Christopher Heil

Introduction to Real Analysis

Chapter 0

Online Expanded Chapter on Notation and Preliminaries

Last Updated: March 28, 2024

Chapter 0
Notation and Preliminaries: Expanded Version

This online chapter is an expanded version of the unnumbered Preliminaries chapter in the text "An Introduction to Real Analysis" by C. Heil. In this Chapter 0 we will review in detail the notation and background information that will be assumed throughout Chapters 1–9 of the main text (though we do assume that the reader has a basic familiarity with logic, sets, real numbers, and functions). For more details and for any omitted proofs of the facts reviewed in Sections 0.1–0.13 of this Chapter 0 we refer to calculus texts (such as [HHW18]), and undergraduate analysis texts (such as [Rud76] or Heil, "A First Course on Real Analysis"). For additional details on the results discussed in Sections 0.14–0.15 of Chapter 0 we refer to linear algebra texts (such as Axler, "Linear Algebra Done Right").

We use the symbol $\square$ to denote the end of a proof, and the symbol $\diamondsuit$ to denote the end of a definition, remark, example, or exercise. We also use $\diamondsuit$ to indicate the end of the statement of a theorem whose proof will be omitted. Some of the more challenging Problems in the text are marked with an asterisk *. A detailed index of symbols employed in the text can be found after Chapter 9 in the main text.

0.1 Numbers

The set of natural numbers is denoted by $\mathbb{N} = \{1, 2, 3, \dots\}$. The set of integers is $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$, $\mathbb{Q}$ denotes the set of rational numbers, $\mathbb{R}$ is the set of real numbers, and $\mathbb{C}$ is the set of complex numbers.

The *real part* of a complex number $z = a+ib$ (where $a, b \in \mathbb{R}$) is $\mathrm{Re}(z) = a$, and its *imaginary part* is $\mathrm{Im}(z) = b$. We say that *z is rational* if both its real and imaginary parts are rational numbers. The *complex conjugate* of z is $\overline{z} = a - ib$. The *modulus*, or *absolute value*, of z is

$$|z| \;=\; \sqrt{z\overline{z}} \;=\; \sqrt{a^2 + b^2}.$$

We have $|\mathrm{Re}(z)| \le |z|$ and $|\mathrm{Im}(z)| \le |z|$ for every complex number z.

If $z \ne 0$ then its *polar form* is $z = re^{i\theta}$ where $r = |z| > 0$ and $\theta \in [0, 2\pi)$. In this case the *argument* of z is $\arg(z) = \theta$. If z is an arbitrary complex number, then there exists a complex number α such that $|\alpha| = 1$ and $z\alpha = |z|$. If $z \ne 0$ then α is uniquely given by $\alpha = e^{-i\theta} = \overline{z}/|z|$, while if $z = 0$ then α can be any complex number that has unit modulus. If z is a real number, then α is simply ± 1.

Some useful identities are

$$z + \overline{z} \;=\; 2\,\mathrm{Re}(z), \qquad z - \overline{z} \;=\; 2i\,\mathrm{Im}(z),$$

and

$$|z+w|^2 \;=\; (z+w)\,(\overline{z}+\overline{w}) \;=\; |z|^2 + 2\,\mathrm{Re}(z\overline{w}) + |w|^2.$$

0.1.1 The Extended Real Line

We append ∞ and $-\infty$ to the real line to form the set of *extended real numbers* $[-\infty, \infty]$:

$$[-\infty, \infty] \;=\; \mathbb{R} \cup \{-\infty\} \cup \{\infty\}.$$

Although ∞ and $-\infty$ are part of the extended real line, *they are not numbers* and should be treated with care. Some texts, such as [Fol99], denote the extended real line by

$$\overline{\mathbb{R}} \;=\; [-\infty, \infty].$$

We extend many of the normal arithmetic operations to $[-\infty, \infty]$. For example, if $-\infty < a \le \infty$ then we set $a + \infty = \infty$. However, $\infty - \infty$ and $-\infty + \infty$ are *undefined*, and are referred to as *indeterminate forms*.

If a is strictly positive extended real number (so $0 < a \le \infty$), then we define

$$a \cdot \infty = \infty, \quad (-a) \cdot \infty = -\infty, \quad a \cdot (-\infty) = -\infty, \quad (-a) \cdot (-\infty) = \infty.$$

Further, we adopt the conventions that

$$0 \cdot (\pm\infty) \;=\; 0 \qquad \text{and} \qquad \frac{1}{\pm\infty} \;=\; 0.$$

If p is an extended real number in the range $1 \le p \le \infty$, then its *dual index* is the unique extended real number p' that satisfies

$$\frac{1}{p} + \frac{1}{p'} \;=\; 1.$$

We have $1 \le p' \le \infty$, and $(p')' = p$. If $1 < p < \infty$, then we can write p' explicitly as

$$p' \;=\; \frac{p}{p-1}.$$

Some examples are $1' = \infty$, $(3/2)' = 3$, $2' = 2$, $3' = 3/2$, and $\infty' = 1$.

More notation related to the extended real line will be defined below in Sections 0.3 and 0.7.

0.2 Sets

If X is a set then we often use lowercase letters such as x, y, z to denote elements of X. We review below some of the terminology and notation that we use for sets.

- The notation $x \in X$ means that x is an *element of* X, or we also say that x *belongs to* X. The notation $x \notin X$ means that x is not an element of X, or that x does not belong to X.

- If every element of a set A is also an element of a set B, then A is a *subset* of B, and in this case we write $A \subseteq B$ (note that this includes the possibility that A could equal B).

- A *proper subset* of a set B is a set $A \subseteq B$ such that $A \ne B$. We indicate this by writing $A \subsetneq B$.

  Note: Some authors use the notation $A \subset B$ to indicate that A is a subset of B. That is, $A \subset B$ is synonymous with our notation $A \subseteq B$. In particular, the notation $A \subset B$ does not indicate that A is a proper subset of B.

- The *empty set* is denoted by $\varnothing$. The empty set is a subset of every set.

- The notation $X = \{x : x \text{ has property P}\}$ means that X is the set of all x that satisfy property P. For example, the union of a collection of sets $\{X_j\}_{j \in J}$ is

  $$\bigcup_{j \in J} X_j \;=\; \bigl\{x : x \in X_j \text{ for some } j \in J\bigr\},$$

  and their intersection is

  $$\bigcap_{j \in J} X_j \;=\; \bigl\{x : x \in X_j \text{ for every } j \in J\bigr\}.$$

- If S is a subset of a set X, then the *complement* of S is

  $$X \backslash S \;=\; \bigl\{x \in X : x \notin S\bigr\}.$$

  We sometimes abbreviate $X \backslash S$ as S^{C} if the set X is understood.

- If A and B are subsets of a set X, then the *relative complement* of A in B is
$$B \backslash A = B \cap A^{\mathrm{C}} = \{x \in B : x \notin A\},$$
and the *symmetric difference* of A and B is
$$A \triangle B = (A \backslash B) \cup (B \backslash A).$$

- *De Morgan's Laws* state that
$$X \setminus \bigcup_{i \in I} X_i = \bigcap_{i \in I} (X \backslash X_i) \qquad \text{and} \qquad X \setminus \bigcap_{i \in I} X_i = \bigcup_{i \in I} (X \backslash X_i).$$

- The *Cartesian product* of sets X and Y is the set of all ordered pairs of elements of X and Y, i.e.,
$$X \times Y = \big\{(x,y) : x \in X,\, y \in Y\big\}.$$

- The *power set* of a set X is the set of all subsets of X. We denote the power set by $\mathcal{P}(X)$:
$$\mathcal{P}(X) = \big\{S : S \subseteq X\big\}.$$

- A collection of sets $\{X_i\}_{i \in I}$ is *disjoint* if $X_i \cap X_j = \varnothing$ whenever $i \neq j$. In particular, two sets A and B are disjoint if $A \cap B = \varnothing$.

- A collection of subsets $\{X_i\}_{i \in I}$ of a set X is a *cover* of a subset E of X if
$$\bigcup_{i \in I} X_i \supseteq E.$$

- A collection of sets $\{X_i\}_{i \in I}$ is a *partition* of a set X if it is both disjoint and covers X. That is $\{X_i\}_{i \in I}$ is a partition if and only if
$$\{X_i\}_{i \in I} \text{ is disjoint} \quad \text{and} \quad \bigcup_{i \in I} X_i = X.$$

0.3 Intervals and Extended Intervals

We introduce names for some special subsets of the real line and the extended real line.

