Real Analysis, Fall 2017–Spring 2018

April 5, 2018

Contents

Chapter 1

The real number line

1.1 Ordered Sets

One basic property of many number systems (natural numbers, integers, rationals, etc) is that they are *ordered*, so we say that "3 is greater than 2", and so on.

1.1.1 Definition. A *total order* on a set S is a relation[1] $\leq$ satisfying the following axioms:

(O1) (Reflexivity) For every element a, it always holds that $a \leq a$.

(O2) (Antisymmetry) If $a \leq b$ and $b \leq a$, then it must be that $a = b$.

(O3) (Transitivity) If $a \leq b$ and $b \leq c$, then it holds that $a \leq c$.

(O4) (Totality) For every pair of elements a and b, either $a \leq b$ or $b \leq a$.

We say S is an *ordered set*.

1.1.2 Example. Find some examples of ordered sets.

1.1.3 Example. Find an example of a *partially ordered* set—a set with a relation satisfying axioms (O1)–(O3) but not (O4).

1.1.4 Problem. Suppose S is an ordered set. Formulate a reasonable definition of strict inequality ($a < b$) in terms of the order relation $\leq$. Then write down a definition equivalent to Definition 1.1.1 using strict inequality as the primitive relation; that is, write down a set of axioms that $<$ should satisfy, in terms of which $\leq$ (suitably defined in terms of $<$) has properties (O1)–(O4).

1.1.5 Definition. Let S be an ordered set, and $A \subseteq S$ a subset. An *upper bound* for A is an element $u \in S$ such that $a \leq u$ for every $a \in A$. If such an element exists, we say A is *bounded above*.

Similarly, a *lower bound* for A is an element $l \in S$ such that $l \leq a$ for every $a \in A$. If such a lower bound exists, we say A is *bounded below*.

1.1.6 Definition. A *least upper bound* or *supremum* of a bounded above set A is an element u_0 of S such that

(i) u_0 is an upper bound for A, and

(ii) $u_0 \leq u$ for every other upper bound u.

We denote a supremum for A (if it exists) by $\sup A$.

Similarly, a *greatest lower bound* or *infimum* of a bounded below set A is an element b_0 of S such that

(i) b_0 is a lower bound for A, and

[1] A relation is a comparision operation between two elements which evaluates to either *true* or *false*.

(ii) $b_0 \geq b$ for every other lower bound b.

We denote an infimum for A (if it exists) by $\inf A$.

1.1.7 Proposition. *If a supremum (or infimum) of A exists, then it is unique.*

1.1.8 Proposition. *If A and B are subsets of an ordered set S which both have a supremum and an infimum and satisfy $A \subseteq B$, then*

$$\inf B \leq \inf A \leq \sup A \leq \sup B. \tag{1.1}$$

1.1.9 Example. Let $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ denote the set of integers, with the usual order. Find some examples of subsets A of $\mathbb{Z}$ such that

(i) A is bounded above and below.

(ii) A is bounded above but not below.

(iii) A is not bounded above and not bounded below.

Which of these sets have a supremum? Which have an infimum?

1.1.10 Example. Repeat Example 1.1.9 with the set $\mathbb{Q}$ of rational numbers in place of $\mathbb{Z}$. The following Lemma may be of use.

1.1.11 Lemma. *There exists no $q \in \mathbb{Q}$ such that $q^2 = 2$.*

Proof hint: Write $q = \frac{a}{b}$ in lowest terms and consider the evenness/oddness of a and b. □

1.1.12 Definition. An ordered set S has the *least upper bound property* if every subset which is bounded above has a supremum. Likewise S has the *greatest lower bound property* if every subset which is bounded below has an infimum.

1.1.13 Example. Does $\mathbb{Z}$ have the least upper bound property? Does $\mathbb{Q}$? Justify your answers with a proof or counterexample.

1.1.14 Theorem. *If S has the least upper bound property, then it has the greatest lower bound property.*

1.2 Fields and ordered fields

Of course the familiar number systems have additional structure. Besides the order, we have addition, subtraction, multiplication and division.

1.2.1 Definition. A *field* is a set $\mathbb{F}$ with two binary operations[2] $+$ and $\cdot$, called *addition* and *multiplication*, respectively, satisfying the following axioms:

(F1) (Associativity of addition) $(a + b) + c = a + (b + c)$ for all a, b, c in $\mathbb{F}$.

(F2) (Additive identity) There exists an element $0 \in \mathbb{F}$ such that $0 + a = a + 0 = a$ for all a.

(F3) (Additive inverses) For each a in $\mathbb{F}$ there exists an element $-a$ such that $(-a) + a = a + (-a) = 0$.

(F4) (Commutativity of addition) $a + b = b + a$ for all a, b in $\mathbb{F}$.

(F5) (Associativity of multiplication) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all a, b, c in $\mathbb{F}$.

(F6) (Multiplicative identity) There exists an element $1 \in \mathbb{F}$ such that $1 \cdot a = a \cdot 1 = a$ for all a.

(F7) (Multiplicative inverses) For all $a \neq 0$, there exists an element a^{-1} in $\mathbb{F}$ such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

[2] A binary operation is a function/operation taking in two elements of $\mathbb{F}$ and returning a third element of $\mathbb{F}$.

(F8) (Commutativity of multiplication) $a \cdot b = b \cdot a$ for all a, b in $\mathbb{F}$.

(F9) (Distributivity) $a \cdot (b + c) = a \cdot b + a \cdot c$.

(F10) (Nontriviality) $0 \neq 1$.

It is customary to omit the $\cdot$ when writing multiplication; in other words, we usually just write ab instead of $a \cdot b$. Additionally, we usually denote $a + (-b)$ simply by $a - b$, and we may also use the notation $\frac{1}{a}$ in place of a^{-1}. It is important to note that subtraction $-$ and division $\div$ are not really distinct operations; they are just syntactic shorthand for addition (resp. multiplication) by an additive (resp. multiplicative) inverse.

We also use the usual shorthand a^n in place of $\underbrace{a \cdots a}_{n \text{ times}}$ and na in place of $\underbrace{a + \cdots + a}_{n \text{ times}}$.

Remark. Though we shall be entirely concerned with fields in this course, you may be familiar with various mathematical objects satisfying fewer of the above axioms. A set with a single operation satisfying axioms (F1)–(F3) is a *group* which is said to be *commutative* or *abelian* if (F4) also holds.

A *ring* is a set with two operations satisfying all of the above except (F7), (F8) and (F10). A *commutative ring* satisfies (F8). According to some conventions, a ring need not satisfy (F6), though such "rings without identity" are sometimes cutely referred to as '*rng*'s. If (F7) holds but not (F8), then $\mathbb{F}$ is called a *division ring*.

Axiom (F10) might be considered optional for fields, but if we allow $0 = 1$ then $\mathbb{F}$ must be the one element set $\{0\}$ (you can prove this after you prove Proposition 1.2.4 below), which for various reasons is best not regarded as a field.

1.2.2 Example. Come up with some examples of fields, some with infinitely many and some with finitely many elements. Can you construct a field with exactly two elements? Three?

1.2.3 Proposition. *The following properties of addition and multiplication hold in any field. (That is, they follow from the axioms above.)*

(i) *(Uniqueness of identities) If an element b in $\mathbb{F}$ satisfies $b + a = a$ for some a, then $b = 0$. Likewise if b satisfies $ba = a$ for some $a \neq 0$, then $b = 1$.*

(ii) *(Uniqueness of inverses) If b satisfies $a + b = 0$, then $b = -a$. Likewise, if b satisfies $ba = 1$ then $b = a^{-1}$.*

(iii) *(Cancellation) If $a + c = b + c$ then $a = b$. Likewise if $c \neq 0$ and $ac = bc$, then $a = b$.*

(iv) *(Inverse of an inverse) $-(-a) = a$ and $(a^{-1})^{-1} = a$.*

1.2.4 Proposition. *In any field, the following properties hold.*

(i) *$0a = 0$ for all a.*

(ii) *If $ab = 0$, then either $a = 0$ or $b = 0$. (We say $\mathbb{F}$ "has no divisors of zero".)*

(iii) *$(-a)b = a(-b) = -(ab)$ for all a and b. In particular $-a = (-1)a$.*

(iv) *$(-a)(-b) = ab$ for all a and b.*

1.2.5 Problem. In a field, show that if $b \neq 0$ and $d \neq 0$ then

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}.$$

1.2.6 Definition. An *ordered field* is a field $\mathbb{F}$ equipped with a total order, so a set with a relation $\leq$ and two operations $+$ and $\cdot$ satisfying axioms (O1)–(O4) and (F1)–(F10), which is additionally required to satisfy the following axioms:

(OF1) (Compatibility of order and addition) If $a \leq b$ then $a + c \leq b + c$ for any c.

(OF2) (Compatibility of order and multiplication) If $a \leq b$ and $0 \leq c$, then $ac \leq bc$.

1.2.7 Example. Which examples from Example 1.2.2 are ordered fields? In case there is not an obvious order, is there any order at all satisfying (OF1) and (OF2)?

1.2.8 Proposition. *The following properties always hold in an ordered field.*

(i) *If* $0 \leq a$ *then* $-a \leq 0$.

(ii) *If* $0 \leq a$ *and* $0 \leq b$ *then* $0 \leq ab$. *(In fact, this is equivalent to (OF2) and is often used in place of it as the other ordered field axiom).*

(iii) *If* $a \leq 0$ *and* $0 \leq b$, *then* $ab \leq 0$.

(iv) $0 \leq a^2$ *for any* a. *In particular* $0 < 1$.

(v) *If* $0 < a \leq b$ *then* $0 < b^{-1} \leq a^{-1}$.

In light of Proposition 1.2.4.(ii) *the above identities hold with strict inequality* $<$ *used in place of inequality* $\leq$.

1.2.9 Problem. Let $\mathbb{F}$ be an ordered field and consider the subset $Z \subset \mathbb{F}$ generated by taking 0, 1, $1+1$, $1+1+1$, etc. along with -1, $-1-1$, $-1-1-1$, etc. Show that this set is in bijection with the set of integers $\mathbb{Z}$.

Likewise, let $Q \subset \mathbb{F}$ be the subset generated by taking the multiplicative inverses of the nonzero elements in Z along with their integer multiples. Show that this set is in bijection with $\mathbb{Q}$.

Thus every ordered field contains a copy of $\mathbb{Q}$, which may be regarded as the "smallest" possible ordered field.

1.2.10 Definition. Let $\mathbb{F}$ be an ordered field. The *absolute value* or *magnitude* of a number $a \in \mathbb{F}$ is defined by

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

1.2.11 Proposition*. *The absolute value satisfies the following properties. For all* a *and* b *in* $\mathbb{F}$*:*

(i) $|a| \geq 0$.

(ii) $|a| = 0$ *if and only if* $a = 0$.

(iii) $|ab| = |a|\,|b|$.

(iv) *(Triangle inequality)* $|a+b| \leq |a| + |b|$.

(v) *(Reverse triangle inequality)* $\big|\,|a| - |b|\,\big| \leq |a-b|$.

Remark. Combining (iv) and (v) of the last proposition gives the useful strings of inequalities:

$$|a| - |b| \leq \big|\,|a| - |b|\,\big| \leq |a+b| \leq |a| + |b|\,, \quad \text{and} \quad |a| - |b| \leq \big|\,|a| - |b|\,\big| \leq |a-b| \leq |a| + |b|\,. \tag{1.2}$$

1.2.12 Definition. The *distance* between numbers a and b in an ordered field $\mathbb{F}$ is the quantity

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

1.2.13 Proposition*. *The distance satisfies the following properties. For all* a, b, *and* c *in* $\mathbb{F}$*:*

(i) $d(a,b) \geq 0$.

(ii) $d(a,b) = 0$ *if and only if* $a = b$.

(iii) *(Symmetry)* $d(a, b) = d(b, a)$.

(iv) *(Triangle inequality)* $d(a, c) \leq d(a, b) + d(b, c)$.

1.2.14 Lemma (Suprema/infima in an ordered field)**.** *Let A be a bounded above subset of an ordered field. Then $u = \sup A$ if and only if*

(i) *$a \leq u$ for all $a \in A$ (i.e., u is an upper bound), and*

(ii) *for every $\varepsilon > 0$, there exists $a \in A$ such that $u - \varepsilon < a$ (i.e., $u - \varepsilon$ fails to be an upper bound).*

Similarly, if A is bounded below, then $b = \inf A$ if and only if

(i) *$b \leq a$ for all $a \in A$, and*

(ii) *for every $\varepsilon > 0$, there exists $a \in A$ such that $a < b + \varepsilon$.*

1.2.15 Definition ($\pm\infty$ notation)**.** As a notation convention, it is useful to introduce the symbols $+\infty$ and $-\infty$ when speaking of suprema and infima in an ordered field. We write $\sup A = +\infty$ if A is not bounded above, and $\inf A = -\infty$ if A is not bounded below. With these conventions $\sup A$ and $\inf A$ are always defined for a nonempty set A, and (1.1) holds identically whenever $A \subseteq B$.

A more formal way to do this is to embed $\mathbb{F}$ into a larger ordered set $\overline{\mathbb{F}} = \mathbb{F} \cup \{+\infty, -\infty\}$ with the order defined so that $-\infty < a < +\infty$ for all $a \in \mathbb{F}$. Note that $\overline{\mathbb{F}}$ is *not* a field, though we may observe the following notation conventions: if $a > 0 \in \mathbb{F}$, then

$$\begin{aligned} &a + (+\infty) = +\infty, \quad a + (-\infty) = -\infty, \quad a(+\infty) = +\infty, \quad a(-\infty) = -\infty, \\ &(-a)(+\infty) = -\infty, \quad (-a)(-\infty) = +\infty, \quad \frac{\pm a}{\pm\infty} = 0. \end{aligned}$$

Expressions such as $+\infty - \infty$ and $\pm\infty / \pm\infty$ are not defined.

1.3 Completeness and the real number field

1.3.1 Definition. An ordered field $\mathbb{F}$ is *complete* if it satisfies the least upper bound property (c.f. Definition 1.1.12), in other words, if for every bounded above subset $A \subset \mathbb{F}$, the supremum (least upper bound) $\sup A$ exists in $\mathbb{F}$.

1.3.2 Theorem† (Characterization/definition of $\mathbb{R}$)**.** *There exists a unique*[3] *complete ordered field called the* real numbers *and denoted by* $\mathbb{R}$.

Remark. We omit the proof of Theorem 1.3.2 for now; we may come back to it later on. However, it is worth mentioning one construction which is possible at this point: define a *Dedekind cut* to be a subset $A \subset \mathbb{Q}$ of the rationals with the properties that

(i) A is neither empty nor all of $\mathbb{Q}$,

(ii) if $q \in A$ and $p < q$, then $p \in A$,

(iii) if $q \in A$ then $q < r$ for some $r \in A$.

[3] Here "uniqueness" means the following: given two complete ordered fields F_1 and F_2, there exists an *isomorphism* (a bijection compatible with the order and field operations) $\phi : F_1 \longrightarrow F_2$. Moreover ϕ is unique. Using ϕ we can regard F_1 and F_2 as being "the same" field.

In other words, a cut is essentially a half infinite open interval in $\mathbb{Q}$; take as an example $\{q \in \mathbb{Q} : q < 2\} = (-\infty, 2)$. It is tempting to want to write $\{q \in \mathbb{Q} : q < \sqrt{2}\}$ as another example, but this is ill-specified since we do not have such a number as $\sqrt{2}$ at this point. The equivalent set may be specified as $\{q \in \mathbb{Q} : q < 0 \text{ or } q^2 < 2\}$. The idea here is that real numbers are represented by the "upper endpoints" of the cuts, though since these are not well-defined, the whole cut stands in as a replacement.

It is then possible to define an order, addition, and multiplication on the set of Dedekind cuts (order and addition are straightforward; multiplication is a little tricky) and verify that they satisfy all the axioms of an ordered field along with completeness, with subfield $\mathbb{Q}$ identified with those cuts of the form $\{q \in \mathbb{Q} : q < p\}$ for $p \in \mathbb{Q}$.

1.3.3 Definition. An ordered field $\mathbb{F}$ is *Archimedean* if for every $a \in \mathbb{F}$, there exists an integer[4] N such that $a \leq N$.

1.3.4 Example*. Show that $\mathbb{Q}$ is Archimedean.

1.3.5 Example (Research Allowed)**.** Find an example of a non-Archimedean field.

1.3.6 Theorem. *As a complete ordered field, $\mathbb{R}$ is Archimedean.*

1.3.7 Proposition. *A field is Archimedean if and only if, for every $a > 0$, there exists a positive integer N such that*

$$0 < \frac{1}{N} < a.$$

Remark. The Archimedean property says that a field has no "infinitely large" elements, and via Proposition 1.3.7, it implies that there are no "infinitely small" elements. The next result gives a technically useful if strange seeming characterization of the zero element.

1.3.8 Corollary. *In an Archimedean field, if $0 \leq a$ and $a < \varepsilon$ for every $0 < \varepsilon$, then $a = 0$.*

1.3.9 Theorem (Density of $\mathbb{Q}$ in $\mathbb{R}$)**.** *Let a and b be real numbers with $a < b$. Then there exists a rational number q such that*

$$a < q < b.$$

We say $\mathbb{Q}$ is dense *in $\mathbb{R}$.*

Remark. This may be a surprising result, especially when juxtaposed with the following one. Recall that an infinite set is said to be *countable* if it is in bijection with the set $\mathbb{N} = \{1, 2, 3, \ldots\}$ of natural numbers.

1.3.10 Theorem.

(i) $\mathbb{Q}$ *is countable.*

(ii) $\mathbb{R}$ *is uncountable.*

One more result at this point will be useful later on, though the proof is rather technical and tricky, so you may go ahead and take it as given rather than trying to prove it.

1.3.11 Theorem† (Positive nth roots)**.** *For every $y > 0$ in $\mathbb{R}$ and $n \in \mathbb{N}$, there exists a unique $x > 0$ in $\mathbb{R}$ such that $x^n = y$.*

The proof is obtained from the following two results, the first of which is more or less straightforward while the second is the tricky one.

1.3.12 Lemma. *For fixed $y > 0$ and $n \in \mathbb{N}$, the set $E = \{t \in \mathbb{R} : 0 < t,\ t^n < y\}$ is nonempty and bounded above.*

1.3.13 Lemma†. *The element $x = \sup E$ satisfies $x^n = y$.*

[4] Here we are identifying a subset of $\mathbb{F}$ with the integers as in Problem 1.2.9.

1.4 Sequences of real numbers

While least upper bounds give an expedient way to express the completeness of $\mathbb{R}$, *sequences* play a much more ubiquitous role in analysis.

