(x)$, so the series is uniformly convergent. □

Example. For $0 \leq x \leq 1$, $\displaystyle\Big|\frac{x}{(n+x^2)^2}\Big| \;\leq\; \frac{1}{n^2}$ and $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$. By the Weierstrass M-test, $\displaystyle\sum_{n=1}^{\infty} \frac{x}{(n+x^2)^2}$ is uniformly convergent on $[0,1]$. Thus we may integrate term by term to get

$$\int_0^1 \Big(\sum_{n=1}^{\infty} \frac{x}{(n+x^2)^2} dx\Big) \;=\; \sum_{n=1}^{\infty} \int_0^1 \frac{x\,dx}{(n+x^2)^2} \;=\; \sum_{n=1}^{\infty} \Big[\frac{-1}{2(n+x^2)}\Big]_0^1 \;=\; \sum_{n=1}^{\infty} \frac{1}{2}\Big(\frac{1}{n} - \frac{1}{n+1}\Big) \;=\; \frac{1}{2}.$$

Example. Fix $a > 1$. Then $|n^{-x}| \le n^{-a}$ if $x \ge a$ and $\sum_{n=1}^{\infty} \frac{1}{n^a} < \infty$, so by the Weierstrass M-test, the *Riemann zeta function*

$$\zeta(x) \;=\; \sum_{n=1}^{\infty} \frac{1}{n^x} \;=\; 1 + \frac{1}{2^x} + \frac{1}{3^x} + \frac{1}{4^x} + \cdots$$

is uniformly convergent on $[a, \infty)$. We may differentiate term by term to get

$$\zeta'(x) \;=\; \sum_{n=1}^{\infty} \frac{-\log n}{n^x} \;=\; \frac{-\log 2}{2^x} + \frac{-\log 3}{3^x} + \frac{-\log 4}{4^x} + \cdots,$$

for all $x \in [a, \infty)$, since $\sum_{n=1}^{\infty} \frac{1}{n^a} |-\log n| < \infty$, and thus for all $x > 1$ since this is valid for all $a > 1$.

5.3 Power series

The most important examples of series of functions are *power series*, that is series of the form

$$\sum_{n=0}^{\infty} a_n x^n \;\equiv\; a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots \qquad \text{where } a_n \in \mathbb{R},$$

corresponding to $f_n(x) = a_n x^n$ in the previous section.

Lemma 5.9. *Suppose that the power series* $\sum_{n=0}^{\infty} a_n x^n$ *converges for* $x = x_0$ *for some* $x_0 \ne 0$. *Then for all* $0 < r < |x_0|$ *the series converges uniformly and absolutely on* $[-r, r]$ *and is continuous on* $[-r, r]$.

Proof. The terms of a convergent series are bounded, so there is some number M such that $|a_n x_0^n| \le M$ for all n. Then for $|x| \le r < |x_0|$

$$|a_n x^n| \;\le\; |a_n r^n| \;=\; |a_n x_0^n| \frac{r^n}{|x_0|^n} \le M\left(\frac{r}{|x_0|}\right)^n.$$

Since $\sum_{n=0}^{\infty} M\left(\frac{r}{|x_0|}\right)^n < \infty$ is a convergent geometric series with ratio $\frac{r}{|x_0|} < 1$, it follows from the Weierstrass M-test that $\sum_{n=0}^{\infty} |a_n x^n|$ is uniformly convergent on $[-r, r]$. $\square$

It now follows easily that a series converges uniformly strictly inside its 'interval of convergence'.

Theorem 5.10. (Uniform convergence of power series) *For a power series* $\sum_{n=0}^{\infty} a_n x^n$, *one of the following is true:*

(a) the series is uniformly and absolutely convergent on the interval $[-r, r]$ *for every* $r > 0$.
(b) the series converges only when $x = 0$,
(c) there is a number $R > 0$ *such that the series is uniformly and absolutely convergent on every interval* $[-r, r]$ *with* $0 < r < R$ *and diverges if* $|x| > R$.

Proof. Let $A = \{|x| : \sum_{n=0}^{\infty} a_n x^n \text{ converges}\}$. If $A = \{0\}$ we are in case (b). If A is unbounded, then the conclusion of case (a) follows from Lemma 5.9. Otherwise, A is bounded and, taking $R = \sup A$, the conclusion of case (c) follows from Lemma 5.9. $\square$

Definition 5.11. (Radius of convergence) *The number R in (c) is called the* radius of convergence *of the series; for convenience we say that the radius of convergence is* 0 *if case (b) holds and is ∞ in case (a).*

In case (c) there is an interval of convergence *on which the series converges, and this will be $(-R,R), (-R,R], [-R,R)$ or $[-R,R]$ depending on whether the series converges at the end points of the intervals.*

Recall that the radius of convergence of a series may often be found using the ratio test.

Proposition 5.12. (Ratio test) *Let $\sum_{n=1}^{\infty} c_n$ be a series with $c_n \neq 0$ for all n. Suppose $\lim_{n\to\infty} \left|\frac{c_{n+1}}{c_n}\right| = L$. Then (a) if $L < 1$ the series is absolutely convergent and thus convergent,*

(b) if $L > 1$ the series is divergent,

(c) if $L = 1$ no conclusion can be drawn.

Proof. Done in earlier courses. $\square$

We now show that the radius of convergence of the 'differentiated' and 'integrated' series are the same as that of a given series.