0.3.1 Intervals

- An *open interval* in the real line $\mathbb{R}$ is any one of the following sets:

$$\begin{aligned}
(a,b) &= \{x \in \mathbb{R} : a < x < b\}, && -\infty < a < b < \infty,\\
(a,\infty) &= \{x \in \mathbb{R} : x > a\}, && a \in \mathbb{R},\\
(-\infty,b) &= \{x \in \mathbb{R} : x < b\}, && b \in \mathbb{R},\\
(-\infty,\infty) &= \mathbb{R}.
\end{aligned}$$

- A *closed interval* in $\mathbb{R}$ is any one of the following sets:

$$\begin{aligned}
[a,b] &= \{x \in \mathbb{R} : a \le x \le b\}, && -\infty < a < b < \infty,\\
[a,\infty) &= \{x \in \mathbb{R} : x \ge a\}, && a \in \mathbb{R},\\
(-\infty,b] &= \{x \in \mathbb{R} : x \le b\}, && b \in \mathbb{R},\\
(-\infty,\infty) &= \mathbb{R}.
\end{aligned}$$

- We refer to $[a,b]$ as a *bounded closed interval*, a *finite closed interval*, or a *compact interval*.

- An *interval* in $\mathbb{R}$ is a set that is either an open interval, a closed interval, or one of the following sets (which are sometimes referred to as "half-open intervals," even though they are neither open nor closed):

$$\begin{aligned}
(a,b] &= \{x \in \mathbb{R} : a < x \le b\}, && -\infty < a < b < \infty,\\
[a,b) &= \{x \in \mathbb{R} : a \le x < b\}, && -\infty < a < b < \infty.
\end{aligned}$$

- The empty set $\varnothing$ and a singleton $\{a\}$ are not intervals, but even so we adopt the following notational conventions:

$$[a,a] = \{a\} \qquad \text{and} \qquad (a,a) = [a,a) = (a,a] = \varnothing.$$

0.3.2 Extended Intervals

We also deal with subsets of the extended real line. An *extended interval* is any one of the following subsets of $[-\infty,\infty]$:

$$\begin{aligned}
(a,\infty] &= (a,\infty) \cup \{\infty\}, && a \in \mathbb{R},\\
[a,\infty] &= [a,\infty) \cup \{\infty\}, && a \in \mathbb{R},\\
[-\infty,b) &= \{-\infty\} \cup (-\infty,b), && b \in \mathbb{R},\\
[-\infty,b] &= \{-\infty\} \cup (-\infty,b], && b \in \mathbb{R},\\
[-\infty,\infty] &= \mathbb{R} \cup \{-\infty\} \cup \{\infty\}.
\end{aligned}$$

An extended interval is not an interval. Whenever we refer to an "interval" without qualification we implicitly *exclude* the extended intervals.

0.4 Equivalence Relations

Informally, we say that $\sim$ is a *relation* on a set X if for each choice of elements x and y in X we have exactly one of the following two possibilities:

$$x \sim y \;\; (x \text{ is related to } y) \qquad \text{or} \qquad x \not\sim y \;\; (x \text{ is not related to } y).$$

Formally, a relation $\sim$ is a set of ordered pairs of elements of X. That is, $\sim$ is a subset of the Cartesian product $X \times X$. If (x, y) belongs to the set $\sim$ then we write $x \sim y$, and if (x, y) does not belong to $\sim$ then we write $x \not\sim y$.

For example, we can define a relation $\sim$ on the set of integers $\mathbb{Z}$ by declaring that two numbers are related if and only if their difference is even, i.e.,

$$m \sim n \quad \iff \quad m - n \text{ is divisible by } 2. \tag{0.1}$$

This is a relation because every pair of integers is either related or not related.

Definition 0.4.1. An *equivalence relation* on a set X is a relation $\sim$ that satisfies the following conditions for all x, y, $z \in X$.

- Reflexivity: $x \sim x$.
- Symmetry: If $x \sim y$ then $y \sim x$.
- Transitivity: If $x \sim y$ and $y \sim z$ then $x \sim z$.

If $\sim$ is an equivalence relation on X, then the *equivalence class* of an element $x \in X$ is the set $[x]$ that contains all elements that are related to x:

$$[x] \;=\; \{y \in X : x \sim y\}. \qquad \diamondsuit$$

The reader should prove that any two equivalence classes are either identical or disjoint. That is, if x and y are two points in X, then either $[x] = [y]$ or $[x] \cap [y] = \varnothing$. The union of all of the equivalence classes $[x]$ is X. Consequently, the set of *distinct* equivalence classes forms a partition of X.

Example 0.4.2. (a) The relation $\sim$ on $\mathbb{Z}$ defined in equation (0.1) is an equivalence relation (see Problem 0.4.3). The equivalence class of an integer $m \in \mathbb{Z}$ is

$$[m] \;=\; \{n \in \mathbb{Z} : m - n \text{ is even}\} \;=\; \{m + 2k : k \in \mathbb{Z}\}.$$

There are only two possibilities: If m is even then the equivalence class of m is $[m] = E$, the set of all even integers, while if m is odd then $[m] = O$, the set of all odd integers. No matter what we choose for m and n, either $[m] = [n]$ or $[m] \cap [n] = \varnothing$. There are only two distinct equivalence classes, the sets E and O, and these two sets form a partition of $\mathbb{Z}$.

(b) If we fix an integer $N > 1$, then we can define an analogous equivalence relation on $\mathbb{Z}$ by declaring that $m \sim n$ if and only if $m - n$ is divisible by N. There are N distinct equivalence classes in this case, which we can list as $[0], [1], \dots, [N-1]$. Although we will not need to make use of this fact, we remark that the *group* known as $\mathbb{Z}_N$ is the set of these equivalence classes, $\mathbb{Z}_N = \bigl\{[0], [1], \dots, [N-1]\bigr\}$, together with an appropriate definition of addition of equivalence classes (specifically, $[m] + [n] = [m+n]$).

(c) We can define a relation on $\mathbb{R}$ by declaring that $x \sim y$ if and only if $x - y$ is rational:

$$x \sim y \quad \iff \quad x - y \in \mathbb{Q}.$$

This is an equivalence relation on $\mathbb{R}$, and the equivalence class of $x \in \mathbb{R}$ is the set of all numbers that differ from x by a rational amount:

$$[x] \;=\; \bigl\{y \in \mathbb{R} : x - y \in \mathbb{Q}\bigr\} \;=\; \bigl\{x + r : r \in \mathbb{Q}\bigr\}. \tag{0.2}$$

Any two equivalence classes are either identical or disjoint. For example, $[0] = \mathbb{Q}$ and $[\sqrt{2}] = \{\sqrt{2} + r : r \in \mathbb{Q}\}$ have no elements in common. This equivalence relation will be useful to us in Section 2.4. $\diamondsuit$

Since the equivalence class $[x]$ defined in equation (0.2) is obtained by translating each element of the set of rationals by x, we often denote it by $x + \mathbb{Q}$ or $\mathbb{Q} + x$. That is, we let

$$x + \mathbb{Q} \;=\; \mathbb{Q} + x \;=\; \bigl\{x + r : r \in \mathbb{Q}\bigr\}.$$

Problems

0.4.3. Prove that the relations defined in Example 0.4.2 are each equivalence relations.

0.4.4. (a) Assume that $\sim$ is an equivalence relation on a set X. Prove that the set of distinct equivalence classes of $\sim$ forms a partition of X.

(b) Suppose that $\{X_i\}_{i \in I}$ is a partition of a set X. For each pair of elements $x, y \in X$, define $x \sim y$ if and only if there exists some $i \in I$ such that x and y both belong to X_i. Prove that $\sim$ is an equivalence relation on X.

0.5 Functions

Let X and Y be sets. We write $f \colon X \to Y$ to mean that f is a function whose *domain* is X and *codomain* (or *target*) is Y. We usually write $f(x)$ to denote the image of x under f, but sometimes we describe the rule for f by writing $x \mapsto f(x)$. For example, if $f \colon \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = 2x$, then we could alternatively say that f is given by the rule $x \mapsto 2x$ for $x \in \mathbb{R}$.

Here is some terminology that we use to describe various properties of a function $f\colon X \to Y$.

- The *direct image* of a set $A \subseteq X$ under f is

$$f(A) \;=\; \{f(x) : x \in A\}.$$

- The *inverse image* of a set $B \subseteq Y$ under f is

$$f^{-1}(B) \;=\; \{x \in X : f(x) \in B\}. \tag{0.3}$$

- The *range* of f is the direct image of its domain X:

$$\operatorname{range}(f) \;=\; f(X) \;=\; \{f(x) : x \in X\}.$$

- f is *surjective*, or *onto*, if $\operatorname{range}(f) = Y$.

- f is *injective*, or *one-to-one*, if $f(a) = f(b)$ implies $a = b$.

- f is a *bijection* if it is both injective and surjective.

- A bijection $f\colon X \to Y$ has an *inverse function* $f^{-1}\colon Y \to X$, defined by the rule $f^{-1}(y) = x$ if $f(x) = y$. The inverse function f^{-1} is also a bijection.