1.4.1 Definition. A *sequence* of real numbers is a function[5] from $\mathbb{N}$ into $\mathbb{R}$. As a matter of notation, if $x : \mathbb{N} \longrightarrow \mathbb{R}$ is a sequence, we prefer to write x_n instead of $x(n)$, and denote the sequence by

$$(x_1, x_2, x_3, \ldots), \quad \text{or} \quad (x_n)_{n=1}^{\infty}, \quad \text{or just} \quad (x_n).$$

It is permissible and often convenient to index a sequence starting from 0 instead of 1, or starting from a number greater than 1.

1.4.2 Definition. A sequence (x_n) in $\mathbb{R}$ is said to be

(i) *bounded* if there exists some $B > 0$ such that $|x_n| \leq B$ for all n.

(ii) *increasing* if $x_n \leq x_{n+1}$ for all n. It is *strictly increasing* if $x_n < x_{n+1}$ for all n.

(iii) *decreasing* if $x_n \geq x_{n+1}$ for all n. It is *strictly decreasing* if $x_n > x_{n+1}$ for all n.

(iv) *monotone* if it is either increasing or decreasing.

(v) *convergent* if there exists some $L \in \mathbb{R}$ with the following property: for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\text{for all } n \geq N, \quad |x_n - L| < \varepsilon.$$

Equivalently, for every $\varepsilon > 0$, the interval $(L - \varepsilon, L + \varepsilon) \subset \mathbb{R}$ contains all but finitely many terms of the sequence. In this case we say L is the *limit* of the sequence (x_n) and write $\lim_{n\to\infty} x_n = L$ or $x_n \to L$.

1.4.3 Exposition*. Explain in plain English what is meant by the limit of a sequence. Address the order of the quantifiers (i.e., "for all" or "there exists"): if the order or type of the quantifiers is changed, why is this a bad definition of limit?

1.4.4 Proposition. *The limit of a sequence, if it exists, is unique.*

1.4.5 Example*.

(i) Show that the constant sequence $x_n = c$ for all n converges and $\lim x_n = c$.

(ii) Show that $x_n = \frac{1+3n}{1+5n}$ has limit $\frac{3}{5}$.

1.4.6 Example.

(i) Show that $x_n = \frac{1}{n} \to 0$.

(ii) If $0 \leq p < 1$, show that $x_n = p^n \to 0$. [Hint: such p can be written as $p = \frac{1}{1+a}$ for $a > 0$. The binomial estimate $(1+a)^n \geq 1 + na$ for $a \geq 0$ is also useful here.]

1.4.7 Proposition. *If a sequence converges, then it is bounded.*

1.4.8 Theorem. *In $\mathbb{R}$, every bounded monotone sequence converges.*

Remark. The previous property of $\mathbb{R}$ is referred to as the *monotone sequence property.* In fact, it is equivalent to completeness: it can be proved that an ordered field in which every bounded monotone sequence converges has the least upper bound property.

1.4.9 Problem. Let $x_n = \sqrt{n^2+1} - n$. Show (x_n) converges and compute its limit.

[5] Recall that a *function* $f : A \longrightarrow B$ is an assignment to every element a of the *domain* A an element $b = f(a)$ of the *target* B.

1.4.10 Theorem. *Let (x_n) and (y_n) be convergent sequences with limits x and y, respectively. Then*

(i) $x_n + y_n \to x + y$.

(ii) $-x_n \to -x$.

(iii) $x_n y_n \to xy$.

(iv) *If $x_n \neq 0$ for all n and $x \neq 0$, then $x_n^{-1} \to x^{-1}$.*

(v) *If $x_n \leq y_n$ for all n, then $x \leq y$.*

Remark. Combining the above, it follows that $x_n - y_n \to x - y$ and $x_n/y_n \to x/y$ (provided $y_n \neq 0$ and $y \neq 0$).

1.4.11 Problem. Let $p(t) = a_0 + \cdots + a_l t^l$ and $q(t) = b_0 + \cdots + b_m t^m$ be polynomials with real coefficients, where $a_l \neq 0$ and $b_m \neq 0$. If $l \leq m$, prove that

$$\lim_{n\to\infty} \frac{p(n)}{q(n)} = \begin{cases} 0, & \text{if } l < m, \text{ and} \\ a_l/b_m, & \text{if } l = m. \end{cases}$$

1.4.12 Problem. Define a sequence inductively by setting $x_1 = \sqrt{2}$ and for $n \geq 2$, $x_n = \sqrt{2 + x_{n-1}}$. Prove that (x_n) converges and find its limit.

1.4.13 Example*. Show that strict inequality cannot be obtained in Theorem 1.4.10.(v), by producing an example where $x_n < y_n$ for all n but $x = y$.

1.4.14 Example* (Harmonic series/sequence)**.** Define a sequence by $x_1 = 1$, $x_2 = 1 + \frac{1}{2}$, and $x_n = 1 + \cdots + \frac{1}{n}$. Show (x_n) is monotone increasing, but unbounded above, hence does not converge. [Possible hint: compare to the sequence $y_1 = 1$, $y_2 = 1 + \frac{1}{2}$, $y_3 = 1 + \frac{1}{2} + \frac{1}{4}$, $y_4 = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4}$, $y_5 = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8}$ etc., where y_n has each term from the corresponding x_n replaced by the largest power of 2^{-1} which is less than or equal to it.]

The following limits are useful to know but tricky to prove at this point[6], so you can take them as given rather than trying to prove them.

1.4.15 Proposition†.

(i) *For any $a > 0$, the sequence $a^{1/n} \to 1$.*

(ii) *The sequence $n^{1/n} \to 1$.*

In addition to bounded monotone sequences, another very useful class of sequences which "ought to converge" are the *Cauchy sequences*.

1.4.16 Definition. A sequence (x_n) is *Cauchy* if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\text{for every } n, m \geq N, \quad |x_n - x_m| < \varepsilon. \tag{1.3}$$

Remark. Intuitively, a Cauchy sequence is one in which the *tails* of the sequence (i.e., the sequences $(x_n)_{n=N}^{\infty}$ for various N) become arbitrarily "bunched up".

1.4.17 Proposition. *Every Cauchy sequence is bounded.*

1.4.18 Proposition. *Every convergent sequence is Cauchy.*

[6] In fact they become quite easy to prove once we have developed the logarithm, but we do not have this yet.

The converse is not true in general; however one of the distiguishing features of $\mathbb{R}$ over $\mathbb{Q}$ as a complete ordered field is the following main result.

1.4.19 Theorem. *In $\mathbb{R}$, every Cauchy sequence converges.*

Proof hint. For each k let $a_k = \sup\{x_n : n \geq k\}$. Then (a_k) is a bounded decreasing sequence. Show that $x_n \to a$, where $a = \lim_k a_k$. □

Remark. This property is known as the *Cauchy completeness* of $\mathbb{R}$. It is possible to show that it is equivalent to both the least upper bound property and to the monotone sequence property.

1.4.20 Example. Show that (1.3) cannot be replaced by "for every $n \geq N$, $|x_n - x_{n+1}| < \varepsilon$", by finding an example of a divergent sequence in $\mathbb{R}$ with the latter property.

Remark. The previous exercise shows that it is generally not enough for pairs of adjacent elements in the sequence to be getting close together; rather, we need the distance between any pair of not-necessarily-adjacent elements in some tail of the sequence to be getting close. On the other hand, if adjacent pairs become close fast enough, then the sequence actually is Cauchy, as the next result shows.

1.4.21 Lemma. *Let (x_n) be a sequence in $\mathbb{R}$ such that for all $n \in \mathbb{N}$,*

$$|x_n - x_{n+1}| \leq \frac{a}{2^n}, \quad \textit{for some } a > 0.$$

Then (x_n) is Cauchy, and therefore convergent. (In fact $a/2^n$ can be replaced by a/b^n for any $b > 1$.)

Remark. Having discussed sequences, we can now mention two more methods of constructing $\mathbb{R}$ from $\mathbb{Q}$. In both cases we consider as elements equivalence classes of sequences (x_n) in $\mathbb{Q}$, with the equivalence relation $(x_n) \sim (y_n)$ if $x_n - y_n \to 0$.

The first method uses equivalence classes of *bounded increasing sequences*, while the second method uses equivalence classes of *Cauchy sequences*. Each element $q \in \mathbb{Q}$ is represented by the equivalence class of the constant sequence $x_n = q$ for all n.

In either method, one has to define a notion of addition, multiplication, and order on sequences, show these are well-defined on equivalence classes, and prove that the set of equivalence classes satisfies all the axioms for an ordered field, along with some version of completeness (usually the monotone sequence property if you are using monotone sequences, and the Cauchy sequence property if you are using Cauchy sequences).

1.4.22 Theorem (Squeeze theorem)**.** *Let (x_n), (y_n) and (z_n) be sequences in $\mathbb{R}$ such that $x_n \leq y_n \leq z_n$ for all n, and suppose that $x_n \to l$ and $z_n \to l$. Then (y_n) also converges and* $\lim_n y_n = l$.

1.4.23 Exposition*. Explain in plain English what is the significance of the completeness of $\mathbb{R}$. Why is this a useful property to have?

1.5 Subsequences

1.5.1 Definition. Let (x_n) be a sequence in $\mathbb{R}$. A *subsequence* of (x_n) is a sequence $(x_{n_k})_{k=1}^{\infty}$ where $n_1 < n_2 < n_3 < \cdots$ form a strictly increasing sequence (n_k) of natural numbers.

If $x_{n_k} \to l$ for some subsequence (x_{n_k}) of (x_n), we say l is a *subsequential limit* of (x_n).

1.5.2 Example. Find examples of a non-convergent sequence (x_n) with

(i) No subsequential limits.

(ii) Exactly one subsequential limit.

(iii) Exactly two subsequential limits.

(iv) Exactly three subsequential limits.

(v) Infinitely many subsequential limits.

1.5.3 Proposition. *A number $l \in \mathbb{R}$ is a subsequential limit of a sequence (x_n) if and only if, for every $\varepsilon > 0$,*

$$|x_n - l| < \varepsilon$$

for infinitely many $n \in \mathbb{N}$.

1.5.4 Proposition. *If $x_n \to x$, then every subsequence of (x_n) converges to x.*

1.5.5 Proposition. *If (x_n) is a Cauchy sequence (not necessarily in $\mathbb{R}$, perhaps in $\mathbb{Q}$), and $x_{n_k} \to x$ for some subsequence (x_{n_k}), then $x_n \to x$.*

1.5.6 Theorem. *Every sequence in $\mathbb{R}$ or $\mathbb{Q}$ has a monotone subsequence (either increasing or decreasing).*

1.5.7 Corollary. *If (x_n) is a sequence in $[a, b] \subset \mathbb{R}$, for some $a \leq b$, then (x_n) has a convergent subsequence in $\mathbb{R}$.*

Prior to discussing limit superior and inferior, it is convenient to make the following definition.

1.5.8 Definition. Given a sequence, we say $x_n \to +\infty$ if, for every $M > 0$, there exists $N \in \mathbb{N}$ such that

$$M < x_n \quad \text{for all } n \geq N.$$

Likewise, we say $x_n \to -\infty$ if for every $M < 0$, there exists $N \in \mathbb{N}$ such that

$$x_n < M \quad \text{for all } n \geq N.$$

In neither case do we regard (x_n) as a convergent sequence in $\mathbb{R}$; however, it is possible to regard it as convergent in the *extended real numbers* $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, regarded as an ordered set (but not a field) as in Definition 1.2.15.

1.5.9 Proposition*. *If (x_n) is a monotone sequence, then $x_n \to l$ for some $l \in \mathbb{R} \cup \{\pm\infty\}$.*

1.5.10 Definition. Let (x_n) be a sequence in $\mathbb{R}$ and let $L = \{l \in \mathbb{R} \cup \{\pm\infty\} : x_{n_k} \to l, \text{ for some subsequence } (x_{n_k})\}$ be the set of its subsequential limits, including possibly $+\infty$ and $-\infty$. The *limit superior* of (x_n) is the supremum

$$\limsup x_n = \sup L,$$

and the *limit inferior* of (x_n) is the infimum

$$\liminf x_n = \inf L.$$

1.5.11 Proposition. *For every sequence (x_n) in $\mathbb{R}$, there exists a subsequence (x_{n_k}) such that*

$$x_{n_k} \to \limsup x_n.$$

Likewise, there exists a subsequence (x_{n_k}) such that $x_{n_k} \to \liminf x_n$.

In other words, in the definition of limit superior, $\limsup x_n = \sup L$ and

The next result justifies the names "limit inferior" and "limit superior".

1.5.12 Proposition. *An equivalent characterization of limit superior and limit inferior are as follows. Let (x_n) be a real sequence and for each $m \in \mathbb{N}$, let $a_m = \inf \{x_n : n \geq m\}$ and $b_m = \sup \{x_n : n \geq m\}$. Then $a = \liminf x_n$ if and only if*

$$a = \lim a_m = \lim_m \inf \{x_n : n \geq m\} = \sup \{\inf \{x_n : n \geq m\}\}.$$

Likewise, $b = \limsup x_n$ if and only if

$$b = \lim b_m = \lim_m \sup \{x_n : n \geq m\} = \inf \{\sup \{x_n : n \geq m\}\}.$$

1.5.13 Proposition. *For a sequence (x_n),*

$$\liminf x_n \leq \limsup x_n$$

with equality if and only if (x_n) converges to this value.

1.6 Series

1.6.1 Definition. A *series* is a formal sum $\sum_{k=0}^{\infty} a_k$ where (a_k) is a sequence of real numbers. It may or may not represent a well-defined number. We say the series *converges* if the sequence (s_n) of *partial sums*

$$s_n = \sum_{k=0}^{n} a_k = a_0 + a_1 + \cdots + a_n$$

converges, and then we set $\sum_{k=0}^{\infty} a_k = \lim_{n\to\infty} s_n$. Otherwise, we say that $\sum_{k=0}^{\infty} a_k$ *diverges*.

As with sequences, it may be convenient for the index k to start at a number other than 0. Also, it is common to use the shorthand $\sum_k a_k$ or just $\sum a_k$ when the range of indices is understood.

1.6.2 Example. Let r be a real number. The *geometric series* $\sum_{n=0}^{\infty} r^n$ converges to $\frac{1}{1-r}$ if $|r| < 1$ and diverges if $|r| \geq 1$.

1.6.3 Theorem. *Let $a = \sum a_n$ and $b = \sum b_n$ be convergent series. Then the series $\sum(a_n + b_n)$ converges to $a+b$. Likewise, for any $c \in \mathbb{R}$, the series $\sum c a_n$ converges to ca.*

In light of Theorem 1.4.19 and Proposition 1.4.18, we recall that sequences converge in $\mathbb{R}$ if and only if they are Cauchy. Applying this to the sequence of partial sums of a series, we obtain the following *Cauchy criterion* for convergence.

1.6.4 Theorem*. *A series $\sum_{k=0}^{\infty} a_k$ converges if and only if, for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that*

$$\left| \sum_{k=m}^{n} a_k \right| = |a_m + \cdots + a_n| < \varepsilon, \quad \text{for all } n \geq m \geq N.$$

1.6.5 Corollary. *If the series $\sum_{n=0}^{\infty} a_n$ converges, then the sequence of terms (a_n) converges to 0.*

Remark. The converse to Corollary 1.6.5 is false in view of the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$, which was shown to diverge in Example 1.4.14.

From the monotone sequence property (Theorem 1.4.8), we have another criterion for convergence in the case that the terms of a sequence are all non-negative.

1.6.6 Theorem. *If $a_k \geq 0$ for all k, then $\sum_{k=0}^{\infty} a_k$ converges if and only if all of the partial sums are bounded; in other words, there exists $M > 0$ such that $s_n = \sum_{k=0}^{n} a_k < M$ for all n.*

Often we can show that a series converges or diverges by comparing it to a known series using the following result.

1.6.7 Theorem (Comparison test)**.** *If $|a_n| \leq b_n$ for all but finitely many n and $\sum b_n$ converges, then $\sum a_n$ also converges (not necessarily to the same value). If $0 \leq c_n \leq a_n$ for all but finitely many n and $\sum c_n$ diverges, then $\sum a_n$ also diverges.*

Returning to examples, we have the following result about so-called *p-series* $\sum_{n=1}^{\infty} \frac{1}{n^p}$.

1.6.8 Theorem. *Let p be a real number. The series $\sum \frac{1}{n^p}$ converges if $p > 1$ and diverges if $p \leq 1$.*

Proof hint. Along the lines of Example 1.4.14, consider the series $\sum_{k=0}^{\infty} 2^k \left(\frac{1}{2^k}\right)^p$ (or $\frac{1}{2}\sum_{k=0}^{\infty} 2^k \left(\frac{1}{2^k}\right)^p$). By Theorem 1.6.6 it suffices to show that the partial sums of both series are either both bounded or both unbounded. □

The outlined proof in fact applies more generally. By a similar argument, one can obtain the following.

1.6.9 Theorem. *Let $a_1 \geq a_2 \geq a_3 \geq \cdots \geq 0$. Then $\sum_n a_n$ converges if and only if the series $\sum_k 2^k a_{2^k}$ converges.*

Besides the comparison theorem, the workhorse convergence theorems in the theory of series are the ratio and root tests.

1.6.10 Theorem (Root test). *Suppose the sequence* $(|a_n|^{1/n})$ *converges to* $\alpha \in \mathbb{R}$. *If* $\alpha < 1$, *then the series* $\sum a_n$ *converges. If* $\alpha > 1$, *then* $\sum a_n$ *diverges. If* $\alpha = 1$, *the test is inconclusive.*

Proof hint. Let $\alpha < \beta < 1$ and show that $a_n < \beta^n$ for n sufficiently large. If $\alpha > \beta > 1$, then $|a_n| > 1$ for infinitely many n. ☐

1.6.11 Theorem (Ratio test). *Suppose that the sequence* $(|a_{n+1}/a_n|)$ *converges to* $\alpha \in \mathbb{R}$. *If* $\alpha < 1$, *then* $\sum a_n$ *converges. If* $\alpha > 1$, *then* $\sum a_n$ *diverges. If* $\alpha = 1$, *the test is inconclusive.*

Proof hint. If $\alpha < \beta < 1$, then $|a_{N+k}| \leq \beta^k |a_N|$ for some N. ☐

Remark. Of course, the sequences $|a_n|^{1/n}$ and $|a_{n+1}/a_n|$ may not converge, but the ratio and root test can be strengthened as follows. The root test holds with $\alpha = \limsup_n |a_n|^{1/n}$, which always exists, and in the ratio test, convergence obtains when $\limsup_n |a_{n+1}/a_n| < 1$ while divergence obtains when $\liminf_n |a_{n+1}/a_n| > 1$.

While the ratio test is often the easiest to apply, the root test is actually stronger, in the sense that for any series the ratio test is conclusive about, the root test will also be conclusive. This is a consequence of the following result.

1.6.12 Theorem†. *For any real sequence* (a_n), *the inequalities*

$$\liminf |a_{n+1}/a_n| \leq \liminf |a_n|^{1/n} \leq \limsup |a_n|^{1/n} \leq \limsup |a_{n+1}/a_n|$$

hold.

