AMSI 2013: MEASURE THEORY
Handout 0
Background on Set Theory and Real Analysis

Marty Ross
martinirossi@gmail.com

January 7, 2013

We give here a very brief survey of set theory and real analysis. We've made no attempt at absolute rigor or logical minimalism. For more details see, for example, *Real Analysis* by Royden (Prentice Hall, 3rd edition, 1988) or *Foundations of Real and Abstract Analysis* by Bridges (Springer-Verlag, 2005). For a more thorough treatment of set theory see, for example, *Topology* by Munkres (Prentice Hall, 2000), *Naive Set Theory* by Halmos (Springer-Verlag, 2nd edition, 1998) or *Basic Set Theory* by Levy (Dover, 2002).

1 SET THEORY

We take as given the notion of a **set** as a collection of objects. These objects, which can be numbers, functions, sets, animals, politicians, whatever, are the **elements** of the set. (We'll sometimes talk of "collections" of sets or "classes" of sets, not with any subtle intent, but simply to avoid the clumsy expression "set of sets"). If x is an element of the set A, we write

$$x \in A\,.$$

We also have the notion of one set being a **subset** or **superset** of another:

$$\begin{cases} A \subseteq B & \text{if} \quad x \in A \implies x \in B\,, \\ A \supseteq B & \text{if} \quad x \in B \implies x \in A\,. \end{cases}$$

A set is determined entirely by its elements,

$$(A \subseteq B \text{ and } A \supseteq B) \quad \iff \quad A = B\,.$$

The **empty set** $\emptyset$ is the (unique) set with no elements. We have the following familiar sets of numbers (discussed briefly in §2 below).

$$\begin{cases} \text{Natural Numbers} & \mathbb{N} = \{1, 2, 3, \dots\} \\ \text{Integers} & \mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\} \\ \text{Rational Numbers} & \mathbb{Q} = \left\{ \dfrac{p}{q} \colon p, q \in \mathbb{Z}, q \neq 0 \right\} \\ \text{Real Numbers} & \mathbb{R} = \{???\} = \overline{\mathbb{Q}} \quad \text{(see §2.1 below)} \\ \text{Complex Numbers} & \mathbb{C} = \{x + yi \colon x, y \in \mathbb{R}\} \end{cases}$$

We construct new sets from old in the standard algebraic manner:

$$\begin{cases} \text{Union} & A \cup B = \{x \colon x \in A \text{ or } x \in B, \text{ or both}\} \\ \text{Intersection} & A \cap B = \{x \colon x \in A \text{ and } x \in B\} \\ \text{Relative Complement} & A \sim B = \{x \colon x \in A \text{ and } x \notin B\} \end{cases}$$

If the context is clear, we'll use the abbeviation $\sim B$ for $A \sim B$. For example, we'll write $\sim \mathbb{Q}$ for the set $\mathbb{R} \sim \mathbb{Q}$ of irrational numbers.

We have the standard algebraic rules:

$$\begin{cases} A \cup B = B \cup A \\ A \cap B = B \cap A \\ (A \cup B) \cup C = A \cup (B \cup C) & \text{(so we can unambiguously write } A \cup B \cup C) \\ (A \cap B) \cap C = A \cap (B \cap C) & \text{(so we can unambiguously write } A \cap B \cap C) \\ \left.\begin{array}{l} A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \end{array}\right\} & \text{(distributive laws)} \\ \sim\sim A = A \\ \left.\begin{array}{l} \sim(A \cap B) = \sim A \cup \sim B \\ \sim(A \cup B) = \sim A \cap \sim B \end{array}\right\} & \text{(De Morgan's laws)} \end{cases}$$

The **direct product** of two sets A and B is defined by

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

Here (a, b) is an **ordered pair**, which we can take as a given notion, but can be formally defined as $\{a, \{a, b\}\}$. As a special case,

$$A^2 = A \times A = \{(a, b) \colon a \in A, b \in A\}.$$

Identifying $(a,(b,c)) \approx ((a,b),c) \approx (a,b,c)$, we have

$$A \times B \times C = (A \times B) \times C = A \times (B \times C) = \{(a,b,c) \colon a \in A, b \in B, c \in C\},$$

and similarly for further products.

The **power set** of a set X is the set of all subsets of X:

$$\wp(X) = \{A \colon A \subseteq X\}.$$

Given a **function** $f \colon X \to Y$ we define

$$\operatorname{graph}(f) = \{(x, f(x)) \colon x \in X\}.$$

(More precisely, we define a function to be such a collection of ordered pairs). Also

$$\begin{cases} \operatorname{domain}(f) = X, \\ \operatorname{range}(f) = f(X) = \{f(x) \colon x \in X\}. \end{cases}$$

If $A \subseteq X$ then

$$f(A) = \{f(x) \colon x \in A\} = \text{the } \textbf{image} \text{ of } A \text{ under } f.$$

If $B \subseteq Y$ then

$$f^{-1}(B) = \{x \colon f(x) \in B\} = \text{the } \textbf{preimage} \text{ or } \textbf{inverse image} \text{ of } B \text{ under } f.$$

The function f is **surjective** if $f(X) = Y$, and f is **injective** if

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

If f is surjective and injective then f is **bijective**. In this case, f has an **inverse function** $f^{-1} \colon Y \to X$, for which

$$x = f^{-1}(y) \iff y = f(x).$$

Note that if f is a bijection then $f^{-1}(B)$ is harmlessly ambiguous: the image of B under f^{-1} is identical to the preimage of B under f.