Proposition 5.13. (Radii of convergence of differentiated and integrated series) *The power series $\sum_{n=0}^{\infty} a_n x^n$, the differentiated series $\sum_{n=0}^{\infty} n\, a_n x^{n-1}$ and the integrated series $\sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$ all have the same radii of convergence.*

Proof. Let $0 < |x| < |x_1|$. There is a constant M such that $\frac{n}{|x|}\left(\frac{|x|}{|x_1|}\right)^n \leq M$ for all $n \in \mathbb{N}$. Then

$$\frac{1}{|x|}|a_n x^n| \;\leq |n\, a_n x^{n-1}| \;=\; \frac{n}{|x|}\left(\frac{|x|}{|x_1|}\right)^n |a_n x_1^n| \;\leq\; M|a_n x_1^n|.$$

Let R and R' be the radii of convergence of the original series and of the differentiated series. If $|x| < R'$ then $\sum_{n=1}^{\infty} |n\, a_n x^{n-1}|$ converges, so by the comparison test $\sum_{n=0}^{\infty} |a_n x^n|$ converges, giving $R \geq R'$. Similarly, if $|x_1| < R$ then $\sum_{n=0}^{\infty} |a_n x_1^n|$ converges so by the comparison test $\sum_{n=1}^{\infty} |n\, a_n x^{n-1}|$ converges for all $|x| < |x_1|$ giving $R' \geq R$. Thus $R' = R$.

The proof for the integrated series is similar; alternatively set $na_n = b_{n-1}$ $(n \geq 1)$ in the differentiated series. $\square$

That power series can be differentiated and integrated term by term inside their interval of convergence now follows from the results for general series of functions.

Corollary 5.14. (Integration and differentiation of power series) *Let the power series* $\sum_{n=0}^{\infty} a_n x^n$ *have radius of convergence R (possibly* ∞*) and write* $S(x) = \sum_{n=0}^{\infty} a_n x^n$ *for the sum. Then for* $|x| < R$*, the sum S is continuous at x, and*

$$\int_0^x S = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1},$$

and

$$S'(x) = \sum_{n=0}^{\infty} n\, a_n\, x^{n-1}.$$

Proof. The integrated and differentiated series also have radius of convergence R by Proposition 5.13. Take r such that $|x| < r < R$. These series are uniformly convergent on $[-r, r]$ by Theorem 5.10. The conclusion is immediate from Corollary 5.7 taking $f_n(x) = a_n x^n$. $\square$

Example. The geometric series

$$1 - x^2 + x^4 - x^6 + \cdots = \frac{1}{1+x^2}$$

has radius of convergence 1 (and interval of convergence $(-1,1)$). Thus for every $0 < r < 1$, the series is uniformly convergent on $[-r, r]$ by Proposition 5.10. By Corollary 5.14, if $|x| < 1$,

$$x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots \; = \; \int_0^x \frac{dx}{1+x^2} = \tan^{-1} x. \tag{5.9}$$

In fact this series also sums to $\tan^{-1} x$ when $x = 1$ which is Gregory's series for $\pi = 4\tan^{-1} 1$.

It may be shown, using trig formulae or complex numbers, that

$$\frac{\pi}{4} = 4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239}.$$

Using the series (5.9) to estimate the two terms in this identity can give very accurate approximations to π; indeed until relatively recently the value of π to the greatest number of decimal places was found in this way.

Example. Fix $0 < a < 1$. The geometric series

$$1 - x + x^2 - x^3 + \cdots = \frac{1}{1+x}$$

has radius of convergence 1 (and interval of convergence $(-1,1)$). Thus for every $0 < r < 1$, the series is uniformly convergent on $[-r, r]$ by Proposition 5.10. By Corollary 5.14, we get the logarithmic series for $|x| < 1$,

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \; = \; \int_0^x \frac{dx}{1+x} = \log(1+x).$$

6 Metric and normed spaces

The aim of this chapter is to show that ideas such as convergence, continuity, etc, extend very easily to a vastly wider setting, that includes N-dimensional Euclidean space $\mathbb{R}^N$, spaces of sequences and spaces of functions. We introduce the very general concept of a metric space, where there is a notion of distance which satisfies some very basic properties, and in particular consider normed spaces where there is an underlying vector space.

6.1 Metric spaces

Any reasonable notion of the distance between points of a set should satisfy three basic and natural conditions.

Definition 6.1. (Metric space) *Let X be a non-empty set. A* metric *or* distance function *d on X is a function $d : X \times X \to \mathbb{R}$ such that*

(M1) $\quad d(x,y) \geq 0 \quad$ *for all $x,y \in X$, with $d(x,y) = 0$ if and only if $x = y$,* $\quad$ (positivity)

(M2) $\quad d(x,y) = d(y,x) \quad$ *for all $x,y \in X$,* $\quad$ (symmetry)

(M3) $\quad d(x,y) \leq d(x,z) + d(z,y) \quad$ *for all $x,y,z \in X$.* $\quad$ (triangle inequality)

The pair (X,d) is called a metric space.

Thinking of $d(x,y)$ as the distance between two points, these three conditions have obvious interpretations. (M1) says that all distances are non-negative and are strictly positive just when the points are distinct. (M2) says that the distance from one point to another is the same as the distance going the other way. The triangle inequality (M3) says that the length of a journey cannot decrease if a specified intermediate point is visited on the way.

For most examples, conditions (M1) and (M2) are virtually obvious but some checking may be needed for (M3).

Examples.

1. $d(x,y) = |x-y|$ is a metric on $\mathbb{R}$.

   (M1) and (M2) are trivial, for (M3) note that for all $x,y,z \in \mathbb{R}$,

   $$d(x,y) = |x-y| = |(x-z)+(z-y)| \leq |x-z| + |z-y| = d(x,z) + d(z,y)$$

   using the triangle inequality for real numbers.

2. $d(x,y) = \left|\frac{1}{x} - \frac{1}{y}\right|$ is a metric on $[1,\infty)$.

   (M1) and (M2) are easy, for (M3), for all $x,y,z \in \mathbb{R}$,

   $$d(x,y) = \left|\tfrac{1}{x} - \tfrac{1}{y}\right| = \left|\left(\tfrac{1}{x} - \tfrac{1}{z}\right) + \left(\tfrac{1}{z} - \tfrac{1}{y}\right)\right| \leq \left|\tfrac{1}{x} - \tfrac{1}{z}\right| + \left|\tfrac{1}{z} - \tfrac{1}{y}\right| = d(x,z) + d(z,y)$$

   again using the triangle inequality for real numbers.

3. Let X be any set and define $d(x,y) = \begin{cases} 0 & (x=y) \\ 1 & (x \neq y) \end{cases}$. Then d is a metric on X called the *discrete metric*.