1.6.13 Problem*. Determine whether the following series converge or diverge:

(i) $\sum \frac{n}{n^2+3}$

(ii) $\sum \frac{1}{n^2+1}$

(iii) $\sum \frac{2^n}{n!}$

(iv) $\sum \frac{1}{2^n+n}$

(v) $\sum 2^{(-1)^n - n}$

None of the theorems or tests so far apply to the *alternating harmonic series* $\sum \frac{(-1)^n}{n}$. This does not compare to a known convergent or divergent series, and both the ratio and root tests are inconclusive. Since the terms in this series alternate between positive and negative, this type of series is known as an *alternating series.*

1.6.14 Theorem (Alternating series). *Let* $a_1 \geq a_2 \geq \cdots \geq 0$. *Then the alternating series* $\sum_n (-1)^n a_n$ *converges if and only if* $a_n \to 0$.

1.7 Absolute convergence

1.7.1 Definition. Let $\sum a_n$ be a series. If $\sum |a_n|$ converges, then (by Theorem 1.6.7) $\sum a_n$ also converges and we say $\sum a_n$ is *absolutely convergent.* If $\sum a_n$ converges but $\sum |a_n|$ diverges, we say $\sum a_n$ is *conditionally convergent.*

Remark. Note that in the comparison, ratio, and root tests, any series which is shown to converge actually converges absolutely.

1.7.2 Example. By Theorem 1.6.14 and Theorem 1.6.8 the alternating harmonic series $\sum \frac{(-1)^n}{n}$ is conditionally convergent.

There are two main results concerning absolute convergence. The first concerns *rearrangements*. By commutativity of addition, any finite sum such as $a_1 + a_2 + a_3$ may be rearranged in any order, say $a_2 + a_3 + a_1$, without changing the result. This is not generally true for series.

1.7.3 Definition. Let $\sum a_n$ be a series. A *rearrangmenent* of $\sum a_n$ is a series $\sum b_n$, where $(b_n) = (a_{i(n)})$ for some bijection $i : \mathbb{N} \longrightarrow \mathbb{N}$.

1.7.4 Theorem†. *If $\sum a_n$ is absolutely convergent, then every rearrangement is also absolutely convergent and converges to the same value.*

The converse is far from true, as evidenced by the following remarkable result.

1.7.5 Theorem†. *Let $\sum a_n$ is conditionally convergent and let α be any real number. Then there exists a rearrangement of $\sum a_n$ which converges to α.*

Remark. In fact, one can choose any $-\infty \leq \alpha \leq \beta \leq +\infty$ and find a rearrangement whose sequence (s_n) of partial sums satisfies $\liminf s_n = \alpha$, $\limsup s_n = \beta$.

The second main result about absolutely convergent series concerns products. It takes a bit of thought to figure out what a sensible definition of the product of two series should be; given $\sum a_n$ and $\sum b_n$, the first instinct is to consider $\sum(a_n b_n)$, but this is not a very meaningful quantity; we wouldn't expect that $\sum(a_n b_n) = (\sum a_n)(\sum b_n)$, for instance.

Instead we take motivation from power series (which we shall come to later on): given $\sum a_n x^n$ and $\sum b_n x^n$, if we take their product formally, without worrying about any kind of convergence, and collect powers of x, we get

$$(\sum a_n x^n)(\sum b_n x^n) = \sum c_n x^n, \quad c_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

1.7.6 Definition. Given two series $\sum a_n$ and $\sum b_n$, define their *Cauchy product* to be the series $\sum c_n$, where $c_n = \sum_{k=0}^{n} a_k b_{n-k}$.

In general, convergence of $\sum a_n$ and $\sum b_n$ does not imply convergence of their product, nor does convergence of the product imply that it converges to the product of $\sum a_n$ and $\sum b_n$.

1.7.7 Theorem†. *Let $a = \sum a_n$ and $b = \sum b_n$ be convergent series, and suppose that $\sum a_n$ converges absolutely. Then the Cauchy product $\sum c_n$ converges, where $c_n = \sum_{k=0}^{n} a_k b_{n-k}$, and*

$$\sum c_n = ab.$$

Chapter 2

Topology of metric spaces

2.1 Metric spaces

The study of sequences, convergence, and of continuous functions in $\mathbb{R}$ really depend only on certain properties of the distance $d(x, y) = |x - y|$. It is useful therefore to work in a more abstract setting, where the results obtained may then apply more broadly. We will work in the setting of *metric spaces*.

2.1.1 Definition. A *metric space* (M, d) is a set M equipped with a real-valued function

$$\begin{aligned} d : M \times M &\longrightarrow \mathbb{R}, \\ (x, y) &\longmapsto d(x, y) \end{aligned}$$

satisfying the following properties:

(M1) (Positivity) $d(x, y) \geq 0$ for all $x, y \in M$.

(M2) (Nondegeneracy) $d(x, y) = 0$ if and only if $x = y$.

(M3) (Symmetry) $d(x, y) = d(y, x)$ for all $x, y \in M$.

(M4) (Triangle inequality) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in M$.

We say d is a *metric on* M.

If $S \subset M$ is any subset, then (S, d) is a metric space in its own right, which we refer to as a *subspace* of M.

2.1.2 Example. In light of Proposition 1.2.13, $\mathbb{R}$ is a metric space with respect to the metric $d(x, y) = |x - y|$.

2.1.3 Example*. Any set S may be equipped with the *discrete metric* δ, defined by

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

Show that (S, δ) is a metric space.

2.1.4 Definition. We define n-dimensional *Euclidean space* to be the set $\mathbb{R}^n = \{x = (x_1, \ldots, x_n) : x_i \in \mathbb{R},\ i = 1, \ldots, n\}$ of ordered n-tuples of real numbers. From linear algebra (or otherwise), we know that $\mathbb{R}^n$ is a *vector space*, meaning it has an *vector addition* operation

$$x + y = (x_1 + y_1, \ldots, x_n + y_n), \quad x, y \in \mathbb{R}^n$$

which is associative, commutative, with identity $0 = (0, \ldots, 0)$, and inverses $-x = (-x_1, \ldots, -x_n)$ (compare axioms (F1)–(F4) of Definition 1.2.1), and a *scalar multiplication* operation

$$ax = (ax_1, \ldots, ax_n), \quad a \in \mathbb{R},\ x \in \mathbb{R}^n,$$

which is appropriately associative and distributive and satisfies $1x = x$. $\mathbb{R}^n$ also has a natural *Euclidean inner product*

$$\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n \in \mathbb{R}$$

which satisfies $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$, $\langle x, y \rangle = \langle y, x \rangle$ and $a\langle x, y \rangle = \langle ax, y \rangle$ for $x, y, z \in \mathbb{R}^n$, $a \in \mathbb{R}$. We define the *Euclidean norm* by

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + \cdots + x_n^2}.$$

2.1.5 Lemma (Cauchy-Schwartz inequality)**.** *For all $x, y \in \mathbb{R}^n$,*

$$|\langle x, y \rangle| \leq \|x\| \, \|y\| \, .$$

[Hint: consider $\langle x - \alpha y, x - \alpha y \rangle$ for a particularly well chosen value of α. Note that using calculus to find a minimizing α is not out of the question here, since the calculus need not enter the actual proof!

2.1.6 Example. Show that each of the following are metrics on $\mathbb{R}^n$:

(i) The *Euclidean (aka ℓ^2) metric*

$$d_2(x, y) = \|x - y\| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

(ii) The ℓ^1 metric

$$d_1(x, y) = |x_1 - y_1| + \cdots + |x_n - y_n| \, .$$

(iii) The ℓ^∞ metric

$$d_\infty(x, y) = \max_{i \in \{1, \ldots, n\}} |x_i - y_i| \, .$$

Unless otherwise specified, we usually consider $\mathbb{R}^n$ with the Euclidean metric.

Remark. The previous example shows that the same set may have the structure of a metric space with respect to several different metrics (we could even consider $\mathbb{R}^n$ with the discrete metric). This is why a particular metric must be specified (or genreally understood) when we assert that a set is a metric space.

The following estimates are routinely useful in $\mathbb{R}^n$, and can be interpreted as saying that the metrics d_1, d_2 and d_∞ on $\mathbb{R}^n$ are comparable to one another (though not generally equal).

2.1.7 Lemma. *For any $z = (z_1, \ldots, z_n) \in \mathbb{R}^n$,*

$$\frac{1}{\sqrt{n}}(|z_1| + \cdots + |z_n|) \leq \sqrt{|z_1|^2 + \cdots + |z_n|^2} \leq (|z_1| + \cdots + |z_n|)$$

and

$$\max_{i \in \{1, \ldots, n\}} |z_i| \leq \sqrt{|z_1|^2 + \cdots + |z_n|^2} \leq \sqrt{n} \, (\max_i |z_i|)$$

In particular, for any $x, y \in \mathbb{R}^n$,

$$\frac{1}{\sqrt{n}} d_1(x, y) \leq d_2(x, y) \leq d_1(x, y), \quad d_\infty(x, y) \leq d_2(x, y) \leq \sqrt{n} \, d_\infty(x, y).$$

2.2 Sequences

Apart from monotonicity, and any statements relating to order, most of the definitions and results about sequences in $\mathbb{R}$ carry over to the general setting of metric spaces.

2.2.1 Definition. Let M be a metric space and $S \subseteq M$. Then S is said to be *bounded* if for some point $p \in M$ there exists a constant $B \geq 0$ such that $d(q, p) \leq B$ for all $q \in S$.

Remark. Using the triangle inequality, it follows that S is bounded if and only if, for *every* point $p \in M$, there exists $B \geq 0$ such that $d(q, p) \leq B$ for all $q \in S$.

2.2.2 Definition. A sequence (x_n) in a metric space M is said to be

(i) *bounded* if the set $\{x_n : n \in \mathbb{N}\} \subset M$ is bounded,

(ii) *convergent*, with *limit* $l \in M$, if there exists $l \in M$ with the property that, for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$d(x_n, l) < \varepsilon \quad \text{for all } n \geq N.$$

We write $x_n \to l$ or $l = \lim_{n\to\infty} x_n$ as usual.

(iii) *Cauchy* if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$d(x_n, x_m) < \varepsilon \quad \text{for all } n, m \geq N.$$

We define a *subsequence* (x_{n_k}) in the usual way, via a strictly increasing sequence $n_1 < n_2 < \cdots$ of indices, and we say $l \in M$ is a *subsequential limit* if $x_{n_k} \to l$ for a subsequence (x_{n_k}).

The proofs of Propositions 1.4.4, 1.4.17, 1.4.18, 1.5.3, 1.5.4, and 1.5.5 rely fundamentally only on the properties (M1)–(M4) of the metric $d(x, y) = |x - y|$ on $\mathbb{R}$. Hence their proofs carry over at once to prove the following.

2.2.3 Theorem. *The following hold for sequences in a metric space M:*

(i) *The limit of a sequence, if it exists, is unique.*

(ii) *Every convergent sequence is Cauchy.*

(iii) *Every Cauchy sequence (and hence by (ii) every convergent sequence) is bounded.*

(iv) $x_n \to l$ *if and only if every subsequence of* (x_n) *converges to* l.

(v) *If* (x_n) *is Cauchy and some subsequence converges to* l, *then* $x_n \to l$.

(vi) $l \in M$ *is a subsequential limit of* (x_n) *if and only if, for every* $\varepsilon > 0$, $d(x_n, l) < \varepsilon$ *for infinitely many* n.

(vii) $l \in M$ *is the limit of* (x_n) *if and only if, for every* $\varepsilon > 0$, $d(x_n, l) < \varepsilon$ *for all but finitely many* n.

The following notion is fundamentally useful in the study of metric spaces.

2.2.4 Definition. For $\varepsilon > 0$, we define the *ε-ball centered at* $p \in M$, or *ε-neighborhood of* p to be the set

$$D_\varepsilon(p) = \{q \in M : d(p, q) < \varepsilon\}.$$

The name "ball" is justified by the example $M = \mathbb{R}^n$, with the Euclidean metric. The term "neighborhood" comes from the idea that points in $D_\varepsilon(p)$ "live close to" p. We will use the two terms interchangably.

Remark. In terms of ε-balls, we can reinterpret the last two items in Theorem 2.2.3 as follows: l is a subsequential limit (respectively, the limit) of (x_n) if and only if for every $\varepsilon > 0$, $D_\varepsilon(l)$ contains infinitely many (resp. all but finitely many) elements of the sequence.

2.2.5 Example. Let S be a metric space with the discrete metric δ (c.f. Example 2.1.3). What do the ε-balls in S look like for various ε? Which sequences are Cauchy in S? Which sequences are convergent?

2.2.6 Example. What do the ε-balls look like in $\mathbb{R}^2$ with respect to the Euclidean metric d_2? How about with respect to d_1 or d_∞?

2.2.7 Definition. Let (x_k) be a sequence in $\mathbb{R}^n$. Temporarily writing superscripts for components, i.e., $x = (x^1, \ldots, x^n) \in \mathbb{R}^n$, we can write each term in the sequence as

$$x_k = (x_k^1, \ldots, x_k^n) \in \mathbb{R}^n.$$

Fixing the superscript i and letting k vary, we get sequences $(x_k^i)_{k=1}^\infty$ in $\mathbb{R}$, for $i = 1, \ldots, n$, which we call the *component sequences* of (x_k).

2.2.8 Proposition. *Let (x_k) be a sequence in $\mathbb{R}^n$.*

(i) *(x_k) converges to $x = (x^1, \ldots, x^n)$ if and only if each component sequence converges:*

$$x_k^i \to x^i, \quad i = 1, \ldots, n.$$

(ii) *(x_k) is Cauchy if and only if each component sequence $(x_k^i)_{k=1}^\infty$ is a Cauchy sequence in $\mathbb{R}$.*

2.2.9 Definition. We say a metric space (M, d) is *complete* if every Cauchy sequence converges in M.

2.2.10 Theorem*.

(i) *Every Euclidean space $\mathbb{R}^n$, $n \in \mathbb{N}$, is a complete metric space.*

(ii) *Any discrete metric space (S, δ) is complete.*

2.2.11 Example. What are some examples of incomplete metric spaces?

The process by which we obtain $\mathbb{R}$ from $\mathbb{Q}$ is an example of a more general operation in the setting of metric spaces, called the *metric completion*. In this general setting, we say a subspace $S \subset M$ is *dense in M* if for every $p \in M$ and for every $\varepsilon > 0$, there is some $q \in S$ such that $d(p, q) < \varepsilon$. In other words, for every $\varepsilon > 0$ and $p \in M$, the ε-ball $D_\varepsilon(p)$ contains some point of S.

2.2.12 Theorem†. *For any metric space (M, d), there exists a complete metric space (M', d') in which M is dense. More precisely, M is in bijection $i : M \longrightarrow S'$ with a dense subspace $S' \subset M'$ under which the metrics agree: $d'(i(p), i(q)) = d(p, q)$.*

Proof sketch. Define M' to be the set of equivalence classes of all Cauchy sequences (x_n) in M, where (x_n) and (y_n) are equivalent if $d(x_n, y_n) \to 0$. In fact one can show that for any two Cauchy sequences (x_n) and (y_n) in M, the sequence $d(x_n, y_n)$ is Cauchy (hence convergent) in $\mathbb{R}$; the metric on M' is then defined by $d'\big((x_n), (y_n)\big) = \lim_n d(x_n, y_n)$. For $p \in M$, take $i(p)$ to be the constant sequence $(p, p, p, \ldots)$. $\square$

2.3 Open and closed sets

2.3.1 Definition. A subset $U \subseteq M$ of a metric space is said to be *open* if for every $p \in U$, there exists some $\varepsilon > 0$ such that the ε-neighborhood of p is contained in U:

$$D_\varepsilon(p) \subseteq U.$$

A set $E \subseteq M$ is said to be *closed* if its complement, $M \setminus E$, is open, where

$$M \setminus E = \{p \in M : p \notin E\}.$$

Remark. Intuitively, open sets are those that do not contain their edges or boundaries[1]; the definition formalizes the idea that no matter how close to the boundary of an open set you get, you can always go a bit farther while remaining inside the set. The condition that ε be strictly positive is important.

The whole space M is itself an open set, since every $D_\varepsilon(p)$ is in M by definition, and the empty set $\emptyset$ is also open, since there are no points to which to apply the condition in Definition 2.3.1, so it holds vacuously. Further examples are afforded by the next results.

2.3.2 Proposition*. *For every $\varepsilon > 0$ and $p \in M$, the ε-ball $D_\varepsilon(p)$ is an open set.*

2.3.3 Proposition*. *For every $p \in M$, the single point set $\{p\}$ is closed.*

2.3.4 Example. Find examples of sets in a metric space which are

(i) Both open and closed.

(ii) Neither open nor closed.

2.3.5 Theorem.

(i) *An arbitrary union of open sets is open; i.e., if $\{U_a : a \in A\}$ is a collection of open sets in a metric space M, then the set $\bigcup_{a\in A} U_a \subseteq M$ is open. The indexing set A need not be finite nor even countable.*

(ii) *A finite intersection of open sets is open; i.e., if $U_1, \dots, U_n$ are open, then $U_1 \cap \cdots \cap U_n$ is open.*

(iii) *A finite union of closed sets is closed.*

(iv) *An arbitrary intersection of closed sets is closed.*

2.3.6 Example. Find examples of

(i) An infinite intersection of open sets in $\mathbb{R}$ which is not open.

(ii) An infinite union of closed sets in $\mathbb{R}$ which is not closed.

2.3.7 Example. What are the open (resp. closed) sets in a discrete metric space (S, δ)?

2.3.8 Definition. A point $p \in M$ is an *limit point* of a set $A \subseteq M$ if, for every $\varepsilon > 0$, there exists some point $q \in A$ such that

$$0 < d(p, q) < \varepsilon.$$

(Note that we cannot take $q = p$ since $d(p, q)$ is required to be positive.) Equivalently, every neighborhood $D_\varepsilon(p)$ contains some point of A other than p itself (which may or may not be in A).

2.3.9 Example. What are the limit points of the following subsets of $\mathbb{R}$?

(i) $(0, 1)$

(ii) $[1, 2] \cup \{3\}$

(iii) $\{n : n \in \mathbb{N}\}$

(iv) $\left\{\frac{1}{n} : n \in \mathbb{N}\right\}$

(v) The set of values $\{x_n : n \in \mathbb{N}\}$ for a sequence (x_n) in $\mathbb{R}$ with distinct terms (i.e., $x_n \neq x_m$ whenever $n \neq m$). What about for an arbitrary sequence?

2.3.10 Proposition. *A point p is an limit point of $A \subseteq M$ if and only if there exists a sequence (x_n) such that $x_n \in A$ and $x_n \neq p$ for all n, but $x_n \to p$.*

2.3.11 Proposition. *The following are equivalent for a subset $E \subseteq M$:*

(i) *E is closed.*

(ii) *E contains all its limit points.*

(iii) *For every sequence (x_n) in E which converges in M, the limit $\lim_n x_n$ lies in E.*

[1] We will formalize the notion of boundary below.

