Real Analysis, Fall 2017

December 4, 2017

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)* $d(a,c) \le d(a,b) + d(b,c)$.

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) *$a \le u$ 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) *$b \le a$ 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 $n \in \mathbb{N}$,*

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

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.

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 $x, y \in \mathbb{R}^n$,*

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

[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

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

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 $z = (z_1, \dots, z_n) \in \mathbb{R}^n$,*

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

and

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

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

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

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 $(p, p, p, \ldots)$. ☐

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 $U_1, \ldots, U_n$ 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 [Hint: it is the union of all the* $D_\varepsilon(p)$ *for interior points* p*].*

(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 $a \le b$ 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 $j \ge 2$. □

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 $a_i \le b_i$.

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 $d(x, y) \le r$ 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 $f : A \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ on 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

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

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

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

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

and the *upper integral* of f by

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

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

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

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

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

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 Theorem. *If a bounded function $f : [a, b] \longrightarrow \mathbb{R}$ is continuous except on a set of measure zero, then f is integrable.*

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 statement we could say that every bounded function which is continuous almost everywhere is integrable.

3.3.18 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?

Remark. In fact the converse to Theorem 3.3.17 holds. With a bit more development, one can show:

3.3.19 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.*

3.3.20 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 $f(x) \le g(x)$ for all $x \in [a, b]$, then*

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

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

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

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

(a) *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 $F'(x) = f(x)$.*

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

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