  Note: Despite the similar notation, an inverse function should not be confused with the *inverse image* defined in equation (0.3). Only a bijection has an inverse function, yet the inverse image $f^{-1}(B)$ is well-defined for *every* function f and set $B \subseteq Y$. Context determines the meaning: If y is an *element* of Y then $f^{-1}(y)$ must mean the image of y under the inverse function f^{-1}, while if B is a *subset* of Y then $f^{-1}(B)$ must mean the inverse image of B under f.

- If $Y = \mathbb{R}$, then we say that f is *real-valued*. If $Y = [-\infty, \infty]$, then f is *extended real-valued*. If $Y = \mathbb{C}$, then f is *complex-valued*.

- Given $S \subseteq X$, the *restriction* of a function $f\colon X \to Y$ to the domain S is the function $f|_S\colon S \to Y$ defined by $(f|_S)(x) = f(x)$ for $x \in S$.

- The *zero function on* X is the function $0\colon X \to \mathbb{R}$ defined by $0(x) = 0$ for every $x \in X$. We use the same symbol 0 to denote the zero function and the number zero.

- The *characteristic function* of a set $A \subseteq X$ is the function $\chi_A\colon X \to \mathbb{R}$ given by

$$\chi_A(x) \;=\; \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

- If the domain of a function f is $\mathbb{R}^d$, then the *translation* of f by a vector $a \in \mathbb{R}^d$ is the function $T_a f$ defined by $T_a f(x) = f(x - a)$ for $x \in \mathbb{R}^d$.

Problems

0.5.1. Given a function $f\colon X \to Y$, prove the following statements.

(a) f is surjective if and only if for every $y \in Y$ there exists some $x \in X$ such that $f(x) = y$.

(b) f is injective if and only if for every $y \in \mathrm{range}(f)$ there exists a unique $x \in X$ such that $f(x) = y$.

(c) f is a bijection if and only if for every $y \in Y$ there exists a unique $x \in X$ such that $f(x) = y$.

0.5.2. Let $f\colon X \to Y$ be a function, let B be a subset of Y, and let $\{B_i\}_{i\in I}$ be a family of subsets of Y. Prove the following statements.

(a) $f^{-1}\left(\bigcup_{i\in I} B_i\right) = \bigcup_{i\in I} f^{-1}(B_i)$.

(b) $f^{-1}\left(\bigcap_{i\in I} B_i\right) = \bigcap_{i\in I} f^{-1}(B_i)$.

(c) $f^{-1}\big(B^{\mathrm{C}}\big) = (f^{-1}(B))^{\mathrm{C}}$.

(d) $f(f^{-1}(B)) \subseteq B$, and equality holds if f is surjective. Show by example that equality need not hold if f is not surjective.

0.5.3. Let $f\colon X \to Y$ be a function, let A be a subset of X, and let $\{A_i\}_{i\in I}$ be a family of subsets of X. Prove the following statements.

(a) $f\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f(A_i)$.

(b) $f\left(\bigcap_{i\in I} A_i\right) \subseteq \bigcap_{i\in I} f(A_i)$, and equality holds if f is injective.

(c) $f(X)\backslash f(A) \subseteq f(A^{\mathrm{C}})$, and equality holds if f is injective.

(d) $A \subseteq f^{-1}(f(A))$, and equality holds if f is injective.

(e) Show by example that equality need not hold in statements (b)–(d) if f is not injective.

0.6 Cardinality

We say that two sets A and B *have the same cardinality* if there exists a bijection f that maps A onto B, i.e., if there is a function $f\colon A \to B$ that is both injective and surjective. Such a function f pairs each element of A with a unique element of B and vice versa, and therefore it is sometimes called a *one-to-one correspondence*.

Example 0.6.1. (a) The function $f\colon [0,2] \to [0,1]$ defined by $f(x) = x/2$ for $0 \le x \le 2$ is a bijection, so the intervals $[0,2]$ and $[0,1]$ have the same cardinality. This shows that a proper subset of a set can have the same cardinality as the set itself (although this is impossible for *finite* sets).

(b) The function $f\colon \mathbb{N} \to \{2,3,4,\dots\}$ defined by $f(n) = n+1$ for $n \in \mathbb{N}$ is a bijection, so the set of natural numbers $\mathbb{N} = \{1,2,3,\dots\}$ has the same cardinality as its proper subset $\{2,3,4,\dots\}$.

(c) The function $f\colon \mathbb{N} \to \mathbb{Z}$ defined by

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

is a bijection of $\mathbb{N}$ onto $\mathbb{Z}$, so the set of integers $\mathbb{Z}$ has the same cardinality as the set of natural numbers $\mathbb{N}$.

(d) If n is a finite positive integer, then there is no way to define a function $f\colon \{1,\dots,n\} \to \mathbb{N}$ that is a bijection. Hence $\{1,\dots,n\}$ and $\mathbb{N}$ do not have the same cardinality. Likewise, if $m \ne n$ are distinct positive integers, then $\{1,\dots,m\}$ and $\{1,\dots,n\}$ do not have the same cardinality. ♢

We use cardinality to define finite sets and infinite sets, as follows.

Definition 0.6.2 (Finite and Infinite Sets). Let X be a set.

(a) We say that X is *finite* if either X is empty or there exists an integer $n > 0$ such that X has the same cardinality as the set $\{1,\dots,n\}$. That is, a nonempty X is finite if for some $n \in \mathbb{N}$ we can find a bijection

$$f\colon \{1,\dots,n\} \to X.$$

In this case we say that X *has n elements*.

(b) We say that X is *infinite* if it is not finite. ♢

We use the following terminology to further distinguish among sets based on cardinality.

Definition 0.6.3 (Countable and Uncountable Sets). We say that a set X is:

(a) *denumerable* or *countably infinite* if it has the same cardinality as the natural numbers, i.e., if there exists a bijection $f\colon \mathbb{N} \to X$,

(b) *countable* if X is *either* finite *or* countably infinite,

(c) *uncountable* if X is not countable. ♢

0.6.1 Examples

Every finite set is countable by definition. Parts (b) and (c) of Example 0.6.1 show that the sets $\mathbb{N}$, $\mathbb{Z}$, and $\{2, 3, 4, \dots\}$ are countable. Here is another countable set.

Example 0.6.4. Consider $\mathbb{N}^2 = \mathbb{N} \times \mathbb{N} = \big\{(j,k) : j,k \in \mathbb{N}\big\}$, the set of all ordered pairs of positive integers. We depict $\mathbb{N}^2$ in table format in Figure 0.1.

$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$	$\cdots$
$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$\cdots$	
$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$\cdots$		
$(4,1)$	$(4,2)$	$(4,3)$	$\cdots$			
$(5,1)$	$(5,2)$	$\cdots$				
$(6,1)$	$\cdots$					
$\cdots$						

Fig. 0.1 The Cartesian product $\mathbb{N}^2 = \mathbb{N} \times \mathbb{N}$ as a table of ordered pairs.

Every ordered pair (j,k) of positive integers j and k appears somewhere in Figure 0.1. In particular, the first row of the table includes all those ordered pairs of positive integers (j,k) whose first component is $j = 1$, the second line lists those pairs whose first component is $j = 2$, and so forth.

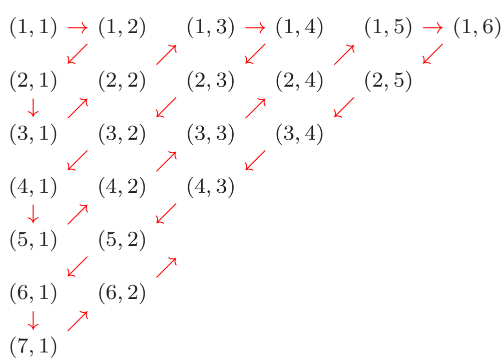

Fig. 0.2 Arrows show the pattern for defining a bijection of $\mathbb{N}$ onto $\mathbb{N}^2$.

Next, as shown in Figure 0.2, we insert arrows into the table in a certain pattern, and we define a bijection $f\colon \mathbb{N} \to \mathbb{N}^2$ by following these arrows.