2.4 Interior and Closure

2.4.1 Definition. For an arbitrary set $A \subseteq M$ in a metric space, we say $p \in A$ is an *interior point* if $D_\varepsilon(p) \subseteq A$ for some $\varepsilon > 0$. The *interior of* A, written A°, is the set of all interior points of A.

2.4.2 Proposition. *For any $A \subseteq M$,*

(i) *A° is open.*

(ii) *A is open if and only if $A = A^\circ$.*

(iii) *if $U \subseteq A$ is any open set contained in A, then $U \subseteq A^\circ$.*

Remark. The first and the last points in the previous result can be interpreted as saying that A° is the largest open set contained in A.

2.4.3 Definition. The *closure* of a set $A \subseteq M$ is the set

$$A^- = A \cup \{p : p \text{ is a limit point of } A\}$$

consisting of A and all its limit points.

2.4.4 Lemma. *Let $A \subseteq M$. Then $p \in A^-$ if and only if for all $\varepsilon > 0$, there exists a point $q \in A$ such that $d(p,q) < \varepsilon$. [Note that we do not exclude the case that $p = q$ here.]*

2.4.5 Proposition. *For any $A \subseteq M$,*

(i) *A^- is closed.*

(ii) *A is closed if and only if $A = A^-$.*

(iii) *If $C \supseteq A$ is any closed set containing A, then $A^- \subseteq C$.*

(iv) *A^- is the complement of the interior of the complement of A:*

$$M \setminus (A^-) = (M \setminus A)^\circ.$$

Remark. The first and the third points in the previous result can be interpreted as saying that A^- is the smallest closed set containing A.

2.4.6 Example*. What are the interior and closure of the following sets in $\mathbb{R}$?

(i) $[0,1]$

(ii) $[0,1) \cup \{2\}$

(iii) $(0,1) \cap \mathbb{Q}$

2.4.7 Problem. Starting with a single set $A \subseteq \mathbb{R}$, what is the largest number of distinct sets you can obtain by iteratively applying the interior and closure operations? (i.e., A, A°, $(A^\circ)^-$, $(A^-)^\circ$, etc.) What is an example of a set generating the maximum number?

2.4.8 Definition. The *boundary* of $A \subseteq M$ is the set

$$\partial A = A^- \setminus A^\circ,$$

and we say $p \in M$ is a *boundary point* of A if $p \in \partial A$.

2.4.9 Proposition. *Let $A \subseteq M$. Then $p \in \partial A$ if and only if for every $\varepsilon > 0$, the ball $D_\varepsilon(p)$ contains both a point of A and a point not in A.*

2.4.10 Example*. What are the boundaries of the sets in Example 2.4.6?

2.5 Continuous functions

2.5.1 Definition. Let (M, d) and (N, d') be metric spaces, and $p \in M$. Then a function[2] $f : A \subseteq M \longrightarrow N$ is said to be *continuous at* $p \in A$ if for every real number $\varepsilon > 0$, there exists a real number $\delta > 0$ such that

$$d(p, q) < \delta \text{ implies } d'\big(f(p), f(q)\big) < \varepsilon.$$

Note that δ may depend on both p and ε. We say $f : A \subseteq M \longrightarrow N$ is simply *continuous* if it is continuous at each $p \in A$.

2.5.2 Example. Show using the definition of continuity that each of the following functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ are continuous.

(i) $f(x) = 1$.

(ii) $f(x) = x$.

(iii) $f(x) = x^2$.

2.5.3 Example. Let (S, δ) be a discrete metric space and M any metric space. Then any function $f : S \longrightarrow M$ is continuous.

2.5.4 Example. Let $f : \mathbb{R} \longrightarrow S$ be a function on $\mathbb{R}$ valued in a discrete metric space. Then f is continuous if and only if it is constant.

There are equivalent characterizations of continuity using sequences and open sets, respectively.

2.5.5 Theorem. *Let (M, d) and (N, d') be metric spaces. A function $f : A \subseteq M \longrightarrow N$ is continuous at $p \in A$ if and only if $f(p_n) \to f(p)$ for every sequence (p_n) in A such that $p_n \to p$.*

2.5.6 Theorem. *$f : M \longrightarrow N$ is continuous if and only if, for every open set $U \subseteq N$, the inverse image*

$$f^{-1}(U) = \{p \in M : f(p) \in U\}$$

is open in M.

2.5.7 Corollary. *$f : M \longrightarrow N$ is continuous if and only if, for every closed set $C \subseteq N$, $f^{-1}(C)$ is closed in M.*

2.5.8 Example. Show that the image, $f(U) = \{f(p) : p \in U\}$, of an open set under a continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ need not be open.

2.5.9 Problem. Let

$$f(x, y) = \begin{cases} \frac{y^4}{x^2+y^2}, & (x, y) \neq 0, \\ 0 & (x, y) = 0. \end{cases}$$

Use Theorem 2.5.5 to prove that f is continuous on $\mathbb{R}^2 \setminus (0, 0)$ and use the $\varepsilon - \delta$ definition to show that f is continuous at $(0, 0)$.

2.5.10 Problem*. Use either the definition or Theorem 2.5.5 to show that the following functions are not continuous at the given points:

(i) $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0, \end{cases}$$

at the point $x = 0$.

[2] Here we allow f to be defined on a *domain* A which may not be all of M.

(ii) $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} \frac{xy^2}{x^2+y^4}, & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

at the point $(x,y) = (0,0)$.

2.5.11 Problem. Show that the unit circle $\{(x,y) : x^2 + y^2 = 1\}$ is closed in $\mathbb{R}^2$ using Corollary 2.5.7.

2.5.12 Theorem. *Let (M_i, d_i) be metric spaces for $i = 1, 2, 3$, and suppose $f : M_1 \longrightarrow M_2$ and $g : M_2 \longrightarrow M_3$ are continuous functions. Then the composition $g \circ f : M_1 \longrightarrow M_3$ is continuous.*

Remark. Try and give three different proofs of Theorem 2.5.12, using Definition 2.5.1, Theorem 2.5.5, and Theorem 2.5.6.

Next we will focus on some results particular to the case where the domain and/or range of the functions are Euclidean spaces.

2.5.13 Theorem. *Let $f, g : (M, d) \longrightarrow \mathbb{R}$ be continuous, real valued functions on a metric space. Then $f + g$ and fg are continuous, and furthermore if $f(p) \neq 0$ for all $p \in M$, then $1/f$ is continuous.*

2.5.14 Proposition. *The coordinate projections*

$$\pi_i : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad \pi_i(x_1, \ldots, x_n) = x_i$$

are continuous for each $i = 1, \ldots, n$.

Let $f : A \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a function. Since the values of f lie in $\mathbb{R}^m$, we can write f as

$$f(x_1, \ldots, x_n) = \big(f_1(x_1, \ldots, x_n),\ f_2(x_1, \ldots, x_n), \ldots,\ f_m(x_1, \ldots, x_n)\big)$$

where the $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ are scalar functions for $i = 1, \ldots, m$, which we will refer to as the *component functions* of f.

2.5.15 Proposition. *$f : A \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is continuous if and only if each of its component functions is continuous.*

Using composition, sums and products of simpler functions, we can show that fairly complicated functions between Euclidean spaces are continuous.

2.5.16 Example. Justify why the following functions are continuous:

(a) $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ given by

$$f(x,y,z) = (xyz^3,\ 4 + xy).$$

(b) $f : \mathbb{R}^2 \setminus \{(0,0)\} \longrightarrow \mathbb{R}^3$ given by

$$f(x,y) = \big(\tfrac{xy}{x^2+y^2},\ \sin(x+y),\ e^{-1/(x^2+y^2)}\big)$$

(Though we have not yet formally defined them in this class, you can assume standard continuity results for the elementary real functions $\sin(t)$ and e^t.)

2.6 Compactness

2.6.1 Definition. Let (M, d) be a metric space, and $A \subseteq M$ a subset. An *open cover* of A is a set $\mathcal{U} = \{U_i : i \in I\}$ of open sets $U_i \subseteq M$ such that

$$A \subseteq \bigcup_{i \in I} U_i.$$

In other words, each U_i is open and for each $p \in A$ there exists some $i \in I$ such that $p \in U_i$. The indexing set I is allowed to be finite or infinite, countable or uncountable.

We say $\mathcal{U}'$ is a *subcover* of $\mathcal{U}$ if each $U_i \in \mathcal{U}'$ is also in $\mathcal{U}$, and if $\mathcal{U}'$ is also a cover of A. In other words, $\mathcal{U}'$ is a subset of $\mathcal{U}$, and each $p \in A$ lies in some open set in $\mathcal{U}'$.

2.6.2 Example. Give an example of an open cover of the set $[0, 1] \subset \mathbb{R}$ which is

(i) finite,

(ii) countably infinite,

(iii) uncountably infinite.

Give an example of an open cover of $[0, 1]$ which has a proper subcover (i.e., throwing out at least one set from the open cover still gives an open cover).

2.6.3 Definition. A set $A \subseteq M$ is *compact* if every open cover of A has a finite subcover.

2.6.4 Example. Show that the following subsets of $\mathbb{R}$ are not compact by constructing an open cover which has no finite subcover (i.e., any finite collection of the open sets of the cover fails to cover the given set):

(i) $(0, 1)$

(ii) $\mathbb{R}$ itself

It is very difficult to use the definition to show that a set actually *is* compact, except in limited cases. Note that while any *particular* open cover may have a finite subcover, this must be true for *every* open cover of a set in order for that set to be compact.

2.6.5 Example. Every single point set $\{p\} \subset M$ is compact, and more generally if $A \subset M$ is a finite set of points, then A is compact.

Passing from finite sets to countably infinite ones, we can produce examples of both compact and non compact sets.

2.6.6 Problem. Let $x_n \to x$ be a convergent sequence in a metric space M. Show that the set $A = \{x_n : n \in \mathbb{N}\} \cup \{x\}$ is compact.

2.6.7 Problem*. Give some examples of countably infinite sets of $\mathbb{R}$ which are not compact, and give justification.

2.6.8 Theorem. *Let A be a compact set in a metric space (M, d). Then every sequence (p_n) in A has a subsequence which converges in A.*

Proof hint. Assume otherwise, and construct an open cover starting with appropriate ε balls around your sequence points, which has no finite subcover. $\square$

The converse direction also holds, but is quite a bit harder to prove:

2.6.9 Theorem† (Bolzano-Weierstrass)**.** *A subset A of a metric space is compact if and only if every sequence (p_n) in A has a convergent subsequence with limit in A.*

2.6.10 Proposition. *In any metric space, a compact set is closed.*

2.6.11 Proposition. *A closed subset of a compact set is compact.*

2.6.12 Theorem. *Let $f : M \to N$ be a continuous function between metric spaces, and let $A \subset M$ be a compact set. Then $f(A) = \{f(p) \in N : p \in A\}$ is a compact set in N.*

2.6.13 Theorem. *Let $a \leq b$ be real numbers. Then $[a, b] \subset \mathbb{R}$ is compact.*

Proof hint. Let $\mathcal{U}$ be an open cover of $[a, b]$ and define

$$A = \{x \in [a, b] : [a, x] \text{ is covered by finitely many } U \in \mathcal{U}\}, \quad \alpha = \sup A.$$

Show that 1) $\alpha \in A$, and 2) $\alpha = b$. □

2.6.14 Corollary (Heine-Borel theorem for $\mathbb{R}$)**.** *A subset $A \subset \mathbb{R}$ is compact if and only if it is closed and bounded.*

The next property of compact sets, due to Cantor, is known variously as the "nested set property" or the "Cantor intersection theorem".

2.6.15 Theorem (Cantor's intersection theorem)**.** *Let K_i, $i \in \mathbb{N}$ be a* nested sequence *of compact sets in a metric space M, meaning*

$$K_1 \supseteq K_2 \supseteq K_3 \supseteq \cdots$$

Then the intersection $\bigcap_{i=1}^{\infty} K_i$ is nonempty.

Proof hint: Proceed by contradiction and consider the open sets $O_j = M \setminus K_j$ for $j \geq 2$. □

2.6.16 Example. The following "Cantor sets" are nonempty:

(i) the classic Cantor "middle thirds" set, obtained by deleting the open middle third of $[0, 1]$, (i.e., replacing $[0, 1]$ by $[0, 1/3] \cup [2/3, 0]$) and then inductively deleting the middle third of each remaining interval in countably many steps.

(ii) more general sets, such as obtained by replacing $[0, 1]$ by $[2/6, 3/6] \cup [4/6, 5/6]$ and so on.

2.6.17 Example. Show that the hypothesis of compactness is necessary in Theorem 2.6.15 by giving an example of a nested sequence of noncompact sets in $\mathbb{R}$ whose intersection is empty.

On our way to generalizing Theorem2.6.13 to $\mathbb{R}^n$, we will use the following notion and an intermediate result.

2.6.18 Definition. An *n-cell* in $\mathbb{R}^n$ is a product

$$I = [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$$

of finite closed intervals, so each a_i and b_i are real numbers with $a_i \leq b_i$.

2.6.19 Lemma. *Let $I_1 \supseteq I_2 \supseteq I_3 \supset \cdots$ be a nested sequence of n cells in $\mathbb{R}^n$. Then the intersection $\bigcap_{i=1}^{\infty} I_i$ is nonempty.*

2.6.20 Theorem (Heine-Borel for $\mathbb{R}^n$)**.** *An n-cell I in $\mathbb{R}^n$ is compact.*

Proof hint: Note that there exists some $r > 0$ such that $d(x, y) \leq r$ for any pair $x, y \in I$. Call this the *raduis* of I. By subdividing all the intervals $[a_i, b_i]$ into $[a_i, c_i]$ and $[c_i, b_i]$, where $c_i = (a_i + b_i)/2$, we can write I as the union of 2^n smaller n-cells of radius $r/2$.

Now suppose $\mathcal{U} = \{U_\alpha\}$ is an open cover of I with no finite subcover. Subdivide I into 2^n smaller cells as above and note that at least one of these is not covered by finitely many of the U_α. Repeat this process and use Lemma 2.6.19 to get a contradiction. □

2.6.21 Corollary. *A set $A \subset \mathbb{R}^n$ is compact if and only if it is closed and bounded.*

One of the most important properties of compact sets is the *extreme value theorem*, which is a simple corollary of Theorems 2.6.12 and Corollary 2.6.14.

2.6.22 Theorem (Extreme value theorem). *Let $f : A \subseteq M \longrightarrow \mathbb{R}$ be a real valued continuous function with domain a compact set A in a metric space M. Then the* extreme values

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

are actually acheived by f on A. In other words there exist points x_0 and x_1 in A such that $f(x_0) = \alpha$ and $f(x_1) = \beta$.

2.7 Connectedness

2.7.1 Definition. Let A be a set in a metric space M. We say A is *disconnected* if there exist open sets U, V such that

(i) $A \subset U \cup V$,

(ii) $A \cap U \neq \emptyset$ and $A \cap V \neq \emptyset$, and

(iii) $U \cap V = \emptyset$.

We say A is *connected* if it is not disconnected.

2.7.2 Example. Show that any two point set $\{p_1, p_2\}$ where $p_1 \neq p_2$ is disconnected. More generally, any finite set of 2 or more points is disconnected.

2.7.3 Example. Let $a < b$ in $\mathbb{R}$. Show that $(a, b) \cap \mathbb{Q}$ is disconnected.

2.7.4 Definition. A nonempty set $A \subseteq \mathbb{R}$ will be called an *interval* if it is of the form

$$[a, b], \quad (a, b), \quad [a, b), \quad \text{or} \quad (a, b],$$

where a and b are extended real numbers, i.e., $a, b \in \mathbb{R} \cup \{\pm\infty\}$. (If one of a or b is $\pm\infty$, then the corresponding half of the interval should be open.) Equivalently, A is an interval if $\inf A < x < \sup A$ implies $x \in A$.

The next two results should confirm your intuition about intervals:

2.7.5 Proposition*. *Every interval in $\mathbb{R}$ is connected.*

2.7.6 Proposition. *If $A \subseteq \mathbb{R}$ is connected, then A is an interval.*

As with compact sets, connected sets have an important relationship with continuous functions.

2.7.7 Theorem. *Let $f : M \longrightarrow N$ be a continuous function between metric spaces, and let $A \subseteq M$. If A is connected, then $f(A) \subseteq N$ is connected.*

2.7.8 Corollary (Intermediate value theorem). *Let $f : M \longrightarrow \mathbb{R}$ be a real-valued continuous function on a metric space M, and suppose $A \subseteq M$ is connected. If $f(p) < f(q)$ for two points $p, q \in A$, then for any $c \in \mathbb{R}$ with $f(p) < c < f(q)$, there exists a third point $r \in A$ such that $f(r) = c$.*

Chapter 3

Single variable calculus

3.1 Limits and derivatives

We now return to functions $f : A \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ in order to talk about differentiation and integration.

3.1.1 Definition. Let $f : A \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ be a function, and let p be a limit point of A. We say $L \in \mathbb{R}$ is the *limit of f at p* if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$0 < |x - p| < \delta \quad \text{implies} \quad |f(x) - L| < \varepsilon.$$

In this case we write $\lim_{x \to p} f(x) = L$, or $f(x) \to L$ as $x \to p$. If no such L exists, then we say the *limit of f at p does not exist.*

Remark. Note that $f(p)$ need not equal L, nor need f be even defined at p. However for the definition to be meaningful, we do require that p be a limit point of A. (If p is an isolated point of A, then the definition becomes vacuuous.)

Note also the similarity of Definition 3.1.1 to the definition of continuity (Definition 2.5.1). In terms of limits, f is continuous at p if and only if $\lim_{x \to p} = f(p)$.

The proof of Theorem 2.5.5 is easily adapted to show the following characterization of functional limits in terms of sequential limits.

3.1.2 Proposition. $L = \lim_{x \to p} f(x)$ *if and only if, for every sequence (p_n) such that $p_n \neq p$ for all n but $p_n \to p$, we have $f(p_n) \to L$.*

3.1.3 Proposition. *Suppose that the limits of two functions f and g exist at $p \in \mathbb{R}$. Then*

(i) $\lim_{x \to p} \big(f(x) + g(x)\big) = \lim_{x \to p} f(x) + \lim_{x \to p} g(x)$

(ii) $\lim_{x \to p} f(x) g(x) = \lim_{x \to p} f(x) \lim_{x \to p} g(x)$

(iii) *Provided $f(x) \neq 0$ for all x and $\lim_{x \to p} f(x) \neq 0$, then $\lim_{x \to p} f(x)^{-1} = \big(\lim_{x \to p} f(x)\big)^{-1}$.*

3.1.4 Definition. Let $f : [a, b] \subset \mathbb{R} \longrightarrow \mathbb{R}$, and suppose $a < x < b$. We say f is *differentiable at x* if

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} \tag{3.1}$$

exists, and then we say $f'(x)$ is the *derivative of f at x*. The limit in (3.1) can alternatively be written as $\lim_{y \to x} \frac{f(y) - f(x)}{y - x}$. Equivalently, f is differentiable at x with derivative $f'(x)$ if the function

$$\phi_{f,x}(h) := f(x + h) - f(x) - f'(x)h$$

satisfies $\lim_{h \to 0} \frac{\phi_{f,x}(h)}{h} = 0$.