Note that

$$\begin{cases} f(A \cup B) = f(A) \cup f(B), \\ f(A \cap B) \subseteq f(A) \cap f(B). \end{cases}$$

The preimage is better behaved:

$$\begin{cases} f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B), \\ f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B), \\ f^{-1}(\sim A) = \sim f^{-1}(A). \end{cases}$$

We also have

$$\begin{cases} f\left(f^{-1}(A)\right) \subseteq A \text{ with equality if } f \text{ is surjective} \\ f^{-1}\left(f(B)\right) \supseteq B \text{ with equality if } f \text{ is injective.} \end{cases}$$

A **finite sequence** $a_1, a_2, \ldots, a_n = \{a_j\}_{j=1}^{n}$ is a function with domain $\{1, 2, \ldots, n\}$. An **infinite sequence** $a_1, a_2, \cdots = \{a_j\}_{j=1}^{\infty}$ is a function with domain $\mathbb{N}$. Thus an infinite sequence $A_1, A_2, \ldots$ of subsets of X is a function $f : \mathbb{N} \to \wp(X)$. Given such a sequence, we can consider infinite (and finite) unions and intersections:

$$\begin{cases} \displaystyle\bigcup_{j=1}^{\infty} A_j = \bigcup_{j \in \mathbb{N}} A_j = \{x \colon x \in A_j \text{ for some } j \in \mathbb{N}\}, \\ \displaystyle\bigcap_{j=1}^{\infty} A_j = \bigcap_{j \in \mathbb{N}} A_j = \{x \colon x \in A_j \text{ for all } j \in \mathbb{N}\}. \end{cases}$$

Here we think of $\mathbb{N}$ as an **indexing set**: we have one set A_j for each $j \in \mathbb{N}$. Other indexing sets are possible – $\mathbb{R}$ is an important example – and we can consider the analogous unions and intersections. Also, if $\mathcal{I}$ is a collection of sets then $\mathcal{I}$ is itself an indexing set for that collection; we write

$$\begin{cases} \displaystyle\bigcup \mathcal{I} = \bigcup_{A \in \mathcal{I}} A = \{x \colon x \in A \text{ for some } A \in \mathcal{I}\}, \\ \displaystyle\bigcap \mathcal{I} = \bigcap_{A \in \mathcal{I}} A = \{x \colon x \in A \text{ for all } A \in \mathcal{I}\}. \end{cases}$$

The rules for images and preimages of functions extend to arbitrary unions and intersections.

For any indexed collection of sets, we have the generalisation of the distributive laws and De Morgan laws above:

$$\begin{cases} \displaystyle A \cap \left(\bigcup_{\alpha \in \mathcal{I}} B_\alpha\right) = \bigcup_{\alpha \in \mathcal{I}} (A \cap B_\alpha), \\ \displaystyle A \cup \left(\bigcap_{\alpha \in \mathcal{I}} B_\alpha\right) = \bigcap_{\alpha \in \mathcal{I}} (A \cup B_\alpha), \\ \displaystyle \sim\left(\bigcap_{\alpha \in \mathcal{I}} B_\alpha\right) = \bigcup_{\alpha \in \mathcal{I}} (\sim B_\alpha), \\ \displaystyle \sim\left(\bigcup_{\alpha \in \mathcal{I}} B_\alpha\right) = \bigcap_{\alpha \in \mathcal{I}} (\sim B_\alpha). \end{cases}$$

Next, we have the notion of a **relation** on a set X. Simple examples are "x is less than y"and "$x-y$ is a rational number". For a general relation R we write xRy if x holds in that relation to y.[1] R is an **equivalence relation** on X if:

$$\begin{cases} xRx & \text{(reflexivity)} \\ xRy \implies yRx & \text{(symmetry)} \\ xRy \text{ and } yRz \implies xRz & \text{(transitivity)} \end{cases}$$

For example, "$x - y$ is a rational number" is an equivalence relation on $\mathbb{R}$ (and on $\mathbb{Q}$).

If R is an equivalence relation on X and $x \in X$ then we define the **equivalence class** of x:

$$E_x = \{y \in X \colon xRy\} .$$

Then, for any x and y in X, either $E_x = E_y$ or $E_x \cap E_y = \emptyset$. Thus, every element of X is in exactly one equivalence class of R. We write

$$X/R = \{E_x \colon x \in X\}$$

for the set of equivalence classes of R.

A relation R on X is a **partial ordering** if

$$\begin{cases} xRy \text{ and } yRx \implies x = y & \text{(antisymmetry)} \\ xRy \text{ and } yRz \implies xRz & \text{(transitivity)} \end{cases}$$

For example, $<$ on $\mathbb{R}$ and $\subseteq$ on $\wp(X)$ are partial orderings (the latter being a reflexive partial ordering).[2]

A partial ordering is **linear** if, as well,

$$\text{for all } x, y \in X, \text{ either } x = y \text{ or } xRy \text{ or } yRx.$$

So, $<$ on $\mathbb{R}$ is linear, but $\subseteq$ on $\wp(X)$ is not linear (if X has more than one element).

A set A is **finite** if either $A = \emptyset$ or for some $n \in \mathbb{N}$ there is a bijection $f \colon \{1, \ldots, n\} \to A$. Subsets of finite sets are finite and the direct product of two finite sets is finite. A finite union of finite sets is finite.