Specifically, we set

$$\begin{aligned} f(1) &= (1,1), \\ f(2) &= (1,2), \\ f(3) &= (2,1), \\ f(4) &= (3,1), \\ f(5) &= (2,2), \\ f(6) &= (1,3), \\ &\vdots \end{aligned}$$

and continuing from there. In other words, once $f(n)$ has been defined to be a particular ordered pair (j,k), then we let $f(n+1)$ be the ordered pair that (j,k) points to next. In this way the outputs $f(1), f(2), f(3), \dots$ give us a list of *every ordered pair* in $\mathbb{N}^2$. Therefore $\mathbb{N}$ and $\mathbb{N}^2$ have the same cardinality, so $\mathbb{N}^2$ is denumerable and hence countable. $\diamondsuit$

As Example 0.6.4 illustrates, if X is a nonempty countable set then we can create a *list* of the elements of X. There are two possibilities. First, a countable set X might be finite, in which case there exists a bijection $f\colon \{1,2,\dots,n\} \to X$ for some positive integer n. Since f is surjective, we therefore have in this case that

$$X = \operatorname{range}(f) = \bigl\{f(1),\ f(2),\ \dots,\ f(n)\bigr\}.$$

Thus the function f gives us a way to list the n elements of X. On the other hand, if X is countably infinite then there is a bijection $f\colon \mathbb{N} \to X$, and hence

$$X = \operatorname{range}(f) = \bigl\{f(1),\ f(2),\ f(3),\ \dots\bigr\}.$$

Thus the elements of X have been again been listed in some order. For example, Example 0.6.4 shows that we can list the elements of $\mathbb{N}^2$ in the following order:

$$\mathbb{N}^2 = \bigl\{(1,1),\ (1,2),\ (2,1),\ (3,1),\ (2,2),\ (1,3),\ (1,4),\ (2,3),\ \dots\bigr\}.$$

Although it may seem more natural to depict $\mathbb{N}^2$ as a "two-dimensional table" (as shown in Figure 0.2), because $\mathbb{N}^2$ is countable it is also true that we can make a "one-dimensional list" of all of the elements of $\mathbb{N}^2$.

0.6.2 Uncountable Sets Exist

Now we show that there exist infinite sets that are not countable.

Example 0.6.5. Let S be the open interval $(0,1)$, which is the set of all real numbers that lie strictly between zero and one:

$$S = (0,1) = \{x \in \mathbb{R} : 0 < x < 1\}.$$

We will use an argument by contradiction to prove that S is not countable. First we recall that every real number can be written in decimal form. In particular, if $0 < x < 1$ then we can write

$$x = 0.d_1 d_2 d_3 \ldots = \sum_{k=1}^{\infty} \frac{d_k}{10^k},$$

where each digit d_k is an integer between 0 and 9. Some numbers have two decimal representations. For example, according to Problem 0.6.8, we have both

$$\frac{1}{2} = 0.5000\ldots = \frac{5}{10} + \sum_{k=2}^{\infty} \frac{0}{10^k}$$

and

$$\frac{1}{2} = 0.4999\ldots = \frac{4}{10} + \sum_{k=2}^{\infty} \frac{9}{10^k}. \tag{0.4}$$

Any number whose decimal representation ends in infinitely many zeros also has a decimal representation that ends in infinitely many nines, but all other real numbers have a unique decimal representation.

Suppose that S were countable. In this case there would exist a bijection $f\colon \mathbb{N} \to S$, and therefore we could make a list of all the elements of S. If we set $x_n = f(n)$, then this implies that

$$S = \operatorname{range}(f) = \{f(1),\ f(2),\ f(3),\ \ldots\} = \{x_1,\, x_2,\, x_3,\, \ldots\}$$

is a list of every real number that lies strictly between 0 and 1. Each number x_n can be represented in decimal form, say,

$$x_n = 0.d_1^n d_2^n d_3^n \ldots,$$

where each digit d_k^n is an integer between 0 and 9.

Now we will create another sequence of digits between 0 and 9. In fact, in order to avoid difficulties arising from the fact that some numbers have two decimal representations, we will always choose digits that are between 1 and 8. To start, let e_1 be any integer between 1 and 8 that does not equal d_1^1 (the first digit of the first number x_1). For example, if the decimal representation of x_1 happened to be $x_1 = 0.72839172\ldots$, then we let e_1 be any digit other than 0, 7, or 9 (so we might take $e_1 = 5$ in this case). Then we let e_2 be any integer between 1 and 8 that does not equal d_2^2 (the second digit of the second number x_2), and so forth. This gives us digits $e_1, e_2, \ldots$, and we let x be the real number whose decimal expansion has exactly those digits:

$$x = 0.e_1e_2e_3\ldots = \sum_{k=1}^{\infty} \frac{e_k}{10^k}.$$

Then x is a real number between 0 and 1, so x is one of the real numbers in the set S. Yet $x \neq x_1$, because the first digit of x (which is e_1) is not equal to the first digit of x_1 (why not—what if x_1 has two decimal representations?). Similarly $x \neq x_2$, because their second digits are different, and so forth. Hence x does not equal any element of S, which is a contradiction. Therefore S cannot be a countable set. ◇

0.6.3 Identifying Countable and Uncountable Sets

We assign the proof of the following properties of countable and uncountable sets to the reader.

Exercise 0.6.6. Let X and Y be sets.

(a) If X is countable and $Y \subseteq X$, then Y is countable.

(b) If X is uncountable and $Y \supseteq X$, then Y is uncountable.

(c) If X is countable and there exists an injection $f\colon Y \to X$, then Y is countable.

(d) If X is uncountable and there exists an injection $f\colon X \to Y$, then Y is uncountable. ◇

Example 0.6.7. (a) Let $\mathbb{Q}^+ = \{r \in \mathbb{Q} : r > 0\}$ be the set of all positive rational numbers. If $r \in \mathbb{Q}^+$, then there is a unique way to write r as a fraction in lowest terms. That is, $r = m/n$ for a unique choice of positive integers m and n that have no common integer factors other than ± 1. Therefore, by setting $f(r) = (m, n)$ we can define an injective map of $\mathbb{Q}^+$ into $\mathbb{N}^2$. Since $\mathbb{N}^2$ is countable and f is injective, we apply Exercise 0.6.6(c) and conclude that $\mathbb{Q}^+$ is countable.

A similar argument shows that $\mathbb{Q}^-$, the set of negative rational numbers, is countable. Problem 0.6.11 tells us that a union of finitely many (or even countably many) countable sets is countable, so it follows that the set of rational numbers is countable because $\mathbb{Q} = \mathbb{Q}^+ \cup \mathbb{Q}^- \cup \{0\}$.

(b) We saw in Example 0.6.5 that $S = (0, 1)$ is uncountable. Since $\mathbb{R} \supseteq S$, Exercise 0.6.6(b) implies that $\mathbb{R}$ is uncountable. Also, since every real number is a complex number we have $\mathbb{R} \subseteq \mathbb{C}$, and therefore $\mathbb{C}$ is uncountable as well.

(c) Let $I = \mathbb{R}\backslash\mathbb{Q}$ be the set of irrational real numbers. Since $\mathbb{R} = I \cup \mathbb{Q}$, if I were countable then $\mathbb{R}$ would be the union of two countable sets, which is countable by Problem 0.6.11. This is a contradiction, so the set of irrationals must be uncountable.

Thus $\mathbb{Q}$ is countable while I is uncountable. This may seem counterintuitive since between any two rational numbers there is an irrational, and between any two irrational numbers there is a rational number! $\diamondsuit$

Problems

0.6.8. Prove equation (0.4) (for the precise definition of an infinite series, see Section 0.12). Hint: If $x = \frac{4}{10} + \sum_{k=2}^{\infty} \frac{9}{10^k}$, then $10x = 4 + \sum_{k=1}^{\infty} \frac{9}{10^k}$.

0.6.9. Given sets A, B, and C, prove the following statements.

(a) A has the same cardinality as A.

(b) If A has the same cardinality as B, then B has the same cardinality as A.

(c) If A has the same cardinality as B and B has the same cardinality as C, then A has the same cardinality as C.

0.6.10. Prove that the closed interval $[0, 1]$ and the open interval $(0, 1)$ have the same cardinality by exhibiting a bijection $f\colon [0, 1] \to (0, 1)$. Hint: Do not try to create a *continuous* function f.

0.6.11. (a) Show that if X and Y are countable sets, then their Cartesian product $X \times Y$ and their union $X \cup Y$ is countable.

(b) Prove that the union of finitely many countable sets $X_1, \dots, X_n$ is countable.

(c) Suppose that $X_1, X_2, \dots$ are countably many sets, each of which is countable. Prove that

$$\bigcup_{k=1}^{\infty} X_k = X_1 \cup X_2 \cup \cdots$$

is countable. Thus the union of countably many countable sets is countable. Hint: Consider Figure 0.1.

0.6.12. Let F be the set of all functions $f\colon \mathbb{R} \to \mathbb{R}$, i.e., F is the set of all functions that map real numbers to real numbers. Prove that F is uncountable, and F does not have the same cardinality as the real line $\mathbb{R}$.