Remark. Note that we require $a < x < b$ in the definition, so the largest set on which f may be differentiable is (a, b). We will not discuss any notion of differentiability at the endpoints a and b.

3.1.5 Proposition. *If f is differentiable at $x \in (a, b)$, then f is continuous at x.*

3.1.6 Example. Give an example of a function $f : [a, b] \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ which is continuous at some $x \in (a, b)$ but not differentiable there.

3.1.7 Example*.

(i) If $f(x) = c$ for some constant for all $x \in [a, b]$, then $f'(x) = 0$ for all $x \in (a, b)$.

(ii) If $f(x) = x$, then $f'(x) = 1$.

3.1.8 Theorem. *If f and g are defined on $[a, b]$ and differentiable at $x \in (a, b)$, then*

(i) $(f + g)'(x) = f'(x) + g'(x)$

(ii) $(fg)'(x) = f(x)g'(x) + g(x)f'(x)$

(iii) *If in addition $f(x) \neq 0$ then $(1/f)'(x) = -\frac{f'(x)}{f(x)^2}$*

(iv) *If $n \in \mathbb{N}$, and $f(x) = x^n$, then $f'(x) = nx^{n-1}$.*

3.1.9 Theorem. *Let $f : [a, b] \subset \mathbb{R} \longrightarrow \mathbb{R}$ and $g : [c, d] \subset \mathbb{R} \longrightarrow \mathbb{R}$ be differentiable at $x \in (a, b)$ and $f(x) \in (c, d)$, respectively. Then $g \circ f$ is differentiable at x and*

$$(g \circ f)'(x) = g'\big(f(x)\big)f'(x)$$

3.1.10 Example. Take for granted that sin and cos are differentiable functions with $\sin'(x) = \cos(x)$, $|\sin(x)| \leq 1$, and $|\cos(x)| \leq 1$ for all $x \in \mathbb{R}$. Investigate the differentiability of the following functions:

$$f(x) = \begin{cases} x \sin\left(\frac{1}{x}\right) & x \neq 0, \\ 0 & x = 0, \end{cases} \qquad g(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & x \neq 0, \\ 0 & x = 0. \end{cases}$$

3.2 Local maxima and minima and the mean value theorem

3.2.1 Definition. A function $f : [a, b] \longrightarrow \mathbb{R}$ has a *local maximum* at $x \in [a, b]$ if, for some $\varepsilon > 0$,

$$0 < |x - y| < \varepsilon \quad \text{implies} \quad f(y) \leq f(x).$$

Likewise, f has a *local minimum* at $x \in [a, b]$ if for some $\varepsilon > 0$,

$$0 < |x - y| < \varepsilon \quad \text{implies} \quad f(y) \geq f(x).$$

We say the maximum or minimum is *strict* if strict inequality holds, i.e., $f(y) < f(x)$ for a strict local maximum or $f(y) > f(x)$ for a strict local minimum. We say f has a *local extremum* at x if it has a local maximum or a local minimum at x.

3.2.2 Theorem. *Let $f : [a, b] \longrightarrow \mathbb{R}$ have a local extremum at $x \in (a, b)$. If f is differentiable at x, then $f'(x) = 0$.*

3.2.3 Example. Give examples which show that

(i) f may have a local extremum at x without $f'(x)$ being defined.

(ii) $f'(x) = 0$ does not imply that x is a local extremum.

3.2.4 Theorem (Mean Value Theorem)**.** *Let $f : [a, b] \longrightarrow \mathbb{R}$ be continuous and differentiable on (a, b). Then there exists $x \in (a, b)$ such that*

$$f(b) - f(a) = (b - a)f'(x).$$

Before proving the mean value theorem, you may wish to prove the following special case, known as *Rolle's Theorem*:

3.2.5 Lemma (Rolle's Theorem)**.** *Let $f : [a, b] \longrightarrow \mathbb{R}$ be continuous and differentiable on (a, b) with $f(a) = f(b)$. Then there exists $x \in (a, b)$ such that $f'(x) = 0$.*

Proof hint. Recall that $[a, b]$ is compact and connected. □

3.2.6 Definition. We say $f : [a, b] \longrightarrow \mathbb{R}$ is *monotone increasing* if

$$x < y \quad \text{implies} \quad f(x) \leq f(y).$$

Likewise we say f is *monotone decreasing* if

$$x < y \quad \text{implies} \quad f(x) \geq f(y).$$

We say f is *strictly monotone increasing* or *strictly monotone decreasing*, respectively, if strict inequality holds (i.e., $f(x) < f(y)$ or $f(x) > f(y)$, respectively).

Finally, we say f is *monotone* (resp. *strictly monotone*) if f is either increasing or decreasing (resp. strictly monotone increasing or strictly monotone decreasing).

3.2.7 Theorem. *Let $f : [a, b] \longrightarrow \mathbb{R}$ be continuous and differentiable on (a, b).*

(i) *If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.*

(ii) *If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is monotone increasing.*

(iii) *If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is monotone decreasing.*

(Moreover, if strict inequality holds, then f is strictly increasing or decreasing).

3.2.8 Theorem (Inverse Function Theorem)**.** *Suppose $f : (a, b) \longrightarrow \mathbb{R}$ is differentiable and either $f'(x) > 0$ for all $x \in (a, b)$ or $f'(x) < 0$ for all $x \in (a, b)$. Then f is a bijection onto its range, say $(c, d) = f\big((a, b)\big)$, and the inverse function $f^{-1} : (c, d) \longrightarrow (a, b)$ is differentiable on (c, d) with*

$$(f^{-1})'(y) = \frac{1}{f'(x)}, \quad \text{where } y = f(x).$$

3.3 Integration

3.3.1 Definition. A *partition* of $[a, b] \subset \mathbb{R}$ is a finite set $P = \{x_i : i = 0, \dots, N\}$ of points such that $a = x_0 < x_1 < \cdots < x_N = b$. We say a partition Q is a *refinement* of P if $P \subseteq Q$. Notice that for any two partitions P_1 and P_2, there is a partition Q which is a refinement of both, namely $Q = P_1 \cup P_2$.

3.3.2 Definition. Let $f : [a, b] \longrightarrow \mathbb{R}$ be a bounded function and $P = \{x_i : i = 0, \dots, N\}$ a partition of $[a, b]$. We define the *lower sum* $L(P, f)$ and *upper sum* $U(P, f)$ by

$$L(P, f) = \sum_{i=1}^{N} \alpha_i(x_i - x_{i-1}), \quad \alpha_i = \inf\{f(x) : x_{i-1} \leq x \leq x_i\}, \quad \text{and}$$

$$U(P, f) = \sum_{i=1}^{N} \beta_i(x_i - x_{i-1}), \quad \beta_i = \sup\{f(x) : x_{i-1} \leq x \leq x_i\},$$

respectively.

3.3.3 Example. Let P be a partition of $[a, b]$.

(i) Compute the lower and upper sums for the constant function $f(x) = c$.

(ii) Compute the lower and upper sums for the function $f(x) = x$.

Notice that the lower sums of a function generally increase under refinement while the upper sums decrease. Indeed, we have:

3.3.4 Proposition. *Let $f : [a, b] \longrightarrow \mathbb{R}$ be a bounded function and let Q be a refinement of P. Then*

$$L(P, f) \leq L(Q, f) \leq U(Q, f) \leq U(P, f).$$

3.3.5 Definition. Let $f : [a, b] \longrightarrow \mathbb{R}$ be a function. Define the *lower integral* of f on $[a, b]$ by

$$\underline{\int_a^b} f(x)\,dx = \sup\left\{L(P, f) : P \text{ is a partition of } [a, b]\right\},$$

and the *upper integral* of f by

$$\overline{\int_a^b} f(x)\,dx = \inf\left\{U(P, f) : P \text{ is a partition of } [a, b]\right\}.$$

If $\underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx$, then we say f is *integrable* on $[a, b]$ and we define its (Riemann) *integral* by

$$\int_a^b f(x)\,dx = \underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx.$$

3.3.6 Proposition.

(i) *For any two partitions P and Q of $[a, b]$ (not necessarily related by refinement),*

$$L(P, f) \leq U(Q, f).$$

(ii)

$$\underline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f(x)\,dx.$$

3.3.7 Proposition. *f is integrable on $[a, b]$ if and only if, for every $\varepsilon > 0$, there exists a partition P of $[a, b]$ such that*

$$U(P, f) - L(P, f) < \varepsilon.$$

3.3.8 Example.

(i) Show that $f(x) = c$ is integrable on $[a, b]$ with $\int_a^b c\,dx = c(b - a)$.

(ii) Show that $f(x) = x$ is integrable on $[a, b]$ with $\int_a^b x\,dx = \frac{1}{2}(b^2 - a^2)$.

(iii) Show that the function $f : [0, 1] \longrightarrow \mathbb{R}$,

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q}, \\ 0 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

is not integrable.

Next we prove that certain nice classes of functions are always integrable.

3.3.9 Theorem. *Every bounded monotone function on* $[a, b]$ *is integrable.*

Proof hint: Consider $U(P, f) - L(P, f)$ for partitions P into fixed-size intervals (i.e., $x_i - x_{i-1} = \delta$ for all i for some fixed δ). □

We prove below that continuous functions are integrable, but before doing so it is useful to consider a slightly stronger notion of continuity. Recall that $f : A \subset \mathbb{R} \longrightarrow \mathbb{R}$ is continuous if and only if

$$\text{for all } x \in A \text{ and all } \varepsilon > 0, \text{ there exists } \delta > 0 \text{ such that}$$
$$y \in A, \quad |y - x| < \delta \quad \implies \quad |f(y) - f(x)| < \varepsilon.$$

In particular, δ may depend on both ε and x (but not y).

3.3.10 Definition. A function $f : A \subset \mathbb{R} \longrightarrow \mathbb{R}$ is *uniformly continuous* if

$$\text{for all } \varepsilon > 0, \text{ there exists } \delta > 0 \text{ such that}$$
$$x, y \in A, \quad |y - x| < \delta \quad \implies \quad |f(y) - f(x)| < \varepsilon.$$

In particular, while δ may depend on ε, it must be independent of x and y.

Remark. The definition of uniform continuity makes sense in the general setting of functions between two metric spaces.

3.3.11 Example. The function $f : \mathbb{R} \longrightarrow \mathbb{R}$, $f(x) = x^2$ is not uniformly continuous.

3.3.12 Proposition. *A continuous function on a compact set is uniformly continuous; in particular, every continuous* $f : [a, b] \longrightarrow \mathbb{R}$ *is uniformly continuous.*

3.3.13 Theorem. *Every continuous function* $f : [a, b] \longrightarrow \mathbb{R}$ *is integrable on* $[a, b]$*.*

We can improve on Theorem 3.3.13 for f which is only piecewise continuous.

3.3.14 Theorem. *If a bounded function* $f : [a, b] \longrightarrow \mathbb{R}$ *is continuous except at finitely many points, then it is integrable.*

In fact, we can do even better:

3.3.15 Definition. Say a set $A \subset \mathbb{R}$ has *measure zero* if it be covered by finitely or countably many open intervals of arbitrarily small total width, i.e., for every $\varepsilon > 0$, there are open intervals $\{I_j = (a_j, b_j) : j \in J\}$ with J finite or countable such that

(i) $A \subset \bigcup_{j \in J} I_j$, and

(ii) $\sum_{j \in J} |I_j| < \varepsilon$, where $|I_j| := b_j - a_j$.

3.3.16 Example. Show the following sets have measure zero in $\mathbb{R}$:

(i) Any finite set $\{x_1, \ldots, x_N\}$.

(ii) $[0, 1] \cap \mathbb{Q}$.

(iii) Any countable set $\{x_n : n \in \mathbb{N}\}$.

(iv) The Cantor middle thirds set from Example 2.6.16.(i). (In fact this set has uncountably many points, so it does not fall under the previous example.)

3.3.17 Example.

(i) Consider the function $f : [0,1] \longrightarrow \mathbb{R}$

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q}, \\ 0 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

from Example 3.3.8.(iii). Where is f continuous?

(ii) Now define $g : [0,1] \longrightarrow \mathbb{R}$ by

$$g(x) = \begin{cases} \frac{1}{q} & x = \frac{p}{q},\ p, q \in \mathbb{N},\ \gcd(p,q) = 1, \\ 0 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

Show that g is integrable. On what set is g continuous?

3.3.18 Theorem† (Riemann-Lebesgue Theorem)**.** *$f : [a,b] \longrightarrow \mathbb{R}$ is Riemann integrable if and only if f is bounded and continuous off of a set of measure zero.*

Remark. It is common terminology to say that a property holds *almost everywhere* if it holds on the complement of a set of measure zero. Thus, in the previous theorem we could say that a function is integrable if and only if it is bounded and continuous almost everywhere.

3.3.19 Theorem. *The Riemann integral has the following properties:*

(i) *If f and g are integrable on $[a,b]$ then so is $cf + dg$ where c and d are constants, and*

$$\int_a^b (c\,f(x) + d\,g(x))\,dx = c\int_a^b f(x)\,dx + d\int_a^b g(x)\,dx.$$

(ii) *If f is integrable on $[a,b]$ and $[b,c]$, then it is integrable on $[a,c]$ and*

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx.$$

(iii) *If f and g are integrable on $[a,b]$ and $f(x) \leq g(x)$ for all $x \in [a,b]$, then*

$$\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx.$$

(iv) *If f is integrable on $[a,b]$ and $|f(x)| \leq M$ for all $x \in [a,b]$, then*

$$\left|\int_a^b f(x)\,dx\right| \leq M(b-a).$$

3.3.20 Theorem (Fundamental Theorem of Calculus)**.**

(i) *Let $f : [a,b] \longrightarrow \mathbb{R}$ be continuous, and for $x \in [a,b]$ define $F(x) = \int_a^x f(t)\,dt$. Then F is differentiable on (a,b) and $F'(x) = f(x)$.*

(ii) *If g is a real-valued function which is differentiable on $[a,b]$ and if g' is continuous on $[a,b]$, then*

$$\int_a^b g'(x)\,dx = g(b) - g(a).$$

Chapter 4

Sequences and series of functions

4.1 Pointwise and uniform convergence

In this section we investigate various modes of convergence for sequences of functions. If a sequence (f_n) of functions converges to a function f under some suitable definition, we would like to know how the limit behaves with respect to various properties and operations, such as continuity, limits, derivatives, and integrals.

4.1.1 Definition. Let (M, d) and (N, d') be metric spaces. A sequence (f_k) of functions $f_k : A \subseteq M \longrightarrow N$ is said to *converge pointwise* to $f : A \longrightarrow N$ if, for each $x \in A$, the sequence $\big(f_k(x)\big)$ of points in N converges to the point $f(x)$. More precisely, for every $x \in A$ and every $\varepsilon > 0$, there exists a natural number K (depending on x and ε) such that

$$d'\big(f_k(x), f(x)\big) < \varepsilon \qquad \text{for all } k \geq K.$$

4.1.2 Example. Find the pointwise limit of the following sequences, if one exists.

(i) $f_k : \mathbb{R} \longrightarrow \mathbb{R}$, $f_k(x) = \chi(x - k)$, where χ is the indicator function of the unit inverval:

$$\chi(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & x \notin [0, 1] \end{cases}$$

(ii) $g_k : \mathbb{R} \longrightarrow \mathbb{R}$, $g_k(x) = \frac{1}{k}\chi(x)$ with χ as above.

(iii) $h_k : [0, 1] \longrightarrow \mathbb{R}$, $h_k(x) = x^k$.

(iv) $e_k : [0, 1] \longrightarrow \mathbb{R}$, $e_k(x) = (-1)^k x^k$.

The preceding examples demonstrate that some undesirable things can happen with pointwise limits. For example, in (i), all of the functions including the limit are integrable on $\mathbb{R}$, but

$$\lim_{k\to\infty} \int_{-\infty}^{\infty} f_k(x)\,dx \neq \int_{-\infty}^{\infty} \lim_{k\to\infty} f_k(x)\,dx.$$

Likewise, in (iii),

$$\lim_{k\to\infty} \lim_{x\to 1} h_k(x) \neq \lim_{x\to 1} \lim_{k\to\infty} h_k(x),$$

and in particular this pointwise limit of a sequence of continuous functions is not continuous. To address these issues we can require a stronger notion of convergence, called *uniform convergence*.

4.1.3 Definition. A sequence (f_k) of functions $f_k : A \subseteq M \longrightarrow N$ is said to *converge uniformly* to $f : A \longrightarrow N$ if, for every $\varepsilon > 0$, there exists a natural number K such that

$$d'\big(f_k(x), f(x)\big) < \varepsilon \quad \text{for all } k \geq K \text{ and } x \in A.$$

Remark. Compare Definitions 4.1.3 and 4.1.1 carefully. For pointwise convergence, K is allowed to depend on both x and ε, while for uniform convergence, K may only depend on ε, and must be chosen "uniformly", indepdenent of x. Observe that a uniformly convergent sequence is pointwise convergent.

As Example 4.1.6 below shows, it is very important to specify the domain A in Definition 4.1.3 when discussing uniform convergence.

4.1.4 Theorem. *The limit of a uniformly convergent sequence of continuous functions is continuous. More precisely, if $f_k : A \longrightarrow N$ are continuous at any $x \in A$ and $f_k \to f$ uniformly, then f is continuous at $x \in A$.*

Remark. Recalling that f is continuous at x if and only if $f(x) = \lim_{y\to x} f(y)$, this can be interpreted as a statement about interchange of limits: if $f_k \to f$ uniformly, then $\lim_{y\to x} \lim_{k\to\infty} f_k(y) = \lim_{k\to\infty} \lim_{y\to x} f_k(y)$.

4.1.5 Example. Which sequences in Example 4.1.2 converge uniformly?

4.1.6 Example. The sequence $f_k(x) = \frac{1+x^k}{1-x}$ converges pointwise on $(-1, 1)$ to the function $f(x) = \frac{1}{1-x}$. Show that this convergence is uniform on $[-a, a]$ for any fixed $a < 1$. Is it uniform on $(-1, 1)$?

Since we often take the range space N to be a complete space such as $\mathbb{R}$ or $\mathbb{R}^n$, we can leverage completeness to characterize uniform convergence in terms of a Cauchy property.