A set A is **countable** if either $A = \emptyset$ or there is an injection $f : A \to \mathbb{N}$; since f need not be surjective, this includes the possibility that A is finite. Subsets of countable sets are countable, and the direct product of two countable sets is countable; moreover, if we assume

[1]Formally, a relation R is a collection of ordered pairs in $X \times X$, and we write xRy if $(x, y) \in R$.

[2]Some texts require that a partial ordering be reflexive; thus, they would consider $\leqslant$, but not $<$, to be a partial ordering on $\mathbb{R}$.

the *countable axiom of choice* (see below) then a countable union of countable sets is again countable. A set A is **countably infinite** if A is countable and not finite. If A is countably infinite then there is a bijection $g:\mathbb{N}\to A$.

A set A is **uncountable** if it is not countable. By the well-known Cantor diagonalization argument, $\wp(\mathbb{N})$ is uncountable. More generally, for any set A, there is no surjection $A\to\wp(A)$, and consequently no injection $\wp(A)\to A$.

For sets A and B we write $|A| = |B|$ if there is a bijection $A\to B$, in which case we say A and B have the **same cardinality**. If there is an injection $A\to B$ we write $|A| \leqslant |B|$, and if $|A| \leqslant |B|$ but it is not the case that $|A| = |B|$ then we write $|A| < |B|$; in the latter case we say A has **lower cardinality** than B. It is clear that $<$ is a partial ordering on sets. The *Schröder-Bernstein theorem* states that if $|A| \leqslant |B|$ and $|B| \leqslant |A|$ then $|A| = |B|$, from which it follows that $\leqslant$ is also a partial ordering on sets. A consequence of the axiom of choice (see below) is that $<$ and $\leqslant$ are in fact linear orderings on sets.

From the above, it is clear that for any set A we have $|A| < |\wp(A)|$. The *continuum hypothesis* is the proposition that no set A has cardinality between that of $\mathbb{N}$ and $\wp(\mathbb{N})$: that is, there is no set A for which $|\mathbb{N}| < |A| < |\wp(A)|$. The continuum hypothesis cannot be proved (or disproved) from the other standard axioms of set theory, and so must itself simply be taken as an axiom if desired. Similarly, we have

The Axiom of Choice. *Suppose $\{A_\alpha\}_{\alpha\in\mathcal{I}}$ is a collection of non-empty and pairwise disjoint sets (that is, $A_\alpha \cap A_\beta = \emptyset$ if $\alpha \neq \beta$). Then there is a set A which contains exactly one element from each A_α.*

The finite axiom of choice (i.e. the case where $\mathcal{I}$ is finite) is trivial. The countable axiom of choice is generally considered intuitive, but nonetheless is neither provable nor disprovable from the other standard axioms of set theory (with or without assuming the continuum hypothesis). Stared at for long enough, the uncountable axiom of choice can make one feel very queasy. See *Halmos*, *Munkres* and *Levy* for details.

An important consequence (in fact, equivalence) of the axiom of choice is:

Hausdorff Maximal Principle. *Suppose R is a partial ordering on a set X. Then there is a maximal linearly ordered $Y \subseteq X$. That is:*

(i) *Y is linearly ordered by R;*

(ii) *If $Y \subseteq Z \subseteq X$ and if Z is linearly ordered by R, then $Z = Y$.*

2 REAL ANALYSIS

2.1 The real number system

We take as given the sets $\mathbb{N}, \mathbb{Z}$ and $\mathbb{Q}$ of, respectively, **natural numbers, integers** and **rational numbers**; we merely note that $\mathbb{N}$ is effectively characterized by **mathematical induction**: if $A \subseteq \mathbb{N}$ and if

$$\begin{cases} 1 \in A \\ j \in A \implies j+1 \in A \end{cases}$$

then $A = \mathbb{N}$. The constructions of $\mathbb{Z}$ and $\mathbb{Q}$ from $\mathbb{N}$ are then quite natural and easy.[3]

The big step of analysis is to fill in the gaps, to extend $\mathbb{Q}$ to the set $\mathbb{R}$ of **real numbers**. This can be done either *constructively* (where we somehow *make* the real numbers out of the rational numbers), or *axiomatically* (where we simply list certain fundamental properties which we are willing to accept as true). Here we'll do neither: we'll be content to recall the signature properties of $\mathbb{R}$, without regard to whether they are assumed or proved.[4]

The quickest method of characterizing $\mathbb{R}$ is with the concept of **least upper bound**. Given a set $A \subseteq \mathbb{R}$, we say A is **bounded above** if there is an **upper bound** M for A: that is, for every a,

$$a \in A \implies a \leqslant M\,.$$

Then α is the **least upper bound** for A if:

$$\begin{cases} \alpha \text{ is an upper bound for } A; \\ \text{If } \beta \text{ is another upper bound for } A \text{ then } \alpha \leqslant \beta\,. \end{cases}$$

Least Upper Bound Property. *If $A \subseteq \mathbb{R}$ is non-empty, and if A is bounded above, then A has a least upper bound α.* In this case, we write

$$\alpha = \sup A\,.$$

Similarly, we can consider sets which are **bounded below**, and a non-empty set A which is bounded below will have a **greatest lower bound** $\beta = \inf A$.

[3] See, for example, Part 1 of *Calculus* by Spivak (Cambridge, 3rd edition, 2006).