   Again (M1) and (M2) are clear. For (M3), either $d(x,y) = 0$ in which case the inequality is obvious, or $d(x,y) = 1$ so $x \neq y$ with either $x \neq z$ so $d(x,z) = 1$ or $z \neq y$ so $d(z,y) = 1$.

4. $d\big((x_1,x_2),(y_1,y_2)\big) \;=\; \max\{|x_1-y_1|,|x_2-y_2|\}$ is a metric on $\mathbb{R}^2$.

   (M1) and (M2) are clear. For (M3), for all $(x_1,x_2),(y_1,y_2),(z_1,z_2) \in \mathbb{R}^2$,

$$\begin{aligned} d\big((x_1,x_2),(y_1,y_2)\big) &= \max\{|x_1-y_1|,|x_2-y_2|\} \\ &\le \max\{|x_1-z_1|+|z_1-y_1|,|x_2-z_2|+|z_2-y_2|\} \\ &\le \max\{|x_1-z_1|,|x_2-z_2|\}+\max\{|z_1-y_1|,|z_2-y_2|\} \\ &= d\big((x_1,x_2),(z_1,z_2)\big)+d\big((z_1,z_2),(y_1,y_2)\big) \end{aligned}$$

   where we have used the triangle inequality for real numbers.

5. $d\big((x_1,x_2),(y_1,y_2)\big) \;=\; +\sqrt{|x_1-y_1|^2+|x_2-y_2|^2}$ is the *Euclidean* or *usual* metric on $\mathbb{R}^2$.

   (M1) and (M2) are clear. The triangle inequality (M3) is an exercise using Cauchy's inequality, or see below.

6.2 Normed spaces

Many of the most important metric spaces are vector spaces (e.g. $\mathbb{R}^N$, spaces of continuous functions), and the natural distance is translation invariant, in the sense that, $d(x+z,y+z) = d(x,y)$ for all x,y,z. In other words the distance between two points x,y depends only on their vector difference. In this situation it is natural to define distances in terms of the 'length' of vectors.

Here, all vector spaces will be over $\mathbb{R}$, i.e. vectors will only be multiplied by real scalars. However, there is no problem in extending the work to vector spaces over $\mathbb{C}$.

Definition 6.2. (Normed space) *Let X be a vector space. A* norm *on X is a function $\|\cdot\| : X \to \mathbb{R}$ such that:*

(N1) $\quad \|x\| \;\ge\; 0$ *for all $x \in X$, with $\|x\| = 0$ if and only if $x = 0$,* (positivity)

(N2) $\quad \|\lambda x\| \;=\; |\lambda|\,\|x\|$ *for every $x \in X$ and every $\lambda \in \mathbb{R}$,* (scalar property)

(N3) $\quad \|x+y\| \;\le\; \|x\|+\|y\|$ *for all $x,y \in X$.* (triangle inequality)

The pair $(X,\|\cdot\|)$ is called a normed space.

We can immediately make two deductions from the triangle inequality for norms.

Lemma 6.3. *In a normed space $(X,\|\cdot\|)$*

(1) $$\Big\|\sum_{i=1}^{n} x_i\Big\| \;\le\; \sum_{i=1}^{n}\|x_i\| \quad \textit{for all } x_i \in X,$$

(2) $$\Big|\|x\|-\|y\|\Big| \;\le\; \|x-y\| \quad \textit{for all } x_i \in X \qquad \text{(reverse triangle inequality).}$$

Proof (1) follows by repeated application of the triangle inequality (N3).
For (2), let $x,y \in X$. Then

$$\|x\| = \|(x-y)+y\| \le \|x-y\|+\|y\|,$$

so

$$\|x\|-\|y\| \le \|x-y\|.$$

Interchanging the roles of x and y completes the proof. $\square$

We think of $\|x\|$ as the 'length' of the vector x in the vector space X, so that $\|x-y\|$ is a 'distance' between the two points x and y. In this way a normed space is automatically a metric space in a natural way, as the following lemma makes precise.

Lemma 6.4. *Let $(X, \|\cdot\|)$ be a normed space. Define $d\colon X \times X \to \mathbb{R}$ by*

$$d(x,y) = \|x-y\|.$$

Then d is a metric on X

Proof. The three properties required for d to be a metric follow easily from the three defining properties of the norm.

Property (M1) follows from (N1) by noting that $d(x,y) = \|x-y\| \geq 0$ with equality if and only if $\|x-y\| = 0$, i.e. $x = y$.

For (M2), $d(x,y) = \|x-y\| = \|-1(y-x)\| = |-1|\;\|(y-x)\| = d(y,x)$, using (N2).

For (M3), $d(x,y) = \|x-y\| = \|(x-z)+(z-y)\| \leq \|(x-z)\| + \|(z-y)\| = d(x,z)+d(z,y)$, using (N3). $\square$

Examples. Here follow some standard examples of normed spaces. In most cases (N1) and (N2) are easy to check, but (N3) will depend on establishing an appropriate inequality which may be more involved.

1. The familiar space $(\mathbb{R}, |\cdot|)$ is a normed space, with (N2) the absolute value multiplication property and (N3) the triangle inequality for real numbers.

2. Norms can be defined on the vector space $\mathbb{R}^N = \{x = (x_1, x_2, \ldots, x_N) : x_i \in \mathbb{R}\}$ of coordinate vectors in many ways, including

   (a) $\|x\| = \max_{1 \leq i \leq N} |x_i|$ (*maximum norm*),

   (b) $\|x\|_1 = \sum_{i=1}^N |x_i|$ (1-*norm* or *'New York' norm* (why?)),

   (c) $\|x\|_2 = \left(\sum_{i=1}^N |x_i|^2\right)^{1/2}$ (2-*norm*, *Euclidean norm* or *usual norm*),

   (d) $\|x\|_p = \left(\sum_{i=1}^N |x_i|^p\right)^{1/p}$ (for $p \geq 1$ only) (p-*norm*).