4.1.7 Theorem. *Let (M, d) and (N, d') be metric spaces, and suppose that N is complete. Then a sequence $f_k : A \subseteq M \longrightarrow N$ of functions is uniformly convergent if and only if it is* uniformly Cauchy*, meaning that for every $\varepsilon > 0$ there exists some $K \in \mathbb{N}$ such that*

$$d'\big(f_k(x), f_\ell(x)\big) < \varepsilon \quad \textit{for all } k, \ell \geq K \textit{ and all } x \in A.$$

4.1.8 Theorem. *Let $f_k : [a, b] \longrightarrow \mathbb{R}$ be a sequence of Riemann integrable functions. If (f_k) is uniformly convergent on $[a, b]$, then $f = \lim_{k\to\infty} f_k$ is integrable and*

$$\lim_{k\to\infty} \int_a^b f_k(x)\, dx = \int_a^b \lim_{k\to\infty} f_k(x)\, dx.$$

One might hope that a similar result holds for derivatives; i.e., if f_k are differentiable at x and $f_k \to f$ uniformly, then f is differentiable at x and $f_k'(x) \to f'(x)$, but this is not generally true.

4.1.9 Example*. Accepting standard facts about the sine function (to be proved later on), let $f_k(x) = \frac{1}{k}\sin(kx)$. Show that each f_k is differentiable and the sequence converges uniformly, but the sequence (f_k') is ill-behaved.

4.1.10 Theorem. *Let $f_k : [a, b] \longrightarrow \mathbb{R}$ be a uniformly convergent sequence of continuously differentiable functions (so each f_k' is continuous on $[a, b]$) with limit f. If the sequence (f_k') of derivative functions converges uniformly to a function g, then f is differentiable and $f' = g$.*

Proof hint. Use Theorem 4.1.8 and the fundamental theorem of calculus. □

Remark. The hypothesis of continuity of the f_k' can be removed, but the proof becomes much more technical.

4.1.11 Example. Let $f_k(x) = (x - 1/k)^2$ on $[0, 1]$. Does f_k converge uniformly? Does Theorem 4.1.8 apply?

4.1.12 Problem*. Let $f_k(x) = x/k$. Does f_k converge uniformly on $[0, 1]$? How about on $\mathbb{R}$?

4.1.13 Problem. Give a proof or find a counterexample:

(a) If $f_k \to f$ and $g_k \to g$ converge uniformly, then $f_k + g_k$ converges uniformly to $f + g$.

(b) If $f_k \to f$ and $g_k \to g$ converge uniformly, then $f_k g_k$ converges uniformly to fg.

4.2 Series of functions

If (f_k) is a sequence of functions whose range is $N = \mathbb{R}$ (or more generally a vector space equipped with a norm such as $\mathbb{R}^n$), then we can speak of the series $\sum_k f_k$.

4.2.1 Definition. We say that a series $\sum_{k=0}^\infty f_k$, where $f_k : A \subseteq M \longrightarrow \mathbb{R}^n$, *converges pointwise* (respectively *uniformly*) to $f : A \longrightarrow \mathbb{R}^n$ provided the sequence partial sums $s_n = \sum_{k=0}^n f_k$ converge pointwise (resp. uniformly) to f.

In light of completeness of $\mathbb{R}^n$, Theorem 4.1.7 takes the following form for series (c.f. Theorem 1.6.4).

4.2.2 Corollary (Uniform Cauchy criterion for series)**.** *A series $\sum f_k$ of functions $f_k : A \subseteq M \longrightarrow \mathbb{R}^n$ converges uniformly if and only if, for every $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that*

$$\left\| \sum_{k=m}^n f_k(x) \right\| < \varepsilon \quad \text{for all } n \geq m \geq K \text{ and all } x \in A.$$

Since finite sums of continuous (resp. integrable) functions are continuous (resp. integrable), we obtain the following two theorems from their sequential counterparts (Theorems 4.1.4 and 4.1.8).

4.2.3 Theorem*. *Uniformly convergent series of continuous functions are continuous. In other words, if for each k the function $f_k : A \subseteq M \longrightarrow \mathbb{R}^n$ is continuous and $\sum_{k=0}^\infty f_k$ converges uniformly to $f : A \longrightarrow \mathbb{R}^n$, then f is continuous.*

4.2.4 Theorem. *Uniformly convergent series of integrable functions are integrable. In other words, if for each k the function $f_k : [a,b] \longrightarrow \mathbb{R}$ is integrable and $\sum_{k=0}^\infty f_k$ converges uniformly to $f : [a,b] \longrightarrow \mathbb{R}$, then f is integrable and*

$$\int_a^b f(x)\,dx = \sum_{k=0}^\infty \int_a^b f_k(x)\,dx.$$

To determine when a series of functions converges uniformly, the next result is often useful.

4.2.5 Theorem (Weierstrass M-test)**.** *Suppose $\sum_{k=0}^\infty M_k$ is a convergent series such that $M_k \geq 0$ for all k. If $f_k : A \subseteq M \longrightarrow \mathbb{R}^n$ is a sequence of functions satisfying $\|f_k(x)\| \leq M_k$ for all $x \in A$, then the series $\sum_{k=0}^\infty f_k$ converges uniformly (and absolutely) to a function $f : A \longrightarrow \mathbb{R}^n$.*

4.2.6 Example. Show that $\sum_{n=1}^\infty \frac{x^n}{n^2 2^n}$ converges to a continuous function on $[-2,2]$. Compute $\int_{-2}^2 f(x)\,dx$, where $f(x)$ is the limit of the series.

4.2.7 Problem. Consider the series $\sum_{n=0}^\infty \frac{x^n}{1+x^n}$. Show that this series converges uniformly on $[0,a]$ for every $a < 1$. Does it converge uniformly on $[0,1)$?

4.3 Power series

4.3.1 Definition. A *power series* is a series of the form $\sum_{n=0}^\infty a_n x^n$ on $\mathbb{R}$, where (a_n) is a sequence of real numbers, which we call the *coefficients* of the power series. More generally, a power series *centered at* $c \in \mathbb{R}$ is a series of the form $\sum_{n=0}^\infty a_n (x-c)^n$.

Remark. Observe that the power series $\sum a_n x^n$ converges on an interval from a to b if and only if the power series $\sum a_n (x-c)^n$ converges on the interval from $a+c$ to $b+c$, so we can prove convergence results for power series assuming $c = 0$ without loss of generality.

4.3.2 Definition. The *radius of convergence* of a power series $\sum a_n x^n$ is the number $R = \frac{1}{\alpha}$, where $\alpha = \limsup_n |a_n|^{1/n}$. If $\alpha = +\infty$, then $R = 0$, and if $\alpha = 0$, then $R = +\infty$.

The next result justifies the name. In light of Theorem 1.6.12, if $\beta = \lim_{n\to\infty} |a_{n+1}| / |a_n|$ exists (which is often easier to compute), then $R = \beta^{-1}$.

4.3.3 Theorem. *Let $\sum a_n x^n$ be a power series with positive radius of convergence $R > 0$. Then it converges to a continuous function on $(-R, R)$, and converges uniformly on every interval $[-R+\varepsilon, R-\varepsilon]$ where $\varepsilon > 0$.*

Remark. The power series may or may not converge uniformly on $(-R, R)$, and it may or may not converge at all at (either or both of) the points $\pm R$.

4.3.4 Example*. Determine the radius of convergence of the following power series, and whether or not they converge at $x = \pm R$.

$$\text{(i)} \sum \frac{x^n}{n!} \qquad \text{(ii)} \sum n x^n \qquad \text{(iii)} \sum \frac{x^n}{n} \qquad \text{(iv)} \sum \frac{x^n}{n^2}$$

Despite the general difficulty of differentiating uniformly convergent sequences (c.f. Example 4.1.9), it turns out that power series can be both integrated *and* differentiated term by term.

4.3.5 Theorem. *Let $f(x) = \sum a_n x^n$ be a power series with radius of convergence R. Then for $|x| < R$,*

$$\int_0^x f(t)\,dt = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}.$$

4.3.6 Theorem. *Let $f(x) = \sum a_n x^n$ be a power series with radius of convergence R. Then f is differentiable on $(-R, R)$ and*

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

Proof hint. Show $\sum n a_n x^{n-1}$ has the same radius of convergence and apply Theorem 4.3.5. $\square$

4.3.7 Corollary. *If $f(x) = \sum a_n x^n$, has radius of convergence $R > 0$, then f has derivatives of all orders on $(-R, R)$, with*

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1) a_n x^{n-k}$$

In particular, $f^{(k)}(0) = k!\, a_k$, so the coefficients are determined by the derivatives of f at 0 by

$$a_k = \frac{f^{(k)}(0)}{k!}.$$

4.3.8 Problem. Prove that $\sum_{n=1}^{\infty} n x^n = \frac{x}{(1-x)^2}$ for $|x| < 1$. Use this to evaluate the following numerical series

(i) $\sum_{n=1}^{\infty} \frac{n}{2^n}$.

(ii) $\sum_{n=1}^{\infty} \frac{(-1)^n n}{3^n}$.

4.4 Trancendental functions

4.4.1 Problem. Define $E(x) = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$ and define $e = E(1) \in \mathbb{R}$.

(i) Show that E is a continuous real-valued function on all of $\mathbb{R}$.

(ii) Show that $E(x+y) = E(x)E(y)$ and $E(0) = 1$; in particular $E(-x) = E(x)^{-1}$.

(iii) Show that $E(x) > 0$ for all x.

(iv) Show that $E'(x) = E(x)$; in particular E is everywhere differentiable to all orders and strictly increasing.

(v) Using (ii), conclude that $E(nx) = E(x)^n$ for all $n \in \mathbb{Z}$ and $E(x/n) = E(x)^{1/n}$ for all $n \in \mathbb{N}$, and generally that $E(ax/b) = E(x)^{a/b}$ for every $\frac{a}{b} \in \mathbb{Q}$. In particular $E(a/b) = e^{a/b}$. This justifies the notation/definition $e^x = E(x)$ for arbitrary $x \in \mathbb{R}$.

4.4.2 Problem. Appeal to the inverse function theorem to define a function $L : (0, \infty) \longrightarrow \mathbb{R}$ such that $L\big(E(x)\big) = x$ and $E\big(L(x)\big) = x$.

(i) Show that $L'(x) = \frac{1}{x}$.

(ii) Show that $L(1) = 0$, that $L(xy) = L(x) + L(y)$, and that $L(x^{-1}) = -L(x)$.

4.4.3 Problem.

(i) Using the above functions, show that $x^a = E\big(aL(x)\big)$ for all $x > 0$ and $a \in \mathbb{Q}$.

(ii) Thus for general $x > 0$ and $a \in \mathbb{R}$, we may define

$$x^a = E\big(aL(x)\big),$$

extending the usual definition when $a \in \mathbb{Q}$. Show that this satisfies $x^{a+b} = x^a x^b$, and $x^0 = 1$.

4.4.4 Problem. Define $S : \mathbb{R} \longrightarrow \mathbb{R}$ and $C : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$S(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}, \qquad C(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}.$$

Show that $S'(x) = C(x)$ and $C'(x) = -S(x)$.

Remark. Using complex-valued series and functions, one can show that $C(x) = \frac{1}{2}\big(E(ix) + E(-ix)\big)$ and $S(x) = \frac{1}{2i}\big(E(ix) - E(-ix)\big)$, or equivalently that $E(ix) = C(x) + iS(x)$. The easily verified fact that $\overline{E(z)} = E(\overline{z})$ (here if $z = x + iy$ then $\overline{z} = x - iy$ denotes the complex conjugate) along with Problem 4.4.1.(ii) shows that $|E(ix)|^2 = E(ix)E(-ix) = 1$, from which it follows that $|S(x)| \leq 1$ and $|C(x)| \leq 1$ and

$$S^2(x) + C^2(x) = 1.$$

The equation $|E(ix)| = 1$ along with $E'(ix) = iE(ix)$ and $|E'(ix)| = 1$ show that $x \longmapsto E(ix)$ traces out the unit circle in the complex plane at unit speed. We define the real number π as the smallest positive number such that $E(i2\pi) = E(0) = 1$. In particular 2π is the circumference of the unit circle. This provides the connection between the functions S and C defined above by power series with the usual trigonometric definitions of sine and cosine.

4.5 The space of continuous functions

In this section we will reinterpret what it means for functions to converge uniformly.

4.5.1 Definition. Let (M, d) be a metric space and $A \subseteq M$. Let

$$\mathcal{B}(A) = \{f : A \longrightarrow \mathbb{R} : \sup\{|f(x)| : x \in A\} < \infty\}$$

denote the set of bounded real valued functions on A, and denote by

$$\mathcal{C}(A) = \{f : f \text{ continuous}\} \subset \mathcal{B}(A)$$

the subset of (bounded) continuous functions. For $f \in \mathcal{B}(A)$ define the *sup norm* by

$$\|f\| = \sup\{|f(x)| : x \in A\} \in [0, \infty).$$

4.5.2 Problem. Show that $\mathcal{B}(A)$ is a vector space over $\mathbb{R}$. More precisely, there are addition and scalar multiplication operations $+ : \mathcal{B}(A) \times \mathcal{B}(A) \longrightarrow \mathcal{B}(A)$ and $\cdot : \mathbb{R} \times \mathcal{B}(A) \longrightarrow \mathcal{B}(A)$ the first of which is associative and commutative, with an identity element and inverses; and the second of which is associative, distributes over addition, and satisfies $1 \cdot f = f$.

Show that $\mathcal{C}(A)$ is a subspace.

4.5.3 Proposition. *The sup norm is a norm, i.e.,*

(i) $\|f\| = 0$ *if and only if* $f \equiv 0$,

(ii) $\|af\| = |a|\,\|f\|$ *for all* $a \in \mathbb{R}$, $f \in \mathcal{B}(A)$,

(iii) $\|f + g\| \leq \|f\| + \|g\|$.

In particular,

$$d_{\sup}(f, g) = \|f - g\| = \sup\{|f(x) - g(x)| : x \in A\}$$

is a metric on $\mathcal{B}(A)$ *and* $\mathcal{C}(A)$.

4.5.4 Proposition. *A sequence* (f_k) *converges with respect to* $d_{\sup}$ *if and only if it converges uniformly.*

4.5.5 Proposition. *The metric spaces* $(\mathcal{B}(A), d_{\sup})$ *and* $(\mathcal{C}(A), d_{\sup})$ *are complete.*

4.6.6 Proposition. *For every* $[a, b] \subset \mathbb{R}$, *integration defines a map* $I : \mathcal{C}([a, b]) \longrightarrow \mathbb{R}$,

$$I(f) = \int_a^b f(x)\,dx,$$

which is continuous with respect to the metric $d_{\sup}$ *on* $\mathcal{C}([a, b])$.

4.6 Compactness in $\mathcal{C}(A)$: the Arzela-Ascoli theorem

Given a sequence (f_k) of real-valued functions on a domain A, it is often of interest to know under what circumstances it will have a (uniformly) convergent subsequence. Recall that this is related to *compactness* (c.f. §2.6): by the Bolzano-Weierstrass Theorem 2.6.8, such a sequence must have a convergent subsequence if (and, by Theorem 2.6.9, only if) it lies in a compact set of $\mathcal{C}(A)$. In this section we prove the Arzela-Ascoli theorem, which identifies compact subsets of $\mathcal{C}(A)$ (just like the Heine-Borel Theorem 2.6.20 identifies compact subsets of $\mathbb{R}^n$).

In order to motivate the conditions in the theorem, we first investigate several examples to see what can go wrong.

4.6.1 Example. Show that the following sequences of bounded continuous functions cannot have any uniformly convergent subsequence.

(i) $f_k(x) = kx(1 - x)$ for $x \in [0, 1]$.

(ii) $g_k(x) = \chi(x - k)$ for $x \in \mathbb{R}$, where $\chi(x) = \begin{cases} x(1-x) & \text{for } x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$

(iii) $h_k(x) = x^k$ for $x \in [0, 1)$.

(iv) $e_k(x) = \sin(2\pi kx)$ for $x \in [0, 1]$.

The first example shows that, in order to have a convergent subsequence, a sequence of functions shouldn't be unbounded. The second and third examples show that compactness of the domain is a reasonable requirement. Finally, the last example illustrates that, even if the sequence is bounded on a compact domain, some additional way of controlling the sequence is necessary.

4.6.2 Definition. A subset $F \subset \mathcal{C}(A)$ of continuous functions is said to be *bounded* if there exists $B \geq 0$ such that $\|f\| \leq B$ for all $f \in F$.

The set is said to be *equicontinuous* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d(x, y) < \delta \quad \text{implies} \quad |f(x) - f(y)| < \varepsilon, \quad \text{for every } f \in F.$$

In particular, the chosen δ must work for all $f \in F$ at the same time.

4.6.3 Theorem (Arzela-Ascoli, sequential version)**.** *Let A be a compact set in a metric space (M, d), and suppose (f_k) is a sequence of functions $f_k : A \longrightarrow \mathbb{R}$ which is bounded and equicontinuous. Then (f_k) has a uniformly convergent subsequence.*

The proof uses a few key observations, which we break out as the following three lemmas. In each one, assume the hypotheses of Theorem 4.6.3.

4.6.4 Lemma. *As a compact set, A has a countable dense subset $E \subset A$. In other words, for every $\varepsilon > 0$ and $x \in A$, there is some $y \in E$ such that $d(x, y) < \varepsilon$.*

Proof hint. For every $n \in \mathbb{N}$, there exists a finite set $E_n = \{x_1, \dots, x_N\}$ such that A is covered by the open $1/n$-balls $D_{1/n}(x_1), \dots, D_{1/n}(x_N)$. □

4.6.5 Lemma. *There exists a subsequence (f_{k_n}) which converges at every point of the countable dense set E. In other words, if $\{x_1, x_2, \dots\} = E$ is an enumeration of E, then $f_{k_n}(x_i)$ converges for each i.*

Proof hint. There is a subsequence that converges at x_1, a further subsequence that converges at x_2 and so on. Then think diagonally. □

4.6.6 Lemma. *The subsequence (f_{k_n}) of Lemma 4.6.5 is uniformly Cauchy.*

While the sequential version above is the one most often used in applications, the following (non-sequential) version is straightforward to conclude from the Bolzano-Weierstrass Theorem.

4.6.7 Theorem (Arzela-Ascoli)**.** *Let $F \subset \mathcal{C}(A)$ be a closed*[1]*, bounded, and equicontinuous subset of the continuous functions on a compact set A. Then F is a compact subspace of $\mathcal{C}(A)$.*

Remark. In fact, it is also possible to show that if every sequence in F has a uniformly convergent subsequence, then F is bounded and equicontinuous, so that the converse of Theorem 4.6.3 also holds; in other words $F \subset \mathcal{C}(A)$ is compact if and only if it is closed, bounded, and equicontinuous.

This is in contrast to the finite dimensional vector space $\mathbb{R}^n$, in which the Heine-Borel theorem says that compact sets are those which are merely closed and bounded. In the infinite dimensional setting of $\mathcal{C}(A)$ closed and bounded isn't enough. (c.f. Example 4.6.1.(iv)).

In practice stronger conditions are typically used to conclude equicontinuity. The most common one is a bound on the derivatives of a set of differentiable functions.