[4] An axiomatic system for the real numbers consists of **algebraic** laws, such as the commutative law $a+b=b+a$ for addition, and laws of **ordering**, such as $a<b \Rightarrow a+c<b+c$. However, $\mathbb{Q}$ will also satisfy all such laws. What is needed is one last axiom, to capture the "no gaps" property of the real numbers. Any of the limit-type properties we now describe could be taken as this final axiom. See *Spivak* for details.

The least upper bound property leads easily to

Archimedean Property. $\mathbb{N}$ *is not bounded above.*

An alternative characterization of $\mathbb{R}$ can be made via infinite sequences. A real-valued sequence $\{a_j\}_{j=1}^{\infty}$ (or $\{a_j\}$ for short) is **increasing** if $a_{j+1} \geqslant a_j$ for all $j \in \mathbb{N}$. (Note that an "increasing" sequence need not be **strictly** increasing, where $a_{j+1} > a_j$). Analogously, we can define $\{a_j\}$ **decreasing**, and $\{a_j\}$ is **monotonic** if $\{a_j\}$ is either increasing or decreasing. The sequence $\{a_j\}$ **converges** to $a \in \mathbb{R}$ if, for every $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that $j \geqslant N \implies |a_j - a| < \epsilon$. In this case, we write $\lim\limits_{j\to\infty} a_j = a$ or $a_j \to a$. Note that if $a_j \to a$, and if for some $L \in \mathbb{R}$ each $a_j \leqslant L$, then also $a \leqslant L$. If a_n increases or decreases to a, we also write $a_n \nearrow a$ or $a_n \searrow a$, accordingly.

We now have

Monotonic Sequence Property. *If $\{a_j\}$ is a monotonic and bounded sequence of real numbers then $\{a_j\}$ converges to some $a \in \mathbb{R}$.*

We can also combine the previous concepts: if $\{a_j\}$ is bounded (above and below), but not necessarily monotonic, we can still define

$$\begin{cases} \limsup\limits_{j\to\infty} a_j = \lim\limits_{j\to\infty} \sup\limits_{k\geqslant j} a_k \,, \\ \\ \liminf\limits_{j\to\infty} a_j = \lim\limits_{j\to\infty} \inf\limits_{k\geqslant j} a_k \,. \end{cases}$$

Note that

$$\lim_{j\to\infty} a_j = a \quad \iff \quad \begin{cases} \limsup\limits_{j\to\infty} a_j = a \\ \\ \liminf\limits_{j\to\infty} a_j = a \,. \end{cases}$$

A similar characterization of $\mathbb{R}$ can be made in terms of the convergence of **infinite series**. An infinite series $\sum\limits_{j=1}^{\infty} a_j$ in $\mathbb{R}$ is defined in terms of the **sequence** $\{s_j\}$ **of partial sums**, where $s_j = \sum\limits_{i=1}^{j} a_i$. Then the series $\sum\limits_{j=1}^{\infty} a_j$ (or $\sum a_j$ for short) **converges** if the corresponding sequence $\{s_j\}$ converges.

Two important families of series are:

$$\begin{cases} \displaystyle\sum_{j=1}^{\infty} \frac{1}{j^p} \quad \text{converges iff } p > 1 & (\textbf{harmonic } \boldsymbol{p}\textbf{-series}), \\ \displaystyle\sum_{j=0}^{\infty} r^j \quad \text{converges iff } |r| < 1 & (\textbf{geometric series}). \end{cases}$$

Of course, in the case of a convergent geometric series, we know the sum:

$$\sum_{j=0}^{\infty} r^j = \frac{1}{1-r} \qquad |r| < 1.$$

A further important result is that for a convergent series of *nonnegative* real numbers, the sum is independent of rearrangement: if $j : \mathbb{N} \to \mathbb{N}$ is a bijection then

$$\sum_{j=1}^{\infty} a_j = \sum_{k=1}^{\infty} a_{j(k)} \qquad a_j \geqslant 0.$$

A series $\sum a_j$ in $\mathbb{R}$ is said to **converge absolutely** if the series $\sum |a_j|$ converges. The standard example $\sum \frac{(-1)^k}{k}$ demonstrates that a convergent series need not be absolutely convergent. However, we do have the converse:

Absolute Convergence Test. *Every absolutely convergent series in* $\mathbb{R}$ *converges.*

It follows that the sum of an absolutely convergent series is independent of rearrangement. In particular, if a doubly-indexed series $\displaystyle\sum_{j,k=1}^{\infty} a_{jk}$ is absolutely convergent then it can be summed in any order, and

$$\sum_{j,k=1}^{\infty} a_{jk} = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_{jk} = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} a_{jk}\,.$$

Similar to infinite series, we can definite an **infinite product** $\displaystyle\prod_{j=1}^{\infty} a_j$ of positive real numbers as the limit of the sequence $\left\{ \displaystyle\prod_{j=1}^{n} a_j \right\}$ of finite products. Then a product $\displaystyle\prod_{j=1}^{\infty} (1+b_j)$ converges iff $\displaystyle\sum_{j=1}^{\infty} b_j$ converges.

We can also characterize $\mathbb{R}$ in terms of **continuous functions**. A function $f : \mathbb{R} \to \mathbb{R}$ is continuous at a if, for every $\epsilon > 0$, there is a $\delta > 0$ such that $|x-a| < \delta \implies |f(x)-f(a)| < \epsilon$.