   Showing (a) and (b) are norms on $\mathbb{R}^N$ are simple exercises (using the triangle inequality for real numbers), (d) is rather harder and depends on Minkowski's inequality.

   We show here that (c) is a norm on $\mathbb{R}^N$:

   For (N1), note that $\|x\|_2 \geq 0$ and $\|x\|_2 = 0$ if and only if $x_i = 0$ for all i, i.e. if and only if $x = 0$.

   For (N2), $\|\lambda x\|_2 = \left(\sum_{i=1}^N |\lambda x_i|^2\right)^{1/2} = \left(\sum_{i=1}^N |\lambda|^2 |x_i|^2\right)^{1/2} = |\lambda|\left(\sum_{i=1}^N |x_i|^2\right)^{1/2} = |\lambda|\|x\|_2$.

   For (N3) we use Cauchy's inequality, that $\left(\sum_{i=1}^N |a_i b_i|\right)^2 \leq \left(\sum_{i=1}^N a_i^2\right)\left(\sum_{i=1}^N b_i^2\right)$ for real numbers a_i, b_i. [This can be proved in exactly the same way as the Cauchy-Schwartz inequality for integrals.] Then

$$\begin{aligned}
\|x+y\|_2^2 &= \textstyle\sum_{i=1}^N (x_i+y_i)^2 = \sum_{i=1}^N x_i^2 + 2\sum_{i=1}^N x_i y_i + \sum_{i=1}^N y_i^2 \\
&\leq \textstyle\sum_{i=1}^N x_i^2 + 2\left(\sum_{i=1}^N x_i^2\right)^{1/2}\left(\sum_{i=1}^N y_i^2\right)^{1/2} + \sum_{i=1}^N y_i^2 \\
&= \textstyle\left(\left(\sum_{i=1}^N x_i^2\right)^{1/2} + \left(\sum_{i=1}^N y_i^2\right)^{1/2}\right)^2 = \left(\|x\|_2 + \|y\|_2\right)^2.
\end{aligned}$$

3. The set $C[0,1]$ of real-valued continuous functions on $[0,1]$ forms a vector space under the obvious operations

$$(f+g)(x) = f(x)+g(x) \quad \text{and} \quad (\lambda f)(x) = \lambda f(x) \quad (x \in [0,1],\ \lambda \in \mathbb{R}).$$

Possible norms on $C[0,1]$ include:

(a) $\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|$ is known as the *supremum norm* or *uniform norm*, since, as we will see, it is closely related to uniform convergence.

For (N1), note that $\|f\|_\infty \geq 0$ and $\|f\|_\infty = 0$ if and only if $\sup_{x\in[0,1]} |f(x)| = 0$, i.e. if and only if $f(x) = 0$ for all $x \in [0,1]$.
For (N2), $\|\lambda f\|_\infty = \sup_{x\in[0,1]} |\lambda f(x)| = \sup_{x\in[0,1]} |\lambda||f(x)| = |\lambda| \sup_{x\in[0,1]} |f(x)| = |\lambda|\|x\|_\infty$.
To verify the triangle inequality (N3),

$$\begin{aligned}\|f+g\|_\infty &= \sup_{x\in[0,1]} |f(x)+g(x)| \leq \sup_{x\in[0,1]}(|f(x)|+|g(x)|)\\ &\leq \sup_{x\in[0,1]} |f(x)| + \sup_{y\in[0,1]} |g(y)| = \|f\|_\infty + \|g\|_\infty.\end{aligned}$$

(b) $\|f\|_1 = \int_0^1 |f(x)|dx$ is known as the 1*-norm*. (Exercise: show it is a norm.) Note that a function may have a large $\|\cdot\|_\infty$-norm but a small $\|\cdot\|_1$-norm.

(c) $\|f\|_p = \Big(\int_0^1 |f(x)|^p dx\Big)^{1/p}$ is the *p-norm*, and is a norm for $p \geq 1$. The triangle inequality requires an integral version of Minkowski's inequality.

4. The space $C^1[0,1]$ of functions on $[0,1]$ with continuous derivative becomes a normed space taking

$$\|f\| = \sup_{x\in[0,1]} |f(x)| + \sup_{x\in[0,1]} |f'(x)|.$$

(Exercise).

5. (For those who are familiar with inner products.) If $\langle\ ,\ \rangle$ is an inner product on a vector space X then $\|x\| = |\langle x,x\rangle|^{1/2}$ defines a norm on X. The triangle inequality follows as in 2(c) above, using the Cauchy-Schwartz inequality in the form $|\langle x,y\rangle| \leq \|x\|\|y\|$.

6.3 Convergence in metric and normed spaces

Metric spaces (and thus normed spaces) have enough structure to be able to define notions such as convergence and continuity. Indeed, the definitions are virtually the same as those we know for sequences of real numbers and real functions. We just have to replace $|x_n - x|$ by $d(x_n,x)$, etc. in the definitions.

Definition 6.5. (Convergence) *In a metric space* (X,d)*, the sequence* $(x_n)_{n=1}^\infty$ converges *to* x*, written* $x_n \to x$ *or* $\lim x_n = x$*, if for all* $\varepsilon > 0$ *there exists* $N \in \mathbb{N}$ *such that* $d(x_n,x) < \varepsilon$ *for all* $n \geq N$.

Thus, in a normed space $(X,\|\cdot\|)$*,* $x_n \to x$ *if for all* $\varepsilon > 0$ *there exists* $N \in \mathbb{N}$ *such that* $\|x_n - x\| < \varepsilon$ *for all* $n \geq N$.

Things become easy if we note that the above definition is precisely the definition that the *sequence of real numbers* $\big(d(x_n,x)\big)_{n=1}^{\infty}$ converges to 0. Thus we may reformulate this definition.

Definition 6.6. (Equivalent definition of convergence) *In a metric space* (X,d)*, the sequence* $(x_n)_{n=1}^{\infty}$ converges *to* x *if* $d(x_n,x) \to 0$ *as* $n \to \infty$.

In a normed space $(X, \|\cdot\|)$ *this becomes* $\|x_n - x\| \to 0$ *as* $n \to \infty$.