4.6.8 Proposition. *Let F be a set of differentiable functions on $[a, b]$ having continuous derivatives, and suppose that there exists $B \geq 0$ such that*

$$\|f'\| = \sup \{|f'(x)| : x \in [a, b]\} \leq B \quad \textit{for every } f \in F.$$

Then F is equicontinuous.

In practice, many problems in geometry or differential equations can be posed as optimization problems with respect to some kind of real-valued function defined on a function space. (Such an object is usually referred to as a *functional*, to avoid confusion.) Compactness results like Arzela-Ascoli can lead to existence results for such problems. Here is a very simple example.

4.6.9 Problem*. Let $F \subset \mathcal{C}([a, b])$ be a set of continuous functions which is uniformly closed, bounded, and equicontinuous. Then there exists $f \in F$ such that

$$\int_a^b f(x)\, dx = \sup\left\{ \int_a^b g(x)\, dx : g \in F \right\}.$$

[1]meaning closed with respect to the uniform metric $d_{\sup}$.

4.7 Density in $\mathcal{C}([a,b])$: Weierstrass approximation

Continuing the analogy between $\mathbb{R}^n$ and $\mathcal{C}(A)$ for compact A, a natural question to ask is, what plays the role of $\mathbb{Q}^n \subset \mathbb{R}^n$? Indeed, it is often useful to know that we can approximate a real vector arbitrarily closely by a rational one, the latter being easier to deal with from both a conceptual as well as computational point of view. Thus we are interested in finding a dense subspace of $\mathcal{C}(A)$, perhaps consisting of functions which are especially "nice" in some way. When $A = [a,b] \subset \mathbb{R}$, there is a particularly nice answer due to Weierstrass.

4.7.1 Theorem (Weierstrass approximation)**.** *Every continuous function on $[a,b]$ can be uniformly approximated by polynomials. In other words, given any $f \in \mathcal{C}([a,b])$, for every $\varepsilon > 0$ there exists a polynomial $P(x) = p_0 + p_1 x + \cdots p_k x^k$ such that*

$$\|P - f\| = \sup\{|P(x) - f(x)| : x \in [a,b]\} < \varepsilon.$$

The proof is somewhat involved, so again we break it down into a series of lemmas. We focus on the case $[a,b] = [0,1]$, as the general case can be reduced to this.

4.7.2 Lemma. *For each $n \in \mathbb{N}$, let $Q_n(x) = c_n(1-x^2)^n$, where $c_n \in \mathbb{R}$ is a constant defined so that $\int_{-1}^1 Q_n(x)\,dx = 1$. Then $c_n < \sqrt{n}$.*

Proof hint. The inequality $(1-x^2)^n \geq 1 - nx^2$ holds on $[0,1]$ by examining the derivative of $(1-x^2)^n - 1 + nx^2$. Then estimate $\int_{-1}^1 (1-x^2)^n\,dx$ from below by $2\int_0^{1/\sqrt{n}} (1-nx^2)\,dx$. □

4.7.3 Lemma. *For every fixed $\delta > 0$, (Q_n) converges uniformly to 0 on $[-1,-\delta] \cup [\delta,1]$.*

4.7.4 Lemma. *Let $f \in \mathcal{C}([0,1])$ and regard it as a (not necessarily continuous) function on all of $\mathbb{R}$ by defining it to be 0 outside of $[0,1]$. Let*

$$P_n(x) = \int_{-1}^1 f(x+t)\,Q_n(t)\,dt \equiv \int_{-x}^{1-x} f(x+t)\,Q_n(t)\,dt.$$

Then P_n is a sequence of polynomials in x.

Proof hint. Show that $P_n(x) = \int_0^1 f(s)\,Q_n(x-s)\,ds$. □

4.7.5 Lemma. *The sequence P_n converges uniformly to f on $[0,1]$.*

Proof hint. $|P_n(x) - f(x)| = \left|\int_{-1}^1 \big(f(x+t) - f(x)\big)\,Q_n(t)\,dt\right|$. Recall that f is actually *uniformly continuous* on $[0,1]$ as it is compact. □

4.7.6 Problem*. Show that there exists a sequence (P_n) of polynomials with $P_n(0) = 0$ for all n such that

$$\sup\{|P_n(x) - |x|| : x \in [-1,1]\} \to 0 \quad \text{as } n \to \infty.$$

Chapter 5

Calculus of multivariable functions

5.1 The total derivative as a linear map

5.1.1 Definition. Recall that a map $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ (or more generally between two vector spaces over the same field) is *linear* if

$$T(x+y) = T(x) + T(y) \quad \text{and} \quad T(ax) = aT(x)$$

for all $x, y \in \mathbb{R}^n$, $a \in \mathbb{R}$. By convention we often omit parentheses, writing Tx instead of $T(x)$.

The set of all such linear maps is denoted $L(\mathbb{R}^n, \mathbb{R}^m)$, and is itself a vector space of dimension nm (check!). For $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ we define the *operator norm* $\|T\|$ by

$$\|T\| = \sup \{\|Tx\| : \|x\| = 1\}.$$

In particular for every $x \in \mathbb{R}^n$,

$$\|Tx\| \le \|T\| \, \|x\| .$$

5.1.2 Proposition. *The operator norm satisfies*

(i) $0 \le \|T\| < \infty$, *and* $\|T\| = 0$ *if and only if* T *is the zero map* ($Tx = 0$ *for every* x).

(ii) $\|aT\| = |a| \, \|T\|$.

(iii) $\|T + S\| \le \|T\| + \|S\|$.

In particular, $d(T, S) = \|T - S\|$ *is a metric on the space* $L(\mathbb{R}^n, \mathbb{R}^m)$.

We next want to define the notion of differentiability for functions $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. Note that the difference quotient, $\frac{f(x+h)-f(x)}{h}$, which we use when $n = m = 1$, does not make sense in general since the numerator is a vector in $\mathbb{R}^m$ and the denominator is a vector in $\mathbb{R}^n$, and in any case it does not make sense to divide vectors. Instead, we take inspiration from the fact that, in the single variable case, $f'(x) \in \mathbb{R}$ defines a linear map from $\mathbb{R}$ to $\mathbb{R}$, namely multiplication $h \longmapsto f'(x)h$. Indeed, this is how the derivative functions in the equivalent formulation

$$\lim_{h \to 0} \left| \frac{f(x+h) - f(x) - f'(x)h}{h} \right| = 0.$$

5.1.3 Definition. We say a function $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is *differentiable* at $x \in A$ if there exists a linear map $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - Th\|}{\|h\|} = 0. \tag{5.1}$$

In other words, for every $\varepsilon > 0$, there exists $\delta > 0$ such that $0 < \|h\| < \delta$ implies

$$\frac{\|f(x+h) - f(x) - Th\|}{\|h\|} < \varepsilon.$$

We write $Df(x) = T \in L(\mathbb{R}^n, \mathbb{R}^m)$ and call this the *total derivative of* f *at* x.

Remark. Note where all the elements in (5.1) live: x and h are vectors in $\mathbb{R}^n$, while $f(x+h)$, $f(x)$ and Th are vectors in $\mathbb{R}^m$. In particular, the norm in the numerator is the Euclidean norm on $\mathbb{R}^n$ while the norm in the denominator is the Euclidean norm in $\mathbb{R}^m$, and the derivative itself, $Df(x)$, is a linear map taking vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$.

The total derivative makes precise the notion of the *best (affine) linear approximation* to f at x, in the sense that the difference

$$f(x+h) - f(x) - Df(x)h = \phi_{f,x}(h)$$

defines a function $\phi_{f,x} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ which satisfies $\lim_{h\to 0} \frac{\|\phi_{f,x}(h)\|}{\|h\|} = 0$. Thus the affine linear function $h \longmapsto f(x) + Df(x)h$ is a good approximation to $f(x+h)$ for small h.

5.1.4 Proposition. *If the derivative $Df(x)$ exists, then it is unique.*

5.1.5 Proposition. *If f is differentiable at x then it is continuous there, i.e., $\lim_{y\to x} f(y) = f(x)$.*

5.1.6 Example. Identify the linear map $Df(x)$ for $f : \mathbb{R} \longrightarrow \mathbb{R}$, $f(x) = x^3$. In other words, how does this linear map act on the input vector $h \in \mathbb{R}^1$?

5.1.7 Example*. Let $\gamma : \mathbb{R} \longrightarrow \mathbb{R}^n$ be a *curve*, with components

$$\gamma(t) = \big(\gamma_1(t), \dots, \gamma_n(t)\big).$$

Show that γ is differentiable at t if and only if each γ_i is, and

$$D\gamma(t) = \big(\gamma_1'(t), \dots, \gamma_n'(t)\big),$$

as a linear map from $\mathbb{R}$ to $\mathbb{R}^n$, which may be identified with a vetor in $\mathbb{R}^n$.

5.1.8 Example. Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ itself be a linear map, so $f(x+y) = f(x) + f(y)$ and $f(ax) = af(x)$ for all $x, y \in \mathbb{R}^n$, $a \in \mathbb{R}$. Show that f is differentiable at every $x \in \mathbb{R}^n$. What is its derivative?

5.1.9 Example. Suppose a map $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ satisfies

$$\|f(x)\| \le B \|x\|^2$$

for some B. Show that f is differentiable at $x = 0$ and compute $Df(0)$.

5.1.10 Theorem. *If f and g are two functions with domain $A \subset \mathbb{R}^n$ and range $\mathbb{R}^m$ which are differentiable at $x \in A$, and $a \in \mathbb{R}$, then*

(i) *$D(f+g)(x) = Df(x) + Dg(x)$, and*

(ii) *$D(af)(x) = aDf(x)$.*

5.1.11 Theorem (Chain rule). *Let $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $g : B \subseteq \mathbb{R}^m \longrightarrow \mathbb{R}^l$ be differentiable at $x \in A$ and $y = f(x) \in B$, respectively. Then $g \circ f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^l$ is differentiable at x and*

$$D(g \circ f)(x) = Dg\big(f(x)\big) \circ Df(x),$$

where the right hand side denotes the composition of the linear map $Df(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ with the linear map $Dg\big(f(x)\big) : \mathbb{R}^m \longrightarrow \mathbb{R}^l$.

5.2 Partial derivatives

5.2.1 Definition. Denote a function $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by

$$f(x) = \big(f_1(x), \dots, f_m(x)\big), \qquad x = (x_1, \dots, x_n),$$

where each *component function* is real-valued: $f_i : A \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ for every $i \in \{1, \ldots, m\}$. Equivalently, $f(x) = \sum_{i=1}^m f_i(x)e_i$, where $\{e_1, \ldots, e_m\}$ are the standard unit basis vectors in $\mathbb{R}^m$, with $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ having the ith entry 1 and all other entries 0.

For each $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$, the jth *partial derivative* of f_i at $x \in \mathbb{R}^n$ is the limit, if it exists,

$$\frac{\partial f_i}{\partial x_j}(x) = \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t} = \lim_{t \to 0} \frac{f_i(x_1, \ldots, x_j + t, \ldots, x_n) - f_i(x_1, \ldots, x_n)}{t}$$

where $e_j \in \{e_1, \ldots, e_n\}$ is the jth standard basis vector in $\mathbb{R}^n$ and $t \in \mathbb{R}$. In other words, $\frac{\partial f_i}{\partial x_j}(x)$ is the derivative in the ordinary sense of the real valued, single variable function $t \longmapsto f_i(x+te_j)$. It can be thought of as the derivative in the ordinary sense of $f_i(x) = f_i(x_1, \ldots, x_n)$ as a function of x_j, with all the other independent variables x_k for $k \neq j$ held constant.

5.2.2 Example. Compute the partial derivatives of the components of $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$, where

$$f(x_1, x_2, x_3) = \big(x_1 \sin(x_2),\, x_3 x_1^2 - e^{x_2},\, \cos^2(x_1 x_2)\big).$$

5.2.3 Proposition. *Suppose $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is differentiable at x. Then each partial derivative $\frac{\partial f_i}{\partial x_j}(x)$ exists, and the matrix of the linear map $Df(x)$ with respect to the standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$ is the $m \times n$* Jacobian matrix

$$Df(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \cdots & \frac{\partial f_2}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}. \tag{5.2}$$

Proof hint. Consider the composition of f with the curves $t \longmapsto x + te_j$ for each j. □

5.2.4 Example. Compute the Jacobian matrix of the function in Example 5.2.2.

It would be reasonable to guess that if the partial derivatives of a function f all exist at a point x, then f should be differentiable at x with derivative (5.2). Unfortuntely *this is not the case*! The partial derivatives might all exist, and yet the function may fail to be differentiable according to Definition 5.1.3. (In other words, the converse to Proposition 5.2.3 is false in general.)

5.2.5 Example. Define $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ by

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & (x, y) \neq (0, 0), \\ 0 & (x, y) = (0, 0). \end{cases}$$

(i) Show that $\frac{\partial f}{\partial x}(0)$ and $\frac{\partial f}{\partial y}(0)$ are both defined.

(ii) Show that, with $T \in L(\mathbb{R}^2, \mathbb{R})$ defined by the matrix $\begin{bmatrix} \frac{\partial f}{\partial x}(0) & \frac{\partial f}{\partial y}(0) \end{bmatrix}$, the limit

$$\lim_{(a,b) \to 0} \frac{\|f(0 + a, 0 + b) - f(0, 0) - T(a, b)\|}{\|(a, b)\|}$$

does not exist. (In fact, f itself is not even continuous at $(0, 0)$: the limit $\lim_{(a,b)\to(0,0)} f(a, b)$ itself does not exist!) In particular, f is not differentiable at $(0, 0)$.

(iii) Compute $\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y)$ for $(x, y) \neq (0, 0)$. Are the partial derivatives continuous at $(0, 0)$?

5.3 Gradient and directional derivatives

5.3.1 Definition. Let $f : A \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a scalar-valued function of n variables, and suppose f is differentiable at x. Then $Df(x) \in L(\mathbb{R}^n, \mathbb{R})$ is an $n \times 1$ matrix

$$Df(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n}(x) \end{bmatrix},$$

which may be uniquely identified with a vector in $\mathbb{R}^n$, called the *gradient* of f at x, defined so that

$$\langle \nabla f(x), h \rangle = Df(x)h, \qquad \text{for every } h \in \mathbb{R}^n.$$

Thus $\nabla f(x) = \big(\frac{\partial f}{\partial x_1}(x), \cdots, \frac{\partial f}{\partial x_n}(x)\big) \in \mathbb{R}^n$, or written as a column vector,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}.$$

5.3.2 Proposition*. *If $\gamma : (a, b) \subset \mathbb{R} \longrightarrow \mathbb{R}^n$ is a differentiable path, and $f : A \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is differentiable at $x = \gamma(t)$, then denoting by $\gamma'(t)$ the vector obtained by applying $D\gamma(t)$ to $1 \in \mathbb{R}$, we have*

$$\frac{d}{dt} f\big(\gamma(t)\big) = \big\langle \nabla f\big(\gamma(t)\big), \gamma'(t) \big\rangle.$$

By composing a real-valued function with various paths and using the previous result we deduce several properties of the gradient, as the next few problems show.

5.3.3 Definition. Let $u \in \mathbb{R}^n$ be a fixed vector with $\|u\| = 1$, and let $f : A \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a scalar valued differentiable function. The *directional derivative* of f at x in the direction u is the limit

$$\lim_{t \to 0} \frac{f(x + tu) - f(x)}{t}.$$

In particular, the directional derivative in the direction of a standard basis vector e_j is just $\frac{\partial f}{\partial x_j}(x)$.

5.3.4 Problem. Show that the directional derivative of f at x in the direction u is given by

$$Df(x)u = \langle \nabla f(x), u \rangle.$$

In particular, it is maximized over u by

$$\max\left\{ \langle \nabla f(x), u \rangle : \|u\| = 1 \right\} = \left\langle \nabla f(x), \frac{\nabla f(x)}{\|\nabla f(x)\|} \right\rangle = \|\nabla f(x)\|.$$

Conclude that the vector $\nabla f(x)$ points in the direction of maximal increase of f at x and has magnitude equal to the directional derivative (rate of increase) in this direction.

5.3.5 Problem. Show that the gradient is orthogonal to level sets of the function f. In other words, if $C = \{x \in A : f(x) = c\}$ for some fixed c and $v \in \mathbb{R}^n$ is a vector tangent to C at $x \in C$, then $\langle \nabla f(x), v \rangle = 0$.

5.3.6 Problem. Say f has a local maximum (resp. minimum) at $x \in A$ if there is some $\delta > 0$ such that

$$f(y) \le f(x) \quad (\text{resp. } f(y) \ge f(x)) \quad \text{for every } y \in D_\delta(x).$$

Show that, if f has a local maximum or minimum at x and it is differentiable there, then $\nabla f(x) = 0$.

5.3.7 Definition. Recall that a set $A \subset \mathbb{R}^n$ is *convex* if for every x and y in A, the line segment

$$(1-t)x + ty \in A, \quad \text{for every } t \in [0,1].$$

The set $\{(1-t)x + ty : t \in [0,1]\}$ is the *line segment* from x to y.

5.3.8 Theorem. *Let $A \subseteq \mathbb{R}^n$ be a convex set and $f : A \longrightarrow \mathbb{R}$ a differentiable function. Then for every x and y in A,*

$$f(y) - f(x) = Df(z)(y-x) = \langle \nabla f(z), (y-x) \rangle$$

for some z lying on the line segement from x to y.

The prior result is of course a generalization of the Mean Value Theorem 3.2.4 to real-valued multivariable functions. When the range of f has dimension greater than 1, it might not be the case that $f(y) - f(x) = Df(z)(y-x)$ for any z between x and y. We could apply the previous theorem to the component functions $f_1, \dots, f_m$, but the points $z_j \in \mathbb{R}^n$ for which $f_j(y) - f_j(x) = Df_j(z_j)(y-x)$ might not agree.

Nevertheless, we can obtain one important result by this method.

5.3.9 Theorem. *If $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a differentiable function on a convex set A and $Df(x) = 0$ for all $x \in A$, then f is constant.*

5.4 Continuity of derivatives

5.4.1 Definition. We say $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is *continuously differentiable* if it is differentiable for all $x \in A$, and the map

$$Df : A \longrightarrow L(\mathbb{R}^n, \mathbb{R}^m), \quad x \longmapsto Df(x)$$

is continuous (with respect to the operator norm on the space $L(\mathbb{R}^n, \mathbb{R}^m)$ of linear maps). We denote the class of continuously differentiable maps from A to $\mathbb{R}^m$ by $\mathcal{C}^1(A; \mathbb{R}^m)$.

5.4.2 Lemma. *A function $f = (f_1, \dots, f_m) : A \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is continuously differentiable if and only if each component function $f_i : A \longrightarrow \mathbb{R}$ is continuously differentiable, where $i \in \{1, \dots, m\}$.*

While we saw in Example 5.2.5 that existence of partial derivatives is not sufficient to imply differentibility, the additionional hypothesis of continuity remedies the matter.