The function f is then continuous if f is continuous at a for each $a \in \mathbb{R}$. Then the no-gaps property of $\mathbb{R}$ is captured by:

Intermediate Value Theorem. *If $f : [a, b] \to \mathbb{R}$ is continuous, and if $f(a) < 0$ and $f(b) > 0$, then there is a real number $c \in (a, b)$ for which $f(c) = 0$.*

We also have the related concept of the **limit** $\lim\limits_{x \to a} f(x)$ of a function f at a, where $x = a$ is precluded from consideration. There is also the natural extension to one-sided limits, $\lim\limits_{x \to a^{\pm}} f(x)$. We then have the standard algebraic limit laws. The notion of a limit also leads to the **differential calculus**. We take as given the standard rules for differentiation: linearity, product and quotient rules, chain rule, and the inverse function theorem.

Further key properties of $\mathbb{R}$ are implied in the results below. As a final property here, we note that, while $\mathbb{Q}$ is countable, $\mathbb{R}$ is uncountable: there is no bijection between $\mathbb{N}$ and $\mathbb{R}$.

2.2 Euclidean Space

Given $\mathbb{R}$ and $m \in \mathbb{N}$, we define ***m*-dimensional Euclidean space**:

$$\mathbb{R}^m = \mathbb{R} \times \cdots \times \mathbb{R} = \{(x_1, \cdots, x_m) \colon x_j \in \mathbb{R} \text{ for } j = 1, \cdots, m\}\,.$$

Of course, choosing $m = 1$ includes $\mathbb{R}$ as a special case. We define the **inner product** $\langle \cdot, \cdot \rangle \colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ by

$$\langle a, b \rangle = \sum_{j=1}^{m} a_j b_j, \qquad a, b \in \mathbb{R}^m\,.$$

We have the associated **norm**,

$$\|a\| = \sqrt{\langle a, a \rangle}\,,$$

and we can then define

$$d(a, b) = \|a - b\|\,,$$

the **distance** between $a, b \in \mathbb{R}^m$. Then, as for $\mathbb{R}$, we can define the concepts of convergent sequences, and **Cauchy** sequences: a sequence $\{a_j\}$ in $\mathbb{R}^m$ is Cauchy if, for every $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that $j, k \geqslant N \implies \|a_j - a_k\| < \epsilon$. The simplest no-gaps property of Euclidean space is then:

Completeness of Euclidean Space. *Every Cauchy sequence in $\mathbb{R}^m$ converges.*

Infinite series in $\mathbb{R}^n$ can be defined just as in $\mathbb{R}$. The results for absolutely convergent series and rearrangement of series hold as before.

The concepts of continuity, limits and differentiation extend to $f:\mathbb{R}^m \to \mathbb{R}^p$, with limits and differentiability defined component by component of $\mathbb{R}^p$. For $f:\mathbb{R}^m \to \mathbb{R}$ we denote the i'th partial derivative of f at a by $D_i f(a)$, and we write $Df = (D_1 f, \ldots, D_m f)$. Then f is differentiable at a if

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - \langle Df(a), h\rangle}{\|h\|} = 0\,.$$

This gives a function $Df:\mathbb{R}^m \to \mathbb{R}^m$, and we then define

$$\begin{aligned} C^1(\mathbb{R}^m) &= \{f\colon \mathbb{R}^m \to \mathbb{R} \text{ such that } f \text{ is differentiable differentiable with } Df \text{ continuous}\} \\ &= \{f\colon \mathbb{R}^m \to \mathbb{R} \text{ such that the partial derivatives } D_i f \text{ are continuous, } i = 1\ldots m\}\ . \end{aligned}$$

The extension to a vector-valued $f:\mathbb{R}^m \to \mathbb{R}^p$ can be made component by component. Then, for $k \in \mathbb{N}$, we define the space $C^k(\mathbb{R}^m)$ of k-times differentiable functions on $\mathbb{R}^m$ with continuous kth derivative.

The theorems for limits of and differentiation of functions on $\mathbb{R}$ have their natural generalisations to functions on $\mathbb{R}^m$. We make special note of

Mean Value Theorem. *Suppose that $f:\mathbb{R}^m \to \mathbb{R}$ is differentiable. Then, for any $a, b \in \mathbb{R}^m$, there is a $c \in \mathbb{R}^m$ on the segment connecting a and b such that*

$$f(b) - f(a) = \langle Df(c), b-a\rangle\ .$$

2.3 Normed spaces and inner product spaces

A real vector space X is a **normed space** if there is a **norm** $\|\cdot\|:X \to \mathbb{R}$, with the following properties:

$$\begin{cases} \|x\| \geqslant 0, \text{ and } \|x\| = 0 \text{ iff } x = 0 \\ \|\alpha x\| = |\alpha|\cdot\|x\|, \text{ for all } x \in X \text{ and } \alpha \in \mathbb{R} \\ \|x+y\| \leqslant \|x\| + \|y\| & \text{(triangle inequality)} \end{cases}$$

Also, X is a (real) **inner product space** if there is an **inner product** $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$, such that

$$\begin{cases} \langle x, x \rangle \geqslant 0, \text{ and } \langle x, x \rangle = 0 \text{ iff } x = 0 \\ \langle x, y \rangle = \langle y, x \rangle \\ \langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \text{ for all } x, y, z \in X \text{ and } \alpha, \beta \in \mathbb{R} \end{cases}$$

The canonical example of an inner product space is $\mathbb{R}^m$. And, any inner product becomes a normed space by defining

$$\|x\| = \sqrt{\langle x, x \rangle}\,.$$

Similar definitions apply to complex vector spaces, with $\mathbb{C}$ replacing the role of $\mathbb{R}$.