Examples.

1. In $(\mathbb{R}^2, \|\cdot\|_2)$, the sequence $\big((\frac{n+1}{n}, \frac{2n}{n+1})\big) \to (1,2)$. To see this,
$$\big\|(\tfrac{n+1}{n}, \tfrac{2n}{n+1}) - (1,2)\big\|_2 = \sqrt{(\tfrac{n+1}{n}-1)^2 + (\tfrac{2n}{n+1}-2)^2} = \sqrt{\tfrac{1}{n^2} + \tfrac{4}{(n+1)^2}} \le \sqrt{\tfrac{5}{n^2}} = \tfrac{\sqrt{5}}{n} \to 0.$$

2. (Uniform convergence) Let $C[0,1]$ be the vector space of continuous functions on $[0,1]$, and let $\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|$ be the uniform norm on $C[0,1]$, with corresponding metric $d(f,g) = \sup\limits_{x\in[0,1]} |f(x) - g(x)|$. Then $f_n \to f$ in $\|\cdot\|_\infty$ if and only if $f_n \to f$ uniformly on $[0,1]$.

   To see this, note $f_n \to f$ in $\|\cdot\|_\infty$ if and only if $\sup_{x\in[0,1]} |f_n(x) - f(x)| = \|f_n - f\|_\infty \to 0$, that is if and only if given $\varepsilon > 0$ there is an N such that for all $n \ge N$, $\sup_{x\in[0,1]} |f_n(x) - f(x)| \le \varepsilon$, i.e. $|f_n(x) - f(x)| \le \varepsilon$ for all $x \in [0,1]$ and $n \ge N$, which is uniform convergence.

3. Let $f_n \in C[0,1]$ be $f_n(x) = \begin{cases} 2nx & (0 \le x \le 1/2n) \\ 2n(\frac{1}{n} - x) & (1/2n \le x \le 1/n) \\ 0 & (1/n \le x \le 1) \end{cases}$.

   Then $\|f_n - 0\|_\infty = \sup_{x\in[0,1]} |f_n(x) - 0| = 1$ for all n, but $\|f_n - 0\|_1 = \int_0^1 |f_n(x) - 0| dx = 1/(2n) \to 0$. Thus $f_n \to 0$ in $\|\cdot\|_1$ but $f_n \not\to 0$ in $\|\cdot\|_\infty$.

4. Let d be the discrete metric on some set X given by $d(x,y) = \begin{cases} 0 & (x = y) \\ 1 & (x \ne y) \end{cases}$.

   Then $x_n \to x$ iff $d(x_n,x) \to 0$ iff there is an N such that $x_n = x$ for all $n \ge N$.

There are many instances where two different metrics (or norms) on the same space are closely enough related to yield the same convergent sequences, as well as sharing other properties. This motivates the definition of equivalent metrics (or norms).

Definition 6.7. (Equivalence of metrics or norms) *We say that two metrics* d *and* d' *on the same set* X *are* equivalent *if there are positive constants* $0 < a \le b$ *such that*

$$a d(x,y) \le d'(x,y) \le b d(x,y) \quad \textit{for all } x,y \in X. \tag{6.10}$$

Similarly, two norms $\|\cdot\|$ *and* $\|\cdot\|'$ *on the same vector space* X *are* equivalent *if*

$$a\|x\| \le \|x\|' \le b\|x\| \quad \textit{for all } x \in X. \tag{6.11}$$

Note that given equivalent norms satisfying (6.11), the associated metrics given by $d(x,y) = \|x-y\|$ and $d'(x,y) = \|x-y\|'$ automatically satisfy (6.10).

Lemma 6.8. *Let* d *and* d' *be equivalent metrics on* X*. If* $x_n, x \in X$ *then* $x_n \to x$ *in* d *if and only if* $x_n \to x$ *in* d'*. Similarly for equivalent norms.*

Proof Using (6.10), if $x_n \to x$ in d then $d'(x_n,x) \leq bd(x_n,x) \to 0$ so $x_n \to x$ in d'. Conversely, if $x_n \to x$ in d' then $d(x_n,x) \leq \frac{1}{a}d'(x_n,x) \to 0$ so $x_n \to x$ in d. □

Corollary 6.9. *The norms $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ on $\mathbb{R}^N$ are all equivalent.*

Proof (for $\mathbb{R}^2$). Note that for $(x_1,x_2) \in \mathbb{R}^2$,

$$\frac{1}{2}\big(|x_1|+|x_2|\big) \leq \max\big\{|x_1|,|x_2|\big\} \leq \big(|x_1|^2+|x_2|^2\big)^{1/2} \leq \big(|x_1|+|x_2|\big),$$

that is

$$\frac{1}{2}\big\|(x_1,x_2)\big\|_1 \leq \big\|(x_1,x_2)\big\|_\infty \leq \big\|(x_1,x_2)\big\|_2 \leq \big\|(x_1,x_2)\big\|_1,$$

i.e. the norms are equivalent. □

It follows that in Example 1 above, we showed that the sequence considered converges in $\|\cdot\|_2$, so it also converges in the equivalent norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$

In fact it can be shown that any two norms on $\mathbb{R}^N$ are equivalent.

Note that we can easily define continuity of a function between metric spaces, either with an 'ε-δ' definition or using sequences, by direct analogue of continuity of real-valued functions on $\mathbb{R}$. That is, if (X,d) and (Y,d') are metric spaces and $f : X \to Y$, then f is *continuous at* $c \in X$ if for every sequence $(x_n)_n$ in X with $x_n \to c$ in d, we have $f(x_n) \to f(c)$ in d'. We do not consider this further here, but with this definition continuity of functions between metric spaces can be developed in a similar way to that for functions on the real numbers.

6.4 Completeness

Recall the General Principle of Convergence: every Cauchy sequence of real numbers converges. Such a statement often holds in other metric or normed spaces; such spaces are called complete. Anything that guarantees that a sequence converges without the need to find the limit is extremely useful.