5.4.3 Theorem. *If all partial derivatives of $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ exist and are continuous on A, then f is continuously differentiable on A.*

Proof hint. By Lemma 5.4.2 it suffices to assume $m = 1$. For $h = (h_1, \dots, h_n) \in \mathbb{R}^n$, denote by h^k the vector $(h_1, \dots, h_k, 0, \dots, 0)$, so in particular $h^0 = 0$ and $h^n = h$. Note that

$$f(x+h) - f(x) = \sum_{k=1}^{n} f(x+h^k) - f(x+h^{k-1}).$$

The mean value theorem shows that $f(x+h^k) - f(x+h^{k-1}) = h_k \frac{\partial f}{\partial x_k}(y_k)$ for some y_k on the line segment between $x + h^{k-1}$ and $x + h^k$. Use continuity of the partial derivatives to estimate $\frac{\partial f}{\partial x_k}(y_k)$ by $\frac{\partial f}{\partial x_k}(x)$ for small h and show that the condition for differentiability is satisfied. □

5.4.4 Problem*. Define the following functions:

(a) $\gamma : \mathbb{R} \longrightarrow \mathbb{R}^2$, $\gamma(t) = (t^2, t^3)$,

(b) $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$, $f(x_1, x_2) = \sin(x_1\, x_2)$,

(c) $g : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$, $g(x_1, x_2) = (x_2^2,\, x_1 x_2,\, x_1^3)$,

(d) $h : \mathbb{R}^3 \longrightarrow \mathbb{R}$, $h(y_1, y_2, y_3) = y_1 e^{y_2 + y_3}$,

(e) $\phi : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, $\phi(x_1, x_2) = (x_1 \cos x_2,\ x_1 \sin x_2)$.

Compute the Jacobian matrices and hence (by Proposition 5.2.3 and Theorem 5.4.3) the total derivatives of each, along with the Jacobian matrices of the compositions $f \circ \gamma$, $h \circ g$, $\phi \circ \gamma$ and $h \circ g \circ \gamma$. Verify the chain rule for each of the compositions.

We also have a useful replacement for the Mean Value Theorem if we suppose continuous differentiability.

5.4.5 Theorem. *Let $f : A \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a continuous differentiable function on a convex set A. Suppose that $\|Df(x)\| \le B$ for all $x \in A$, for some $B \ge 0$. (Here we are using the operator norm on the linear maps $Df(x)$.) Then for all $x, y \in A$,*

$$\|f(y) - f(x)\| \le B \, \|y - x\| \,. \tag{5.3}$$

Proof sketch. The proof will involve integrating vector valued functions, which we did not yet talk about, but here is the short version: if $g : [a,b] \longrightarrow \mathbb{R}^m$ has the property that all its component functions $g_i : [a,b] \longrightarrow \mathbb{R}$ are integrable, then we say g is integrable and set

$$\int_a^b g(t)\, dt = \big(\textstyle\int_a^b g_1(t)\, dt,\ \int_a^b g_2(t)\, dt,\ \ldots,\ \int_a^b g_m(t)\, dt \big) \in \mathbb{R}^m.$$

The integral satisfies the inequality $\left\| \int_a^b g(t)\, dt \right\| \le \int_a^b \|g(t)\|\, dt$, and the fundamental theorem of calculus holds: $\int_a^b g'(t)\, dt = g(b) - g(a)$ (since it holds component-wise), where $g'(t) = \big(g_1'(t), \ldots, g_m'(t)\big)$.

To get the estimate (5.3), let $g(t) = f\big(\gamma(t)\big)$, where $\gamma(t) = (1-t)x + ty$. □

5.4.6 Definition. If $f \in \mathcal{C}^1(A; \mathbb{R}^m)$, then the *second partial derivatives* of $f = (f_1, \ldots, f_m)$, if they exist, are defined by the limits

$$\frac{\partial^2}{\partial x_j \partial x_k} f_i(x) = \lim_{t \to 0} \frac{\frac{\partial}{\partial x_k} f_i(x + t e_j) - \frac{\partial}{\partial x_k} f_i(x)}{t}, \quad t \in \mathbb{R}, \quad 1 \le i \le m, \quad 1 \le j, k \le n.$$

If these all exist and are continuous in A, then we say f is a *twice continuously differentiable*, or $\mathcal{C}^2$ *function*, and denote the set of such maps by $\mathcal{C}^2(A; \mathbb{R}^m)$.

For simplicity, in the next few results we suppose that $n = 2$ and $m = 1$ (the general case is similar). The following Lemma is a kind of mean value theorem for second derivatives, which will be used in the proof of Theorem 5.4.9 below.

5.4.7 Lemma. *Suppose f is defined on an open set $A \subset \mathbb{R}^2$ and that $\frac{\partial}{\partial x_1} f$ and $\frac{\partial^2}{\partial x_2 \partial x_1} f$ exist at every point in A. Suppose also that $R \subset A$ is a closed rectangle having opposite vertices (a, b) and $(a + h, b + k)$. Then*

$$f(a+h, b+k) - f(a+h, b) - f(a, b+k) + f(a, b) = h\, k \frac{\partial^2}{\partial x_2 \partial x_1} f(x_1, x_2)$$

for some $(x_1, x_2) \in R$.

Proof hint. Consider $g(t) = f(t, b+k) - f(t, b)$. □

Generally speaking, even if they both exist, the second partial derivatives $\frac{\partial^2}{\partial x_i \partial x_j} f$ and $\frac{\partial^2}{\partial x_j \partial x_i} f$ need not agree.

5.4.8 Example. Let

$$f(x, y) = \begin{cases} \frac{xy(x^2 - y^2)}{x^2 + y^2} & (x, y) \neq (0, 0), \\ 0 & (x, y) = (0, 0). \end{cases}$$

Show that $\frac{\partial^2}{\partial x \partial y} f(0,0) \neq \frac{\partial^2}{\partial y \partial x} f(0,0)$, though both are defined.

Once again, the assumption of continuity saves the day.

5.4.9 Theorem. *Suppose that $f \in \mathcal{C}^2(A;\mathbb{R})$ for some open set $A \subset \mathbb{R}^2$. Then*

$$\frac{\partial^2}{\partial x_1 \partial x_2} f(x) = \frac{\partial^2}{\partial x_2 \partial x_1} f(x)$$

for all $x \in A$.

Remark. What about the total second derivative? Consider what kind of object it must be. Supposing that $f : A \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is differentiable on A, then the assignment $x \longmapsto Df(x)$ is a map

$$Df : A \subset \mathbb{R}^n \longrightarrow L(\mathbb{R}^n, \mathbb{R}^m).$$

While the range is not an ordinary Euclidean space, it is a vector space (of dimension nm), and comes equipped with the operator norm of Definition 5.1.1. The notion of differentiability makes sense for such a map, but note that the derivative of Df at $x \in A$, if it exists, is a linear map

$$D^2 f(x) = D(Df)(x) \in L\big(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)\big), \tag{5.4}$$

in other words, a linear map *from $\mathbb{R}^n$ into the space of linear maps* from $\mathbb{R}^n$ to $\mathbb{R}^m$. Such an object is equivalent to (and better thought of as) a *bilinear map*

$$D^2 f(x) : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad (h,k) \longmapsto D^2 f(x)(h,k) = \big(D(Df)(x)h\big)k$$

of two arguments, h and k in $\mathbb{R}^n$, where h is the argument from the first instance of $\mathbb{R}^n$ in (5.4) and k is the argument from the second instance of $\mathbb{R}^n$, and where *bilinearity* means that the map is linear with respect to both arguments:

$$\begin{aligned} D^2 f(x)(a_1 h_1 + a_2 h_2, k) &= a_1 D^2 f(x)(h_1, k) + a_2 D^2 f(x)(h_2, k), \\ D^2 f(x)(h, b_1 k_1 + b_2 k_2) &= b_1 D^2 f(x)(h, k_1) + b_2 D^2 f(x)(h, k_2) \qquad \text{for all } a_1, a_2, b_1, b_2 \in \mathbb{R}. \end{aligned}$$

Just as a linear map is represented by a matrix with respect to bases for its domain and range, a bilinear map would be represented by a 3 dimensional ($n \times n \times m$) array of numbers, which in this case are the partial derivatives $\frac{\partial f_i}{\partial x_j \partial x_k}$. Of course we don't usually try to write this down since writing 3 dimensional arrays of numbers is a pain!

The hypothesis of continuous second differentiability, i.e., $f \in \mathcal{C}^2(A;\mathbb{R}^m)$ implies that the bilinear map $D^2 f(x)$ is *symmetric* for every $x \in A$, meaning that

$$D^2 f(x)(h,k) = D^2 f(x)(k,h).$$

This is the total derivative version of Theorem 5.4.9.

Higher total derivatives are similarly expressed as *multilinear maps* $\mathbb{R}^n \times \cdots \times \mathbb{R}^n \longrightarrow \mathbb{R}^m$, though we shall not say much more about that here. The best way to analyze such objects is through the development of *tensor products* of vector spaces, a topic we refer to advanced linear algebra texts.

5.5 Contraction mappings and applications

We will return to the total derivative in §5.6, where we will generalize the inverse function theorem to higher dimensions. First though, we digress in order to develop a key result from which the inverse function theorem follows, and give another of its applications to the existence of solutions to ordinary differential equations.

5.5.1 Definition. Let $\Phi : A \longrightarrow A$ be any map from a set A to itself. We say $x \in A$ is a *fixed point* of Φ if

$$\Phi(x) = x.$$

Many existence proofs in analysis can be framed in terms of fixed points: if we want to deduce the existece of some object, for example a solution to an ordinary differential equation or the inverse of a differentiable function, we may be able to set up a map for which the solution we seek is a fixed point. There are many fixed point theorems, of which the following, known as the *Banach fixed point theorem*, or *contraction mapping principle* is among the most important.

5.5.2 Definition. Let (M, d) be a metric space and $A \subseteq M$ a subset. We say a map $\Phi : A \longrightarrow A$ is a *contraction mapping* if there exists a real number $0 \le c < 1$ such that

$$d\big(\Phi(x), \Phi(y)\big) \le c\, d(x, y), \quad \text{for all } x, y \in A.$$

5.5.3 Theorem (Contraction mapping principle)**.** *If M is complete and $A \subset M$ is closed, then a contraction mapping $\Phi : A \longrightarrow A$ has a unique fixed point.*

Proof hint. Pick any $x_0 \in A$ and consider the sequence $x_n = \Phi(x_{n-1})$. The proof of Lemma 1.4.21 may also be of some inspiration. □

Among the applications of the contraction mapping principle are two fundamental results of analysis: the multivariable inverse function theorem, which we consider in the following section, and the Picard existence and uniqueness theorem for ODE, which we consider here.

5.5.4 Definition. A *first order ordinary differential equation* (ODE) is an equation of the form

$$u'(t) = F\big(t, u(t)\big), \tag{5.5}$$

where $u : (a, b) \subset \mathbb{R} \longrightarrow \mathbb{R}$ is a differentiable function, and where $F : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ is some function of two variables. We say u is a *solution* to the equation *on the interval* (a, b) if (5.5) is satisfied for some interval (a, b), which we typically want to be as large as possible. There are often infinitely many solutions to the equation (5.5), so we usually also require that u satisfy a given *initial condition*

$$u(t_0) = u_0, \tag{5.6}$$

for some fixed $t_0 \in \mathbb{R}$ and $u_0 \in \mathbb{R}$.

More generally, a *first order system* is an equation of the form (5.5), (5.6), where $u : (a, b) \subset \mathbb{R} \longrightarrow \mathbb{R}^n$ for some $n > 1$, where $F : \mathbb{R} \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ and $u_0 \in \mathbb{R}^n$.

Remark. If u solves (5.5) with the initial condition (5.6), note that

$$u(t) = u_0 + \int_{t_0}^{t} F\big(s, u(s)\big)\, ds$$

by the fundamental theorem of calculus.

5.5.5 Problem. Show that the second order (scalar) ODE

$$u'' = -u$$

can be transformed into an equivalent first order system $(u, v)' = (v, -u)$. (What are its solutions?) More generally, any nth order scalar ODE may be transformed into an equivalent first order system in $\mathbb{R}^n$.

5.5.6 Theorem (Picard existence and uniqueness)**.** *Suppose $F : \mathbb{R} \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $(t, x) \longmapsto F(t, x)$ is continuous in t and* Lipschitz *in x, meaning there exists some $B \ge 0$ such that*

$$\|F(t, x) - F(t, y)\| \le B\, \|x - y\| \quad \textit{for all } t \in \mathbb{R} \textit{ and } x, y \in \mathbb{R}^n \tag{5.7}$$

Then there exists $\varepsilon > 0$ such that the ODE

$$u'(t) = F\big(t, u(t)\big), \quad u(t_0) = u_0$$

has a unique continuous solution on the interval $(t_0 - \varepsilon, t_0 + \varepsilon)$.

In fact, it is sufficient that (5.7) hold in a neighborhood of (t_0, u_0), i.e., there exist $\delta_1, \delta_2 > 0$ such that $\|F(t, x) - F(t, y)\| \le B\, \|x - y\|$ for all $t \in (t_0 - \delta_1, t_0 + \delta_1)$ and $x, y \in D_{\delta_2}(u_0) = \{z : \|u_0 - z\| < \delta_2\}$.

Proof hint. Note that a solution is a fixed point of the map

$$\Phi(u) = u_0 + \int_{t_0}^{t} F\big(s, u(s)\big)\, ds.$$

Arrange ε so that $\Phi : \mathcal{C}\big((t_0-\varepsilon, t_0+\varepsilon), D_{\delta_2}(u_0)\big) \longrightarrow \mathcal{C}\big((t_0-\varepsilon, t_0+\varepsilon), D_{\delta_2}(u_0)\big)$ is a contraction mapping. □

Remark. This theorem is remarkable for its generality: we assumed no particular structure of F beyond the fairly weak estimates (5.7). In particular it may be highly nonlinear in u. The contrast with partial differential equations is stark—indeed, there exist first order *linear* PDE (with complex coefficients) for which no solutions exist, even locally (i.e., on arbitrarily small open sets)!

The theorem is also remarkable for being relatively constructive. Indeed, from the proof of Theorem 5.5.3 it follows that the solution can be obtained via a sequence of approximate solutions, starting with $u(t) = u_0$ (the constant approximation), and then defined inductively by

$$u_n(t) = u_0 + \int_{t_0}^{t} F\big(s, u_{n-1}(s)\big)\, ds.$$

While Theorem 5.5.6 guarantees a solution for short times $t \in [t_0-\varepsilon, t_0+\varepsilon]$, we could wonder if it is always possible to extend such a solution to have domain $t \in \mathbb{R}$. This is not always possible (if F is nonlinear in u), as the following example shows.

5.5.7 Example. Find a solution to the scalar ODE

$$u' = u^2, \quad u(0) = 1.$$

Show that the domain of u cannot be extended to $t \geq 1$.

5.6 The Inverse and Implicit Function Theorems

We now proceed to generalize the inverse function theorem (3.2.8) to functions from $\mathbb{R}^n$ to itself (note that $m = n$ is necessary here), along with an important corollary, the "implicit function theorem", which allows us to deduce when an equation of the form $F(x, y) = 0$ for $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ implicitly defines a function $y = y(x)$.

5.6.1 Theorem (Inverse Function Theorem)**.** *Let $A \subseteq \mathbb{R}^n$ and $f \in \mathcal{C}^1(A; \mathbb{R}^n)$ a continuously differentiable map. Suppose that $Df(x_0)$ is invertible as a linear map for some $x_0 \in A$. Then f is* invertible near x_0*, meaning there exist open sets $O \ni x_0$ and $U \ni f(x_0)$ such that $f : O \longrightarrow U$ admits an inverse $f^{-1} : U \longrightarrow O$; moreover f^{-1} is continuously differentiable in U with derivative at $y = f(x)$ given by*

$$Df^{-1}(y) = \big(Df(x)\big)^{-1} \in L(\mathbb{R}^n, \mathbb{R}^n). \tag{5.8}$$

We break the proof up into a few Lemmas.

5.6.2 Lemma. *Set $g(x) = Df(x_0)^{-1} f(x)$. Then g is differentiable at x_0 with $Dg(x_0) = I$, where I denotes the identity map. Moreover, f is invertible near x_0 if and only if g is.*

Similarly, set $h(x) = f(x + x_0) - f(x_0)$. Then h satisfies $h(0) = 0$ and h is differentible at 0 *with $Dh(0) = Df(x_0)$. Moreover, f is invertible near x_0 if and only if h is invertible near* 0.

By Lemma 5.6.2 we can assume without loss of generality that $x_0 = 0$, $f(0) = 0 = y_0$ and that $Df(0) = I$. Then we seek to invert f near 0.

5.6.3 Lemma. *Suppose $f(0) = 0$ and $Df(0) = I$. Then*

$$f(x) = x + R(x),$$

where R is a $\mathcal{C}^1$ function satisfying $R(0) = 0$ and $DR(0) = 0$. In particular, there exists $\delta > 0$ such that $\|DR(x)\| \leq \frac{1}{2}$ for $\|x\| \leq \delta$, i.e., $x \in D_\delta^-(0)$.

Given a fixed y, we want to solve for x such that $y = f(x) = x + R(x)$, which is equivalent to

$$x = y - R(x),$$

in other words, x is a fixed point of the map $x \longmapsto y - R(x)$.

5.6.4 Lemma. *Fix y and set $\Phi_y(x) = y - R(x)$. If $\|y\| \le \delta/2$, then Φ_y is a contraction mapping on $D_\delta^-(0)$. Thus f admits a unique inverse $f^{-1} : D_{\delta/2}^-(0) \ni y \longrightarrow x \in D_\delta^-(0)$ such that $x = \Phi_y(x)$.*

Proof hint. Use the mean value inequality (5.3). □

5.6.5 Lemma. *Since $D_{\delta/2}^-(0)$ and $D_\delta^-(0)$ are compact and $f : D_\delta^-(0) \longrightarrow D_{\delta/2}^-(0)$ is continuous, $f^{-1} : D_{\delta/2}^-(0) \longrightarrow D_{\delta/2}^-(0)$ is continuous.*

To get open sets as in the Theorem statement, set $U = D_{\delta/2}(0)$ and $O = f^{-1}(U)$.

5.6.6 Lemma. *f^{-1} is differentiable at 0 with derivative $Df^{-1}(0) = \big(Df(0)\big)^{-1}$.*

This proves (5.8) for the point x_0. To get (5.8) for all $y \in U$, we proceed with one final argument.

5.6.7 Lemma. *For every $x \in O$,*

$$\|Df(x)h\| = \|h + DR(x)h\| \ge \tfrac{1}{2}\,\|h\|\,,$$

and it follows that $Df(x)$ is invertible for every $x \in O$. From the previous results, f^{-1} is differentible at $y = f(x)$ with derivative satisfying (5.8). In particular, $y \longmapsto Df^{-1}(y)$ is continuous, so $f^{-1} \in \mathcal{C}^1(U; O)$.