In any inner product space we have the **Cauchy-Schwarz inequality**:

$$\langle x, y \rangle \leqslant \|x\| \cdot \|y\|\,.$$

For any set X we define

$$B(X) = \{\text{bounded functions } f : X \to \mathbb{R}\}\ .$$

Then, $B(X)$ is a normed space with respect to the **sup norm**:

$$\|f\| = \sup\{|f(x)| : x \in X\}\,.$$

2.4 Metric spaces

A **metric space** (X, d) is a set X together with a **distance function** $d : X \times X \to \mathbb{R}$ satisfying:

$$\begin{cases} d(x, y) \geqslant 0, \text{ and } d(x, y) = 0 \text{ iff } x = y \\ d(x, y) = d(y, x) \\ d(x, z) \leqslant d(x, y) + d(y, z). \qquad \text{(triangle inequality)} \end{cases}$$

We write X for (X, d) if the associated metric d is clear. Any normed space, and $\mathbb{R}^m$ in particular, is naturally a metric space by defining

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

A subset $A \subseteq X$ of a metric space X is a metric **subspace** with the **induced metric** d_A in the obvious manner. On any set X, we can define the **discrete metric** d_s:

$$d_s(x, y) = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y. \end{cases}$$

We can define convergent and Cauchy sequences in a metric space X, as in Euclidean space. The space X is said to be **complete** if every Cauchy sequence is convergent. Thus, $\mathbb{R}^m$ is complete; in contrast, the interval $(0,1)$ with the induced metric is not complete. Any discrete metric space is complete.

A complete normed space is called a **Banach space**, and a complete inner product space is called a **Hilbert space**; thus Euclidean space is both a Banach space and a Hilbert space. In a normed space, we also have the notion of infinite series being convergent; then a normed space X is a Banach space iff every absolutely convergent series is convergent. The results for rearranging absolutely convergent series also hold in a Banach space.

For any set X, the space $B(X)$ of bounded functions on X is a Banach space. If $f_n \to f$ in $B(X)$ we say that f_n **converges uniformly** to f. We also have the *Weierstrass M-test*: if $\|f_n\| \leqslant M_n$ with $\sum M_n$ convergent, then $\sum f_n$ converges, and converges uniformly to its limit.

Given a metric space X, $x \in X$ and $r \in \mathbb{R}$, we define, respectively, the **open ball** and **closed ball** of radius r about x:

$$\begin{cases} B_r(x) = \{y \in X : d(x,y) < r\}\,, \\ C_r(x) = \{y \in X : d(x,y) \leqslant r\}\,. \end{cases}$$

A set $U \subseteq X$ is **open** if for every $x \in U$ there is an $r > 0$ such that $B_r(x) \subseteq U$. A set $C \subseteq X$ is **closed** if its **complement** $X \sim C$ is open. Equivalently, C is closed if whenever a sequence $x_j \to x$ in X then $x \in C$. Note that open balls are open and closed balls are closed, and that X and $\emptyset$ are both open and closed. A closed subset of a complete metric space is complete. With respect to the discrete metric, all sets are both open and closed.

If $A \subseteq X$ is non-empty and $x \in X$, the **distance** from x to A is defined by

$$\operatorname{dist}(x, A) = \inf\{d(a,x) : a \in A\}\,.$$

Given $r \in \mathbb{R}$, $\{x \in X : \operatorname{dist}(x,A) < r\}$ is open and $\{x \in X : \operatorname{dist}(x,A) \leqslant r\}$ is closed. The **diameter** of a non-empty set A is defined by

$$\operatorname{diam}(A) = \sup\{d(a,b) : a, b \in A\}$$

if this sup exists, and $\operatorname{diam}(A) = \infty$ otherwise; A is **bounded** if $\operatorname{diam}(A) < \infty$. By convention, we define $\operatorname{diam}(\emptyset) = 0$.

The **interior** A° of $A \subseteq X$ is the union of all open sets contained in A, and the **closure** $\overline{A}$ of A is the intersection of all closed sets containing A. Thus

$$\sim A^\circ = \overline{\sim A}\,.$$

The **boundary** of $A \subseteq X$ is defined by

$$\partial A = \overline{A} \cap \overline{\sim A}\,.$$

Note that $\overline{A}$ is the union of A together with any limit $x \in X$ of a sequence $\{x_j\}$ in A which converges; so, $\overline{A}$ consists of all $x \in X$ that are a distance 0 from A. We say A is **dense** in X if $\overline{A} = X$. It follows easily from the properties of $\mathbb{R}$ that $\mathbb{Q}$ is dense in $\mathbb{R}$ and, more generally, $\mathbb{Q}^m$ is dense in $\mathbb{R}^m$.

A metric space X is **compact** if any of the following equivalent conditions hold:

- Every sequence in X has a convergent **subsequence**[5] (**Bolzano-Weierstrass**).
- X is complete and **totally bounded**.[6]
- From any collection $\{O_\alpha\}$ of open sets which covers X, there is a finite subcollection which also covers X (**Heine-Borel**).
- If $\{C_\alpha\}$ is a collection of closed sets in X such that any finite subcollection from $\{C_\alpha\}$ has non-empty intersection, then all of $\{C_\alpha\}$ has non-empty intersection.