Hint: Suppose there were a bijection $a\colon \mathbb{R} \to F$. To simplify the notation, for each $x \in \mathbb{R}$ write a_x instead of $a(x)$. Then a_x is a *function* that maps real numbers to real numbers. The assumption that a is a bijection means that *every* function $f\colon \mathbb{R} \to \mathbb{R}$ is precisely one of these functions a_x. Consider the function defined by $f(x) = a_x(x) + 1$ for $x \in \mathbb{R}$.

0.7 Extended Real-Valued Functions

A function that maps a set X into the real line $\mathbb{R}$ is called a *real-valued function*, and a function that maps X into the extended real line $[-\infty, \infty]$ is an *extended real-valued function*. Every real-valued function is an extended real-valued function, but an extended real-valued function need not be real-valued. For example, if we set

$$f(x) \;=\; \begin{cases} 1/x, & x > 0, \\ \infty, & x = 0, \end{cases} \qquad \text{and} \qquad g(x) \;=\; \begin{cases} 1/x, & x > 0, \\ 0, & x = 0, \end{cases}$$

then f is extended real-valued but not real-valued, while g is both real-valued and extended real-valued.

An extended real-valued function f is *nonnegative* if $f(x) \ge 0$ for every x in its domain, where we use the convention that $0 \le \infty$ (indeed, $a < \infty$ for every real number a).

If $f\colon X \to [-\infty, \infty]$, then to avoid multiplicities of parentheses, brackets, and braces, we often write $f^{-1}(a,b) = f^{-1}((a,b))$, $f^{-1}[a,\infty) = f^{-1}([a,\infty))$, and so forth. We also use shorthands such as

$$\begin{aligned}
\{f > a\} \;&=\; \{x \in X : f(x) > a\} \;=\; f^{-1}(a,\infty], \\
\{f \ge a\} \;&=\; \{x \in X : f(x) \ge a\} \;=\; f^{-1}[a,\infty], \\
\{f = a\} \;&=\; \{x \in X : f(x) = a\} \;=\; f^{-1}\{a\}, \\
\{a < f < b\} \;&=\; \{x \in X : a < f(x) < b\} \;=\; f^{-1}(a,b) \\
\{f \ge g\} \;&=\; \{x \in X : f(x) \ge g(x)\}, \\
\{f = g\} \;&=\; \{x \in X : f(x) = g(x)\},
\end{aligned}$$

and so forth.

0.7.1 Positive and Negative Parts

Let $f\colon X \to [-\infty, \infty]$ be an extended real-valued function. We associate to f the two extended real-valued functions f^+ and f^- defined by

$$f^+(x) \;=\; \max\{f(x), 0\} \qquad \text{and} \qquad f^-(x) \;=\; \max\{-f(x), 0\}.$$

We call f^+ the *positive part* and f^- the *negative part* of f (see the illustration in Figure 0.3). They are each nonnegative, and we have the equalities

$$f(x) \;=\; f^+(x) \;-\; f^-(x) \qquad \text{and} \qquad |f(x)| \;=\; f^+(x) \;+\; f^-(x).$$

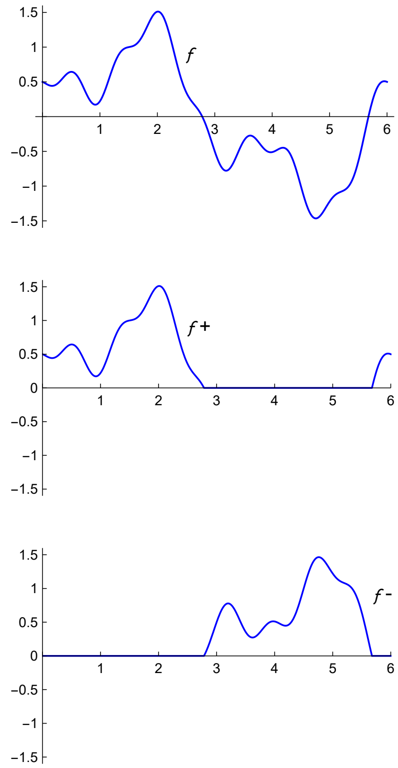

Fig. 0.3 A function f (top), its positive part f^+ (middle), and its negative part f^- (bottom).

Observe that even though $f^+(x)$ or $f^-(x)$ can be ∞, they cannot *simultaneously* be ∞. Therefore the expression $f^+(x)-f^-(x)$ is never an indeterminate form.

0.7.2 Monotone Functions

If $f: S \to [-\infty, \infty]$ is an extended real-valued function on a set $S \subseteq \mathbb{R}$, then we say that f is *monotone increasing* on S if for all $x, y \in S$ we have

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

We say that f is *strictly increasing* on S if for all $x, y \in S$,

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

Monotone decreasing and *strictly decreasing* functions are defined similarly.

Often the domain S of a monotone function is some type of interval. If $[a,b]$ is a finite closed interval and $f\colon [a,b] \to \mathbb{R}$ is a monotone increasing, real-valued function on $[a,b]$, then f is bounded since $f(a)$ and $f(b)$ must be finite real numbers and $f(a) \le f(x) \le f(b)$ for every $x \in [a,b]$. However, a monotone increasing function whose domain is any other type of interval can be unbounded, even if f never takes the values $\pm\infty$. For example, $f(x) = \frac{1}{x}$ is an unbounded, strictly decreasing function on $(0,1]$, even though $f(x)$ is finite for every $x \in (0,1]$.

0.7.3 Notation for Extended Real-Valued and Complex-Valued Functions

Most of the functions that we will encounter in this text will either be real-valued, extended real-valued, or complex-valued. A function of the form $f\colon X \to \mathbb{R}$ is said to be *real-valued*, a function of the form $f\colon X \to [-\infty,\infty]$ is *extended real-valued*, and a function of the form $f\colon X \to \mathbb{C}$ is *complex-valued*. We have the inclusions $\mathbb{R} \subseteq [-\infty,\infty]$ and $\mathbb{R} \subseteq \mathbb{C}$, so every real-valued function is both an extended real-valued and a complex-valued function. However, neither $[-\infty,\infty]$ nor $\mathbb{C}$ is a subset of the other, so an extended real-valued function need not be a complex-valued function, and a complex-valued function need not be an extended real-valued function. Hence there are usually two cases to consider:

- extended real-valued functions of the form $f\colon X \to [-\infty,\infty]$, and
- complex-valued functions of the form $f\colon X \to \mathbb{C}$.

To avoid excessive duplication, we introduce a notation that will allow us to consider both cases together.

Notation 0.7.1 (Scalars and the Symbol $\overline{\mathbf{F}}$). We let the symbol $\overline{\mathbf{F}}$ denote a choice of either the extended real line $[-\infty,\infty]$ or the complex plane $\mathbb{C}$. Associated with this choice, we make the following declarations.

- If $\overline{\mathbf{F}} = [-\infty,\infty]$, then the word *scalar* means a *finite real number* $c \in \mathbb{R}$.
- If $\overline{\mathbf{F}} = \mathbb{C}$, then the word *scalar* means a *complex number* $c \in \mathbb{C}$.

Note that a *scalar* cannot be $\pm\infty$; instead, a scalar is always a real or complex number. ♢

Thus, for example, when we write $f\colon X \to \overline{\mathbf{F}}$, we mean that f is either an extended real-valued or a complex-valued function on X. Both possibilities include real-valued functions as a special case.

Remark 0.7.2. The letter "$\overline{\mathbf{F}}$" here is related to the name "field." In many circumstances in analysis, we want to be able to use either the real line $\mathbb{R}$ or the complex plane $\mathbb{C}$ as our field. In these cases, it is not uncommon to use the symbol $\mathbb{F}$ or $\mathbf{F}$ to denote a choice of $\mathbb{R}$ or $\mathbb{C}$. For example, this notation is used in both [Heil11] and [Heil18].

However, in this text the essential choice is between the extended real line $\overline{\mathbb{R}} = [-\infty, \infty]$ and the complex plane $\mathbb{C}$. The extended real line is not a field, but it is related to the field $\mathbb{R}$. Hence fields are still the issue, and this is reason for the choice of the letter "F" in this context. Thus, in this text, $\overline{\mathbf{F}}$ denotes a choice of $\overline{\mathbb{R}} = [-\infty, \infty]$ or $\mathbb{C}$, but the reader should be aware that other notations are used in other texts, such as the use of $\mathbb{F}$ to denote a choice of $\mathbb{R}$ or $\mathbb{C}$. Additionally, some texts focus solely on one field, or move interchangeably between fields as convenient without adopting a notation to denote a choice of fields (or their extended versions).