Definition 6.10. *(Cauchy Sequence) A sequence (x_n) in a metric space (X,d) is called a* Cauchy sequence *(or just* Cauchy*) if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n,x_m) < \varepsilon$ for all $n,m \geq N$.*

Thus (x_n) is Cauchy in a normed space $(X,\|\cdot\|)$ if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $\|x_n - x_m\| < \varepsilon$ for all $n,m \geq N$.

Example.

1. Let $f_n(x) = \sin(x+1/n)$ in $\big(C[0,1],\|\cdot\|_1\big)$.Then

$$\|f_n - f_m\|_1 = \int_0^1 \big|\sin(x+1/n)-\sin(x+1/m)\big|dx \leq \int_0^1 \big|1/n-1/m\big|dx \leq \frac{1}{n}+\frac{1}{m} < \varepsilon$$

if $n,m \geq 2/\varepsilon =$ 'N'. Thus (f_n) is Cauchy.

Proposition 6.11. *In a metric or normed space, every convergent sequence is a Cauchy sequence.*

Proof. Suppose $x_n \to x$ in (X,d). Given $\varepsilon > 0$ we may find $N \in \mathbb{N}$ such that $d(x_n,x) < \varepsilon/2$ if $n \geq N$. Then if $n,m \geq N$,

$$d(x_n,x_m) \leq d(x_n,x) + d(x,x_m) < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

using the triangle inequality, so (x_n) is Cauchy. $\square$

The converse of this proposition is in general *not* true. The Cauchy sequences may be thought of sequences that are 'trying to converge' and a space is complete if there is somewhere for them to converge to.

Definition 6.12. (Complete metric or normed space) *A metric space* (X,d) *or normed space* $(X,\|\cdot\|)$ *is called* complete *if every Cauchy sequence in the space converges to some point in the space.* [*A complete normed space is called a* Banach space.]

Of course, we already know that the real numbers are complete under the usual distance.

Proposition 6.13. (General Principle of Convergence) *Every Cauchy sequence of real numbers converges, i.e. the space* $(\mathbb{R},|\cdot|)$ *is complete.*

It is useful to have the completeness property in a normed space since we can assert that limits of sequences exist without explicitly having to find the limit.

Many important normed spaces are complete and we will give some of these below. Indeed the completeness of many spaces follows without too much difficulty from the completeness of the real numbers.

Important Examples 6.11.

1. *Let* $[a,b] \subset \mathbb{R}$ *be a closed interval. Then* $([a,b],|\cdot|)$ *is complete.*

   Proof: Note that if (x_n) is a Cauchy sequence in $([a,b],|\cdot|)$ then it is certainly a Cauchy sequence of real numbers, so $x_n \to x$ for some $x \in \mathbb{R}$. Then $x - a = \lim_{n\to\infty}(x_n - a) \geq 0$ and $b - x = \lim_{n\to\infty}(b - x_n) \geq 0$, so $x \in [a,b]$. $\square$

2. *The spaces* $(\mathbb{R}^N,\|\cdot\|_2)$, $(\mathbb{R}^N,\|\cdot\|_1)$, *and* $(\mathbb{R}^N,\|\cdot\|_\infty)$ *are all complete.*

   Proof . We first consider $(\mathbb{R}^N,\|\cdot\|_1)$; to keep coordinate notation reasonably simple we just prove completeness of $(\mathbb{R}^2,\|\cdot\|_1)$, though the calculation is virtually the same for larger N. Let $(x_n) = \big((x_n^1,x_n^2)\big)$ be a Cauchy sequence in $\|\cdot\|_1$. Thus given $\varepsilon > 0$, there is an N such that if $n,m \geq N$ then

   $$\varepsilon > \|x_n - x_m\|_1 = \big\|(x_n^1,x_n^2) - (x_m^1,x_m^2)\big\|_1 = \big\|(x_n^1 - x_m^1, x_n^2 - x_m^2)\big\|_1 = \textstyle\sum_{i=1}^2 \big|x_n^i - x_m^i\big|.$$

   In particular, $\varepsilon > |x_n^i - x_m^i\big|$ for each $i = 1,2$, i.e. $(x_n^i)_{n=1}^\infty$ is a Cauchy sequence of real numbers, so by the GPC, $x_n^i \to x^i$ for some $x^i \in \mathbb{R}$ for $i = 1,2$. Let $x = (x^1,x^2)$. Then

   $$\|x_n - x\|_1 \ = \ \textstyle\sum_{i=1}^2 |x_n^i - x^i| \to 0,$$

   i.e. $x_n \to x$ in $\|\cdot\|_1$. Thus $(\mathbb{R}^2,\|\cdot\|_1)$ is complete.

   We noted in Corollary (6.9) that the norms $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ are equivalent. It follows easily from the definition of equivalence that if a space is complete under one norm then it is complete under any equivalent norm, so $\mathbb{R}^N$ is also complete under $\|\cdot\|_2$ and $\|\cdot\|_\infty$. [In fact $\mathbb{R}^N$ is complete under any norm that might be defined on it.] $\square$

3. *The space* $(C[0,1], \|\cdot\|_\infty)$ *is complete.*

   [Recall that $C[0,1]$ is the space of continuous functions on $[0,1]$ and $\|f(x)\| = \sup_{y\in[0,1]} |f(y)|$; of course $[0,1]$ can be replaced by any other closed interval $[a,b]$.]

   Proof: Suppose $(f_n)_n$ is a Cauchy sequence in $(C[0,1], \|\cdot\|_\infty)$. For each $x \in [0,1]$,

   $$|f_n(x) - f_m(x)| \le \sup_{y\in[0,1]} |f_n(y) - f_m(y)| = \|f_n - f_m\|_\infty,$$

   so given $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that for all $n, m \ge N$,

   $$|f_n(x) - f_m(x)| < \varepsilon \qquad \text{for every } x \in [0,1]. \tag{6.12}$$

   Thus for each x, the sequence $\big(f_n(x)\big)_n$ is a Cauchy sequence of real numbers, so by the completeness of $\mathbb{R}$ (i.e. the GPC) it converges to a limit which we call $f(x)$.