Note, in particular, that a compact metric space is bounded. An important result is that a subset of $\mathbb{R}^m$ is compact iff it is closed and bounded; in particular, any closed and bounded interval $[a, b]$ in $\mathbb{R}$ is compact. If a set X is given the discrete metric, then X is compact iff X is finite.

A real-valued function $f : X \to \mathbb{R}$ on a metric space X is **continuous** if any of the following equivalent conditions hold:

- For any $x \in X$ and $\epsilon > 0$, there is a $\delta > 0$ such that $d(x, y) < \delta \Longrightarrow |f(x) - f(y)| < \epsilon$.
- Whenever $x_j \to x$ in X, then $f(x_j) \to f(x)$.
- For any open $U \subseteq \mathbb{R}$, the inverse image $f^{-1}(U)$ is open in X.
- For any closed $C \subseteq \mathbb{R}$, the inverse image $f^{-1}(C)$ is closed in X.

If $A \subseteq X$ then the function $x \mapsto \mathrm{dist}(x, A)$ is continuous. If X is a compact metric space and $f : X \to \mathbb{R}$ is continuous, then f is **uniformly continuous**: for any $\epsilon > 0$ there is a $\delta > 0$ such that $d(x, y) < \delta \Longrightarrow |f(x) - f(y)| < \epsilon$. The point is that ϵ can be chosen independent of x and y; the function $f(x) = x^2$ on $\mathbb{R}$ is a standard example showing continuity does not imply uniformly continuity.

[5] A subsequence $\{x_{j_k}\}_{k=1}^{\infty}$ of sequence $\{x_j\}_{j=1}^{\infty}$ is determined by a strictly increasing function (sequence) $k \mapsto j_k$ of natural numbers.

[6] X is totally bounded if, for any $\epsilon > 0$, X is contained in finitely many open balls of radius ϵ.

2.5 Topological spaces

A **topological space** $(X, \mathcal{T})$ is a set X together with a collection $\mathcal{T} = \{U_\alpha\}$ of subsets of X, designated as the **open sets**. This collection must satisfy:

$$\begin{cases} \emptyset \text{ and } X \text{ are open} \\ \text{If } A \text{ and } B \text{ are open then } A \cap B \text{ is open} \\ \text{The union of any collection of open sets is open} \end{cases}$$

We write X for the topological space if the topology $\mathcal{T}$ is clear. A **base** for the topology $\mathcal{T}$ is a subcollection $\mathcal{B}$ of open sets such that: for each $U \in \mathcal{T}$ and each $x \in U$ there is a $B \in \mathcal{B}$ such that $x \in B \subseteq U$. Any base completely determines the topology $\mathcal{T}$. A collection $\mathcal{B}$ of subsets of X will be a base for a (unique) topology on X if $\cup\mathcal{B} = X$ and if, for any $U, V \in \mathcal{B}$, there is a $W \in \mathcal{B}$ with $W \subseteq U \cap V$.

For $x \in X$, a **base at** x is a collection $\mathcal{B}_x$ of open sets such that: for each $U \in \mathcal{T}$ containing x there is a $B \in \mathcal{B}_x$ such that $x \in B \subseteq U$. A metric space X, with open sets as defined above, is a topological space; the open balls are a base for X, and the open balls centred at a given x are a base at x.

The **discrete topology** $\mathcal{T}_s = \wp(X)$ on X consists of all subsets of X; this is the topology derived from the discrete metric. At the other extreme, the **indiscrete topology** $\mathcal{T}_i = \{\emptyset, X\}$ consists solely of $\emptyset$ and X.

Closedness, compactness and continuity in a topological space can be defined as for metric spaces above, where we restrict to those characterizations made in terms of open sets or closed sets. In particular, if X and Y are topological spaces then $f : X \to Y$ is defined to be continuous if $f^{-1}(U)$ is open in X for every open $U \subseteq Y$; if X and Y are metric spaces, this is equivalent to $f(x_j) \to f(x)$ in Y whenever $x_j \to x$ in X, and is equivalent to the $\epsilon - \delta$ definition of continuity. If X is discrete then any $f : X \to Y$ is continuous; at the other extreme, if X is indiscrete, then f is continuous iff f is constant.

If $A \subseteq X$ is a subset of a topological space X, then A can be given the **subspace topology** by declaring

$$U \text{ is open in } A \quad \Longleftrightarrow \quad U = V \cap A \text{ for some open } V \subseteq X.$$

The corresponding property then holds for closed sets. If $A \subseteq X$ is open (closed) in X then $U \subseteq A$ is open (closed) in A iff U is open (closed) in X. If X is a metric space, then the induced metric d_A give the subspace topology on $A \subseteq X$.

We note the following elementary properties:

- A closed subset of a compact topological space is compact.
- If $f\!:\!X\!\to\!Y$ and $g\!:\!Y\!\to\!Z$ are continuous then $g\circ f\!:\!X\!\to\!Z$ is continuous.
- If $f\!:\!X\!\to\!Y$ is continuous then

$$\begin{cases} f\!:\!X\!\to\!B \text{ is continuous for any } B\supseteq f(X) \\ \text{the } \textbf{restriction } f_{|_A}\!:\!A\!\to\!Y \text{ of } f \text{ is continuous for any } A\subseteq X. \end{cases}$$

- If $A\subseteq X$ is closed, $f:A\to\mathbb{R}$ is continuous and $f=0$ on ∂A, then $g:X\to\mathbb{R}$ is continuous, where
$$g(x)=\begin{cases} f(x) & x\in A, \\ 0 & x\in\sim A. \end{cases}$$
- If $f\!:\!X\!\to\!Y$ is continuous and X is compact then $f(X)$ is compact.
- If $f:X\to\mathbb{R}$ is continuous and X is compact then f is bounded above (and below), and f achieves its maximum (and minimum) on X.
- If $\{f_j\!:\!X\!\to\!\mathbb{R}\}$ is a sequence of continuous functions converging uniformly to f, then f is continuous.