The reader should also be aware that there is another notion, useful in topological contexts, of the *one-point compactification* of $\mathbb{R}$ or of $\mathbb{C}$. This is a distinct concept that will not be used in this text. In particular, for our purposes in real analysis, it is not useful to try to define a "complex infinity." $\diamondsuit$

0.8 Sequences

Let J be a fixed set. Given a set X and points $x_j \in X$ for $j \in J$, we write $\{x_j\}_{j \in J}$ to denote the sequence of elements x_j indexed by the set J. We call J an *index set* in this context, and we refer to x_j as the *jth component* of the sequence $\{x_j\}_{j \in J}$. If we know that the x_j are real or complex *numbers*, then we often write a sequence as $(x_j)_{j \in J}$ instead of $\{x_j\}_{j \in J}$. If the index set J is understood then we may write $\{x_j\}$, $\{x_j\}_j$, (x_j), or $(x_j)_j$, as appropriate.

Technically, a sequence $\{x_j\}_{j \in J}$ is shorthand for the function $x\colon J \to X$ whose rule is

$$x(j) = x_j, \qquad \text{for } j \in J.$$

Consequently the components x_j of a sequence need not be distinct—it is possible that we might have $x_i = x_j$ for some $i \neq j$.

Often the index set J is countable. The two most common situations are the finite index set $J = \{1, \dots, d\}$ and the countably infinite index set $J = \mathbb{N}$. If $J = \{1, \dots, d\}$ then we often write a sequence in list form as

$$\{x_n\}_{n=1}^d = \{x_1, \dots, x_d\} \qquad \text{or} \qquad (x_n)_{n=1}^d = (x_1, \dots, x_d).$$

Similarly, if $J = \mathbb{N} = \{1, 2, \dots\}$ then we often write

$$\{x_n\}_{n \in \mathbb{N}} = \{x_1, x_2, \dots\} \qquad \text{or} \qquad (x_n)_{n \in \mathbb{N}} = (x_1, x_2, \dots).$$

A *subsequence* of a countable sequence $\{x_n\}_{n \in \mathbb{N}} = \{x_1, x_2, \dots\}$ is a sequence of the form

$$\{x_{n_k}\}_{n\in\mathbb{N}} \;=\; \{x_{n_1}, x_{n_2}, \dots\} \quad \text{where } n_1 < n_2 < \cdots.$$

For example,

$$\{x_2, x_3, x_5, x_7, x_{11}, \dots\}$$

is a subsequence of $\{x_1, x_2, \dots\}$.

We say that a countable sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ is *monotone increasing* if $x_n \le x_{n+1}$ for every n, and *strictly increasing* if $x_n < x_{n+1}$ for every n. We define monotone decreasing and strictly decreasing sequences similarly.

0.8.1 The Kronecker Delta and the Standard Basis Vectors

If i and j are indices in an index set J (typically $J = \mathbb{N}$), then the *Kronecker delta* of i and j is the number δ_{ij} defined by the rule

$$\delta_{ij} \;=\; \begin{cases} 1, & \text{if } i = j,\\ 0, & \text{if } i \ne j. \end{cases}$$

If needed, we may insert a comma in the Kronecker delta symbol for clarity. For example, $\delta_{3,5}$ denotes the Kronecker delta with $i = 3$ and $j = 5$, so $\delta_{3,5} = 0$. However, in most abstract situations it is clear that i and j represent two different quantities, and in those circumstances we omit the comma and just write δ_{ij} instead of $\delta_{i,j}$.

If $n \in \mathbb{N}$ is a positive integer, then we let δ_n denote the sequence

$$\delta_n \;=\; (\delta_{nk})_{k\in\mathbb{N}} \;=\; (0, \dots, 0, 1, 0, 0, \dots).$$

That is, the nth component of the sequence δ_n is 1, while all other components are zero. We call δ_n the *nth standard basis vector*, and we refer to the family $\{\delta_n\}_{n\in\mathbb{N}}$ as the *sequence of standard basis vectors*, or simply the *standard basis*.

Problems

0.8.1. Prove that each of the following sets A, B, C, and D are uncountable.

(a) A is the set of sequences $x = (x_1, x_2, \dots)$ where each x_k is an integer between 0 and 9.

(b) B is the set of sequences $x = (x_1, x_2, \dots)$ where each x_k is an integer between 0 and 4.

(c) C is the set of sequences $x = (x_1, x_2, \dots)$ where each x_k is either 0 or 1.

(d) D is the set of sequences $x = (x_1, x_2, \dots)$ where each x_k is either 0 or 2.

0.9 Suprema and Infima

We introduce some terminology related to a set S of real numbers.

0.9.1 Bounded Sets and Maximum and Minimum Elements

- S is *bounded above* if there exists a real number M such that $x \le M$ for every $x \in S$. Any such number M is called an *upper bound* for S.
- S is *bounded below* if there exists a real number m such that $m \le x$ for every $x \in S$. Any such number m is called a *lower bound* for S.
- S is *bounded* if it is bounded both above and below. Equivalently, S is bounded if and only if there is a real number $M \ge 0$ such that $|x| \le M$ for all $x \in S$.
- x is a *maximum element* of S if $x \in S$ and $s \le x$ for every $s \in S$.
- x is a *minimum element* of S if $x \in S$ and $s \ge x$ for every $s \in S$.

Not every set of real numbers S has a maximum or minimum element, even if it is bounded. For example, the open interval $I = (0, 1)$ has no maximum or minimum element.

0.9.2 Supremum and Infimum

Often, a more useful notion than a maximum or minimum element is the supremum or infimum of a set. We consider these next.

Definition 0.9.1 (Supremum). Let S be a nonempty set of real numbers. We say that the number $u \in \mathbb{R}$ is the *supremum*, or *least upper bound*, of S if the following two statements hold.

(a) u is an upper bound for S, i.e., $s \le u$ for every $s \in S$.

(b) If v is any upper bound for S, then $u \le v$. That is, if $v \in \mathbb{R}$ and $s \le v$ for every $s \in S$, then $u \le v$.

We denote the supremum of S, if one exists, by $u = \sup(S)$.

If $S = (x_n)_{n\in\mathbb{N}}$ is countable, then we often write $\sup_n x_n$ or $\sup x_n$ to denote the supremum instead of $\sup(S)$. ◊

For example, the supremum of the open interval $S = (0,1)$ is $\sup(S) = 1$.

The supremum of a set S need not belong to the set. In fact, the supremum belongs to S if and only if S has a maximum element (see Problem 0.9.5).

It is not obvious whether every set that is bounded above has a supremum. We take the existence of suprema as an axiom, as follows.

Axiom 0.9.2 (Supremum Property of the Real Line). Let S be a nonempty subset of $\mathbb{R}$. If S is bounded above, then there exists a real number $x = \sup(S)$ that is the supremum of S. ◊

Here are two immediate but useful facts about the supremum of a set.

Lemma 0.9.3. *Let S be a nonempty subset of $\mathbb{R}$ that is bounded above.*

(a) *If x is* **any** *element of S, then*

$$x \leq \sup(S).$$

(b) *For each $\varepsilon > 0$ there exists* **some** *(at least one) $x \in S$ such that*

$$\sup(S) - \varepsilon < x.$$

Proof. For convenience of notation, let $u = \sup(S)$. Since S is bounded above, we have by the Supremum Axiom that u is a finite real number.

(a) By definition, u is an upper bound for S. Therefore $x \leq u$ for every $x \in S$.

(b) By definition, u is the *least* upper bound for S, so the number $u - \varepsilon$ is not an upper bound for S. Therefore there must exist some $x \in S$ such that $u - \varepsilon < x$. □

The *infimum*, or *greatest lower bound*, of S is defined in an entirely analogous manner, and is denoted by $\inf(S)$. Statements analogous to the ones made for suprema hold for infima.

0.9.3 Supremum and Infimum for Unbounded Sets

We extend the definition of a supremum to sets that are not bounded above by *declaring* that $\sup(S) = \infty$ if S is not bounded above. We also declare that $\sup(\varnothing) = -\infty$. Using these conventions, every set $S \subseteq \mathbb{R}$ has a supremum (although it might be $\pm\infty$), and the following statements hold.

- If S is empty, then $\sup(S) = -\infty$.
- If S is nonempty and bounded above, then $-\infty < \sup(S) < \infty$.
- If S is not bounded above, then $\sup(S) = \infty$.

0.9.4 Properties of Suprema

To illustrate the use of suprema, we prove the following result. Further results about suprema and infima (including facts about unbounded sets of numbers) are given in the problems for this section.