   Letting $m \to \infty$ in (6.12) it follows that if $n \ge N$ then $|f_n(x) - f(x)| \le \varepsilon$ for every $x \in [0,1]$. Thus $f_n \to f$ uniformly on $[0,1]$. As the uniform limit of a sequence of continuous functions is continuous (Theorem 5.3) $f \in C[0,1]$, and also $|f_n(x) - f(x)|_\infty \le \varepsilon$ for all x if $n \ge N$, so $f_n \to f$ in $\|\cdot\|_\infty$. $\square$

Here is an example of how we can use completeness to show that a sequence of functions converges where it would be impossible to find an explicit function that is the limit.

Example. *Define a sequence of continuous functions* $f_n : [0,1] \to \mathbb{R}$ *by*

$$f_0(x) = x^3, \qquad f_n(x) = \frac{x}{2}\sin\big(f_{n-1}(x)\big) + \frac{1}{2}\cos x \quad (n = 1, 2, \ldots). \tag{6.13}$$

Then f_n *converges in* $\|\cdot\|_\infty$.

Check. For positive integers n, m, (6.13) gives

$$|f_n(x) - f_m(x)| = \frac{x}{2}\left|\sin\big(f_{n-1}(x)\big) - \sin\big(f_{m-1}(x)\big)\right| \le \frac{x}{2}\left|f_{n-1}(x) - f_{m-1}(x)\right|$$

so taking suprema,

$$\|f_n(x) - f_m(x)\|_\infty \ \le\ \frac{1}{2}\ \|f_{n-1}(x) - f_{m-1}(x)\|_\infty.$$

We can iterate this m times to get

$$\|f_n(x) - f_m(x)\|_\infty \ \le\ \frac{1}{2^m}\ \|f_{n-m}(x) - f_0(x)\|_\infty \ \le\ \frac{1}{2^m},$$

where the final inequality follows since $0 \le f_n(x) \le 1$ for all n and all $x \in [0,1]$ (by a simple induction using (6.13)).Thus, given $\varepsilon > 0$, choose N such that $1/2^N < \varepsilon$. Then if $n, m \ge N$, $\|f_n(x) - f_m(x)\|_\infty \ \le 1/2^N < \varepsilon$. Hence f_n is a Cauchy sequence in the complete normed space $(C[0,1], \|\cdot\|_\infty)$, so f_n converges in the norm $\|\cdot\|_\infty$ to some $f \in C[0,1]$. $\square$

6.5 The contraction mapping theorem

This final section concerns a very important and useful theorem valid in any complete metric or normed space that can be used to show that a unique solution exists to a wide range of problems, including simple equations, systems of equations, matrix equations, differential equations, etc, etc. It also provides an algorithm for finding the solution.

We will work with a function f that contracts distances on a complete metric space (X,d).

Definition 6.14. *Let (X,d) be a metric space. A function $f : X \to X$ is a* contraction *if there is a constant $0 < c < 1$ such that*

$$d\big(f(x), f(y)\big) \le c\, d(x,y) \qquad \text{for all } x,y \in X. \tag{6.14}$$

Given a function $f : X \to X$ we write $f^n : X \to X$ for the nth iterate of f, defined by $f^n(x) = f\big(f\big(f(\cdots f(x)\cdots)\big)\big)$ where f is applied n times.

We will prove the contraction mapping theorem, also known as Banach's fixed point theorem.

Theorem 6.15. (Contraction mapping theorem) *Let f be a contraction on a complete metric space (X,d). Then f has a unique fixed point x_0, i.e. there exists exactly one $x_0 \in X$ such that $f(x_0) = x_0$.*

Furthermore, for every $x \in X$, we have $d\big(f^n(x), x_0\big) \le c^n d\big(x, x_0\big) \to 0$ as $n \to \infty$.

Proof. Take any $x \in X$. We show that the sequence $(f^n(x))_{n=1}^{\infty}$ is Cauchy. First, using (6.14) repeatedly, for all $n \in \mathbb{N}$,

$$\begin{aligned} d\big(f^n(x), f^{n-1}(x)\big) &\le c d\big(f^{n-1}(x), f^{n-2}(x)\big) \le c^2 d\big(f^{n-2}(x), f^{n-3}(x)\big) \\ &\le \cdots \le c^{n-2} d\big(f^2(x), f^1(x)\big) \le c^{n-1} d\big(f^1(x), x\big). \end{aligned} \tag{6.15}$$

Now let $n > m$ be integers. Applying the triangle inequality repeatedly

$$\begin{aligned} d\big(f^n(x), f^m(x)\big) &\le d\big(f^n(x), f^{n-1}(x)\big) + d\big(f^{n-1}(x), f^{n-2}(x)\big) + \cdots + d\big(f^{m+1}(x), f^m(x)\big) \\ &\le \big[c^{n-1} + c^{n-2} + \cdots + c^m\big] d\big(f(x), x\big) \\ &< c^m \frac{d(f(x),x)}{1-c} \end{aligned}$$

using (6.15) and summing the geometric series. Given $\varepsilon > 0$, choose N large enough so that $c^N d(f(x),x)/(1-c) < \varepsilon$. Then $d\big(f^n(x), f^m(x)\big) < \varepsilon$ if $n,m \ge N$. Thus $(f^n(x))_{n=1}^{\infty}$ is a Cauchy sequence in the complete metric space (X,d) so $f^n(x) \to x_0$ for some $x_0 \in X$.

Using the triangle inequality and (6.14),

$$d(f(x_0), x_0) \le d(f(x_0), f^n(x)) + d(f^n(x), x_0) \le c d(x_0, f^{n-1}(x)) + d(f^n(x), x_0) \to 0$$

as $n \to \infty$. It follows that $d(f(x_0), x_0) = 0$ and so $f(x_0) = x_0$, that is x_0 is a fixed point of f.

Suppose that x_1 is also a fixed point. Then by (6.14)

$$d(x_0, x_1) = d(f(x_0), f(x_1)) \le c d(x_0, x_1),$$

so $d(x_0, x_1) = 0$ since $0 < c < 1$, so $x_0 = x_1$ giving uniqueness of the fixed point.