For a topological space X, define

$$C(X)=\{f\!: f \text{ is a continuous real-valued function on } X\}\,.$$

We then define $C_0(X)$ to be the set of functions in $C(X)$ with **compact support**.[7] A consequence of the above results is that $C(X)\cap B(X)$ and $C_0(X)\subseteq B(X)$ are Banach spaces. In the latter case, we can write $\|f\|=\max\{|f(x)|: x\in X\}$. Of course, if X is compact then $C_0(X)=C(X)$. A deep and important result is:

Stone-Weierstrass Theorem. *Suppose that $X\subseteq\mathbb{R}^m$ is compact, and let $\mathcal{P}(X)$ be the space of polynomials $p\!:\!\mathbb{R}^m\!\to\!\mathbb{R}$ restricted to X. Then $\overline{\mathcal{P}(X)}=C(X)$.*

A topological space X is **locally compact** if for every $x\in X$ there in an open set U containing x with $\overline{U}$ compact. Clearly, $\mathbb{R}^m$ is locally compact. If X is an infinite metric space then $C(X)$ is not locally compact; X itself is not locally compact if given the discrete metric.

[7] The support of a function $f\!:\!X\!\to\!\mathbb{R}$ is defined as $\mathrm{spt}(f)=\overline{f^{-1}(\{0\})}$.

If X and Y are topological spaces then the product $X \times Y$ can be given the **product topology**. A base for the product topology is the collection of all sets of the form $U \times V$, where U is open in X and V is open in Y. The product topology is the weakest topology (i.e. the topology including the fewest "open" sets) such that the **projections** $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are continuous. If X and Y are compact (locally compact) then $X \times Y$ is compact (locally compact).

A topological space X is **metrizable** if there is a metric d on X that gives rise to exactly the open sets of X; of course this will be the case when the topology arises from a metric in the first place. Not all topological spaces are metrizable; as a trivial example, if X contains at least two points, then the indiscrete topology on X is not metrizable (see below). For many more interesting examples, see *Munkres* or *Counterexamples in Topology* by Steen and Seebach (Dover, 1995).

Even if a topological space is metrizable, there will be many metrics which give rise to the same topology. If X and Y are metrizable, then so is $X \times Y$: if d_X metrizes X and d_Y metrizes Y, then

$$d((a,b),(c,d)) = d_X(a,c) + d_Y(b,d)$$

defines a metric which gives rise to the product topology on $X \times Y$. Note that this procedure does *not* give the standard metric on $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$, introduced previously: "balls" with respect to this product metric are diamond-shaped. However, the product metric and the standard metric do give rise to the same topology on $\mathbb{R}^2$ (and the metrics are also equivalent in stronger, geometric, senses).

Countability properties and **separation properties** indicate the extent to which a topological space behaves like a metric space. A topological space X is **separable** if it has a countable dense subset. X is **first countable** if there is a countable base at every $x \in X$, and X is **second countable** if X has a countable base. We noted above that $\mathbb{R}^m$ is separable. It also follows easily from the Stone-Weirstrass theorem that if $X \subseteq \mathbb{R}^m$ is compact then $C(X)$ is separable. Further:

- Any metric space is first countable.
- A metric space is second countable iff it is separable.

A topological space X is **Hausdorff** if, for any distinct points $c, d \in X$, there are disjoint open sets A and B with $c \in A$ and $d \in B$. A Hausdorff space is **normal** if, for any disjoint closed sets C and D, there are disjoint open sets A and B with $C \subseteq A$ and $D \subseteq B$. We note that:

- Any compact subset of a Hausdorff space is closed.
- Any singleton subset $\{a\}$ of a Hausdorff space is closed.
- Compact Hausdorff spaces are normal.
- Metrizable topological spaces are Hausdorff and normal.

Note that if X is indiscrete and contains at least two points, then X is not Hausdorff: it is thus immediate that (X, d_i) is not metrizable. A much deeper result is:

Urysohn's Lemma. *Suppose C and D are disjoint closed subsets of a normal Hausdorff space X. Then there is a continuous function $f : X \to [0,1]$ such that $f = 0$ on C and $f = 1$ on D.*

Note that Urysohn's lemma is easy to prove for a metric space: if C and D are non-empty, we can simply define

$$f(x) = \frac{\mathrm{dist}(x, C)}{\mathrm{dist}(x, C) + \mathrm{dist}(x, D)}\,.$$

If X is locally compact and Hausdorff then Urysohn's lemma still holds, as long as we assume one of the disjoint closed sets, C or D, is compact.

One application of Urysohn's lemma is to prove that many normal spaces are in fact metrizable:

Urysohn Metrization Theorem. *If X is Hausdorff, normal and second countable then X is metrizable.*

Urysohn's metrization theorem continues to hold if X is locally compact, Hausdorff and second countable.