Lemma 0.9.4. *If* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *are two bounded sequences of real numbers, then*

$$\sup_{n\in\mathbb{N}} (x_n + y_n) \;\le\; \sup_{n\in\mathbb{N}} x_n \;+\; \sup_{n\in\mathbb{N}} y_n.$$

Proof. For simplicity of notation, set $u = \sup x_n$ and $v = \sup y_n$. Since $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are bounded, we know that u and v are finite real numbers. Since u is an upper bound for the x_n, we have $x_n \le u$ for every n. Similarly, $y_n \le v$ for every n. Therefore $x_n + y_n \le u + v$ for every n. Hence $u + v$ is an upper bound for the sequence $(x_n + y_n)_{n\in\mathbb{N}}$, so this sequence is bounded above and therefore has a finite supremum, say $w = \sup (x_n + y_n)$. By definition, w is the *least* upper bound for $(x_n + y_n)_{n\in\mathbb{N}}$. Since we have shown that $u+v$ is *an* upper bound for $(x_n+y_n)_{n\in\mathbb{N}}$, we must therefore have $w \le u + v$, which is exactly what we wanted to prove. □

Problems

0.9.5. Let S be a nonempty set of real numbers. Prove the following statements about suprema, and formulate and prove analogous results for infima.

Hint: We are not assuming that S is bounded, so first prove these results assuming S is bounded, and then separately consider the case of an unbounded set S.

(a) If S has a maximum element x, then $x = \sup(S)$.

(b) If $t \in \mathbb{R}$, then $\sup(S + t) = \sup(S) + t$, where $S + t = \{x + t : x \in S\}$.

(c) If $c \ge 0$, then $\sup(cS) = c \sup(S)$, where $cS = \{cx : x \in S\}$.

(d) If $a_n \le b_n$ for every n, then $\sup a_n \le \sup b_n$.

0.9.6. Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be two sequences of real numbers (not necessarily bounded).

(a) Show that if $c > 0$, then

$$\sup_{n\in\mathbb{N}} cx_n \;=\; c \sup_{n\in\mathbb{N}} x_n \qquad \text{and} \qquad \sup_{n\in\mathbb{N}} (-cx_n) \;=\; -c \inf_{n\in\mathbb{N}} x_n.$$

Remark: Recall our convention that $c \cdot (\pm\infty) = \pm\infty$ and $-c \cdot (\pm\infty) = \mp\infty$ for every positive real number c.

(b) Prove that

$$\inf_{n\in\mathbb{N}} x_n \,+\, \inf_{n\in\mathbb{N}} y_n \;\le\; \inf_{n\in\mathbb{N}} (x_n+y_n) \;\le\; \sup_{n\in\mathbb{N}} (x_n+y_n) \;\le\; \sup_{n\in\mathbb{N}} x_n \,+\, \sup_{n\in\mathbb{N}} y_n.$$

Show by example that any of the inequalities on the preceding line can be strict.

0.9.7. Let A and B be nonempty sets of real numbers. Prove that

$$\sup(A+B) \;=\; \sup(A) \;+\; \sup(B),$$

where $A+B = \{a+b : a \in A, b \in B\}$. Why does this not contradict Problem 0.9.6(b)?

0.9.8. Let S be a bounded, nonempty set of real numbers. Given a real number u, prove that $u = \sup(S)$ if and only if both of the following two statements hold:

(a) there does not exist any $s \in S$ such that $u < s$, and

(b) if $v < u$, then there exists some $s \in S$ such that $v < s$.

0.9.9. Let $\{E_n\}_{n\in\mathbb{N}}$ be a collection of nonempty bounded subsets of $\mathbb{R}$, and for each n let $s_n = \sup(E_n)$. Prove that if $\cup_{n=1}^{\infty} E_n$ is bounded, then

$$\sup\biggl(\bigcup_{n=1}^{\infty} E_n\biggr) \;=\; \sup\{s_n : n \in \mathbb{N}\}.$$

0.10 Convergent Sequences of Numbers

We review some terminology and facts related to convergence of sequences $(x_n)_{n\in\mathbb{N}}$ of real or complex numbers.

Definition 0.10.1 (Convergent Sequences). We say that a sequence of real or complex numbers $(x_n)_{n\in\mathbb{N}}$ *converges* if there exists some real or complex number x such that for every $\varepsilon > 0$ there is an $N > 0$ such that

$$n \,\ge\, N \quad\Longrightarrow\quad |x - x_n| \,<\, \varepsilon.$$

In this case we say that x_n *converges to* x *as* $n \to \infty$ and write

$$x_n \to x \qquad\text{or}\qquad \lim_{n\to\infty} x_n \,=\, x \qquad\text{or}\qquad \lim x_n \,=\, x. \qquad \diamondsuit$$

All convergent sequences are bounded (but not all bounded sequences converge).

Lemma 0.10.2. *If $(x_n)_{n\in\mathbb{N}}$ is a convergent sequence of real or complex numbers then it is bounded, i.e.,* $\sup |x_n| < \infty$.

Proof. Suppose that $x_n \to x$. Considering $\varepsilon = 1$, there must exists some $N > 0$ such that $|x - x_n| < 1$ for all $n \geq N$. Therefore, for $n \geq N$ we have

$$|x_n| \;=\; |x_n - x_N + x_N| \;\leq\; |x_n - x_N| \,+\, |x_N| \;\leq\; 1 + |x_N|.$$

Hence, for an arbitrary $n \in \mathbb{N}$,

$$|x_n| \;\leq\; \max\bigl\{|x_1|, \dots, |x_{N-1}|,\, |x_N| + 1\bigr\}.$$

Since the right-hand side of the line above is a constant that is independent of n, we see that the sequence $(x_n)_{n\in\mathbb{N}}$ is bounded. □

Here are some basic properties of convergent sequences.

Exercise 0.10.3. Assume that $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are convergent sequences of real or complex numbers. Prove the following statements.

(a) If c is a real or complex number then $(cx_n)_{n\in\mathbb{N}}$ converges, and

$$\lim_{n\to\infty} cx_n \;=\; c \lim_{n\to\infty} x_n.$$

(b) $(x_n + y_n)_{n\in\mathbb{N}}$ converges, and

$$\lim_{n\to\infty} (x_n + y_n) \;=\; \lim_{n\to\infty} x_n \;+\; \lim_{n\to\infty} y_n.$$

(c) $(x_n y_n)_{n\in\mathbb{N}}$ converges, and

$$\lim_{n\to\infty} x_n y_n \;=\; \Bigl(\lim_{n\to\infty} x_n\Bigr)\Bigl(\lim_{n\to\infty} y_n\Bigr).$$

(d) If $y_n \neq 0$ for every n and $\lim\limits_{n\to\infty} y_n \neq 0$, then $(1/y_n)_{n\in\mathbb{N}}$ converges, and

$$\lim_{n\to\infty} \frac{1}{y_n} \;=\; \frac{1}{\lim\limits_{n\to\infty} y_n}.$$

(e) If $y_n \neq 0$ for every n and $\lim\limits_{n\to\infty} y_n \neq 0$, then $(x_n/y_n)_{n\in\mathbb{N}}$ converges, and

$$\lim_{n\to\infty} \frac{x_n}{y_n} \;=\; \frac{\lim\limits_{n\to\infty} x_n}{\lim\limits_{n\to\infty} y_n}. \qquad \diamondsuit$$

0.10.1 Cauchy Sequences

If $(x_n)_{n\in\mathbb{N}}$ is a convergent sequence then there some number x such that x_n gets closer and closer to x as n increases. Closely related is the idea of a sequence where the points x_m and x_n get closer to *each other* as m and n

increase. We say that a sequence that has that property is a *Cauchy sequence*. Here is the precise definition.

Definition 0.10.4 (Cauchy Sequence). A sequence $(x_n)_{n\in\mathbb{N}}$ of real or complex numbers is a *Cauchy sequence* if for every $\varepsilon > 0$ there exists an integer $N > 0$ such that

$$m, n \geq N \quad \Longrightarrow \quad |x_m - x_n| < \varepsilon. \qquad \diamondsuit$$

It follows immediately from the definition that every convergent sequence of real or complex numbers is a Cauchy sequence (this is Problem 0.10.18). According to the following result, which is a consequence of the Supremum Property of the real line, the converse holds as well (this is Problem 0.10.19, or see [Rud76, Thm. 3.11]).

Theorem 0.10.5 (Cauchy Sequences of Scalars are Convergent). *If $(x_n)_{n\in\mathbb{N}}$ is a sequence of real or complex numbers, then*

$$(x_n)_{n\in\mathbb{N}} \text{ is convergent} \quad \Longleftrightarrow \quad (x_n)_{n\in\mathbb{N}} \text{ is Cauchy.} \qquad \diamondsuit$$

0.10.2 Divergence to Infinity

We introduce some terminology for a sequence of *real numbers* that increases without bound. We do not say that such a sequence converges but instead say that it *diverges to infinity.*

Definition 0.10.6 (Divergence to Infinity). If $(x_n)_{n\in\mathbb{N}}$ is a sequence of real numbers, then we say that $(x_n)_{n\in\mathbb{N}}$ *diverges to* ∞ if for each real number $R > 0$ there is an integer $N > 0$ such that $x_n > R$ for all $n \geq N$. In this case we write

$$x_n \to \infty, \qquad \lim_{n\to\infty} x_n = \infty, \qquad \text{or} \qquad \lim x_n = \infty.$$

We define *divergence to* $-\infty$ similarly. $\diamondsuit$

0.10.3 Convergence in the Extended Real Sense

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of *real numbers*. We say that

$$\lim_{n\to\infty} x_n \textbf{ \textit{exists}}$$

or that

$$(x_n)_{n\in\mathbb{N}} \textbf{ \textit{converges in the extended real sense}}$$

if either:

- x_n converges to a real number x as $n \to \infty$, *or*
- x_n diverges to ∞ as $n \to \infty$, *or*
- x_n diverges to $-\infty$ as $n \to \infty$.