Finally, for every $x \in X$, applying (6.14) repeatedly,

$$d\big(f^n(x), x_0\big) = d\big(f^n(x), f^n(x_0)\big) \le c^n d(x, x_0) \to 0$$

as $n \to \infty$. $\square$

Example 6.16. (Roots of simple equations) *The equation* $x^3 + 3x - 1 = 0$ *has a unique real solution* $x \simeq 0.32219$.

Rearranging the equation gives $$x = \frac{1}{x^2+3}.$$

The mapping $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 1/(x^2+3)$ is a contraction on the complete space $(\mathbb{R}, |\cdot|)$, since

$$\begin{aligned} |f(x) - f(y)| &= \left| \frac{1}{x^2+3} - \frac{1}{y^2+3} \right| = \left| \frac{(y-x)(y+x)}{(x^2+3)(y^2+3)} \right| \\ &\leq |x-y| \left(\frac{|x|}{x^2+3} \frac{1}{y^2+3} + \frac{|y|}{y^2+3} \frac{1}{x^2+3} \right) \\ &\leq \tfrac{2}{3}|x-y|. \end{aligned}$$

By the CMT there is a unique $x \in \mathbb{R}$ such that $f(x) = x$, ti.e. the equation has a unique solution in $\mathbb{R}$. Moreover, for any $y \in \mathbb{R}$, the sequence $(f^n y)_n$ converges to the solution, so taking $y = 0.5$ we get the sequence of approximations

$$0.5, \quad 0.308, \quad 0.323, \quad 0.3221, \quad 0.32219, \ \ldots. \quad \square$$

Example 6.17. (Simultaneous nonlinear equations) *The simultaneous equations*

$$4x - \sin y = 5, \qquad \cos^2 x + 4y = 3,$$

have a unique solution $(x,y) \in \mathbb{R}^2$.

We can rearrange these equations as

$$\tfrac{1}{5}x + \tfrac{1}{5}\sin y + 1 = x, \quad -\tfrac{1}{6}\cos^2 x + \tfrac{1}{3}y + \tfrac{1}{2} = y. \tag{6.16}$$

Recalling that $\|(x,y)\|_\infty = \max\{|x|,|y|\}$, define f on $(\mathbb{R}^2, \|\cdot\|_\infty)$ by

$$f(x,y) = \big(\tfrac{1}{5}x + \tfrac{1}{5}\sin y + 1, -\tfrac{1}{6}\cos^2 x + \tfrac{1}{3}y + \tfrac{1}{2}\big).$$

Then

$$f(x,y) - f(x',y') = \big(\tfrac{1}{5}(x-x') + \tfrac{1}{5}(\sin y - \sin y'), -\tfrac{1}{6}(\cos^2 x - \cos^2 x') + \tfrac{1}{3}(y-y')\big),$$

so a simple estimate gives

$$\|f(x,y) - f(x',y')\|_\infty \leq \max\{\tfrac{2}{5}, \tfrac{2}{3}\} \max\{|x-x'|, |y-y'|\} = \tfrac{2}{3}\|(x,y) - (x',y')\|_\infty.$$

Hence f is a contraction on the complete normed space $(\mathbb{R}^2, \|\cdot\|_\infty)$, so by the CMT there is a unique $(x,y) \in \mathbb{R}^2$ such that $f(x,y) = (x,y)$, that is a unique solution to (6.16).

As before, $f^n(x_1,y_1)$ gives a sequence of increasingly good approximations to this solution for any initial (x_1,y_1) $\square$

We next turn to the question of whether a given differential equation has a solution and if so whether it is unique. This does not always happen even for simple DEs, for example,

$$\frac{dy}{dx} = 2|y|^{1/2} \qquad \text{with } y(0) = 0$$

has $y = 0$ and $y = x^2$ both as solutions, in fact, it has infinitely many solutions. However, the CMT may used to show that that a wide variety of differential equations have unique solutions.

Example 6.18. (Differential equations) *The differential equation*

$$\frac{df}{dx} = \tfrac{1}{4}\cos\big(f(x)\big) + x^3 \qquad \textit{with } f(0) = 0$$

has a unique solution for $0 \le x \le 1$.

This differential equation is equivalent to the integral equation

$$f(t) = \tfrac{1}{4}\int_0^x \cos\big(f(u)\big)du + \tfrac{1}{4}x^4, \tag{6.17}$$

which incorporates the initial conditions of the DE. Define a mapping ϕ on $(C[0,1], \|\cdot\|_\infty)$ by

$$\big(\phi(f)\big)(x) = \tfrac{1}{4}\int_0^x \cos\big(f(u)\big)du + \tfrac{1}{4}x^4.$$

Then

$$\begin{aligned}
\big|\big(\phi(f)\big)(x) - \big(\phi(g)\big)(x)\big| &= \Big|\tfrac{1}{4}\int_0^x \Big(\cos\big(f(u)\big) - \cos\big(g(u)\big)\Big)du\Big| \\
&\le \tfrac{1}{4}\int_0^x \big|\cos\big(f(u)\big) - \cos\big(g(u)\big)\big|du \\
&\le \tfrac{1}{4}\int_0^x \big|f(u) - g(u)\big|du \\
&\le \tfrac{1}{4}x\,\|f-g\|_\infty \;\le\; \tfrac{1}{4}\|f-g\|_\infty \quad \text{ since } x \in [0,1].
\end{aligned}$$

In particular, $\|\phi(f) - \phi(g)\|_\infty \le \frac{1}{4}\|f-g\|_\infty$. Hence ϕ is a contraction on the complete normed space $(C[0,1], \|\cdot\|_\infty)$. By the CMT there is a unique fixed point ϕ, i.e a unique function that satisifies $\phi(f) = f$, that is the solution to the integral equation (6.17) and thus to the differential equation. We can use iteration to find a sequence of approximations $\phi^n(g)$ to this solution f starting with any initial $g \in C[0,1]$ and this can be done numerically. $\square$