Remark 0.10.7. In some circumstances in mathematics it is appropriate to introduce an analogue of ∞ for the complex plane. For example, this is done when we consider the topological "one-point compactification of the complex plane." Those notions are not appropriate *for the purposes of this text*, and hence we will not consider any analogue of "convergence in the extended real sense" for sequences of complex numbers. Consequently, if $(x_n)_{n\in\mathbb{N}}$ is a sequence of complex numbers, then its limit exists if and only if x_n converges to some *complex scalar* (not $\pm\infty$ or any notion of a "complex infinity"). $\diamondsuit$

0.10.4 Conventions

A sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers can converge, converge in the extended real sense, or not converge at all. A sequence $(x_n)_{n\in\mathbb{N}}$ of complex numbers can converge or not converge. The terminology in these two situations is similar but slightly different, yet it is usually clear from context what we mean when we say that a given sequence of scalars $(x_n)_{n\in\mathbb{N}}$ converges or that a limit exists. However, to be completely precise, we list the technical details here.

Notation 0.10.8 (Existence of a Limit of Scalars). Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of (real or complex) scalars.

- When we say that a generic sequence of scalars $(x_n)_{n\in\mathbb{N}}$ ***converges***, we mean that it converges to a scalar value. This applies to sequences of real numbers and to sequences of complex numbers.
- When we say that a sequence of real numbers ***exists***, this means that it exists in the extended real sense. When we say that a sequence of real numbers ***converges***, we mean that it converges to a finite real number.
- When we say that a sequence of complex numbers ***exists*** or that the sequence ***converges***, this means that it converges to a complex number.
- We do not use a concept of "complex infinity" in this text, and hence there is no notion of "divergence to infinity" for a sequence of *complex numbers*. Therefore, if $(x_n)_{n\in\mathbb{N}}$ is a sequence of complex numbers, then this sequence converges if and only the limit of the sequence exists and is a complex number. $\diamondsuit$

Example 0.10.9. Every monotone increasing sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ converges in the extended real sense, and in this case $\lim x_n = \sup x_n$. Similarly, a monotone decreasing sequence of real numbers converges in the extended real sense and its limit equals its infimum. $\diamondsuit$

Remark 0.10.10. Sometimes we need to consider sequences of extended real numbers, instead of just sequences of real numbers. The concepts that we have introduced extend to this setting. For example, if we allow each x_n to be an extended real number, then it is still true that a monotone increasing sequence of extended real numbers $(x_n)_{n\in\mathbb{N}}$ converges in the extended real sense. $\diamondsuit$

0.10.5 Pointwise Convergence of Functions

If X is a set and $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of extended real-valued or complex-valued functions whose domain is X, then we say that f_n *converges pointwise to a function* f if the limit of $f_n(x)$ exists for every x and

$$f(x) \;=\; \lim_{n\to\infty} f_n(x), \qquad \text{for every } x \in X.$$

In this case we write $f_n(x) \to f(x)$ *for every* $x \in X$ or $f_n \to f$ *pointwise.*

0.10.6 Monotone Sequences of Functions

If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of extended real-valued functions whose domain is a set X, then we say that $\{f_n\}_{n\in\mathbb{N}}$ is a *monotone increasing sequence* of functions if the sequence $\{f_n(x)\}_{n\in\mathbb{N}}$ is monotone increasing for each $x \in X$; that is:

$$f_1(x) \;\le\; f_2(x) \;\le\; \cdots \qquad \text{for every } x \in X.$$

In this case $f(x) = \lim\limits_{n\to\infty} f_n(x)$ exists for each $x \in X$ in the extended real sense, and we say that f_n *increases pointwise to* f. We denote this by writing $f_n \nearrow f$ on X.

Problems

0.10.11. Assume $(x_n)_{n\in\mathbb{N}}$ is a sequence of real or complex numbers, and there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ that converges to x. Prove that $x_n \to x$.

0.10.12. Prove the *Squeezing Theorem*: If a_n, b_n, and c_n are real numbers such that $a_n \le b_n \le c_n$ for each $n \in \mathbb{N}$ and $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} c_n$, then the sequence $(b_n)_{n\in\mathbb{N}}$ converges and $\lim\limits_{n\to\infty} b_n = \lim\limits_{n\to\infty} a_n$.

0.10.13. Assume that S is a nonempty set of real numbers that is bounded above. Prove that there exist numbers $x_n \in S$ such that $\lim\limits_{n\to\infty} x_n = \sup(S)$.

0.10.14. Assume that $(x_n)_{n\in\mathbb{N}}$ is a monotone increasing sequence of real numbers, i.e., each x_n is real and $x_1 \le x_2 \le \cdots$. Prove that $(x_n)_{n\in\mathbb{N}}$ converges if and only if $(x_n)_{n\in\mathbb{N}}$ is bounded, and in this case we have

$$\lim_{n\to\infty} x_n \;=\; \sup_{n\in\mathbb{N}} x_n.$$

Formulate and prove an analogous result for monotone decreasing sequences.

0.10.15. For each $n \in \mathbb{N}$, let

$$x_n \;=\; \sum_{k=1}^{n} \frac{1}{k} \;=\; 1 + \frac{1}{2} + \cdots + \frac{1}{n}.$$

Prove that the sequence $(x_n)_{n\in\mathbb{N}}$ is monotone increasing but unbounded, and therefore does not converge, but rather diverges to ∞. Even so, show that

$$\lim_{n\to\infty} \bigl(x_{n+1} - x_n\bigr) \;=\; 0.$$

Hint: $\frac{1}{2} \ge \frac{1}{2}$, $\frac{1}{2} + \frac{1}{3} + \frac{1}{4} \ge \frac{1}{2} + \frac{1}{4} + \frac{1}{4} = \frac{1}{2} + \frac{1}{2}$, $\frac{1}{2} + \cdots + \frac{1}{8} \ge \frac{1}{2} + \frac{1}{2} + \frac{1}{2}$.

0.10.16. Prove that if a sequence of numbers $(x_n)_{n\in\mathbb{N}}$ satisfies

$$|x_n - x_{n+1}| \;\le\; 2^{-n} \qquad \text{for every } n \in \mathbb{N},$$

then $(x_n)_{n\in\mathbb{N}}$ is Cauchy.

0.10.17. Suppose that $(x_n)_{n\in\mathbb{N}}$ is a sequence of real numbers that does not diverge to infinity.

(a) Prove that there is a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ that is bounded above.

(b) Must $(x_n)_{n\in\mathbb{N}}$ be bounded above? Either prove this, or exhibit a counterexample.

0.10.18. Prove the "easy" direction of Theorem 0.10.5, i.e., show that every convergent sequence of scalars is Cauchy.

0.10.19.* Use Axiom 0.9.2 to prove the "hard" direction of Theorem 0.10.5.

0.11 Limsup and Liminf

Not every sequence of real numbers converges. Consequently, instead of trying to use limits it is sometimes more useful to deal with the following weaker notions.

Definition 0.11.1 (Limsup and Liminf). The *limit superior*, or *limsup*, of a sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ is

$$\limsup_{n\to\infty} x_n \;=\; \inf_{n\in\mathbb{N}} \biggl(\sup_{m\ge n} x_m\biggr).$$

Likewise, the *limit inferior*, or *liminf*, of $(x_n)_{n\in\mathbb{N}}$ is

$$\liminf_{n\to\infty} x_n \;=\; \sup_{n\in\mathbb{N}} \biggl(\inf_{m\ge n} x_m\biggr). \qquad \diamondsuit$$

We sometimes use the abbreviated notations $\limsup x_n$ or $\liminf x_n$ to denote a limsup or a liminf.

The liminf and limsup of every sequence of real numbers exists in the extended real sense. That is, if $(x_n)_{n\in\mathbb{N}}$ is any sequence of real numbers then its liminf and limsup are extended real numbers in the range

$$-\infty \;\le\; \liminf_{n\to\infty} x_n \;\le\; \limsup_{n\to\infty} x_n \;\le\; \infty.$$

0.11.1 Examples

We start with a bounded sequence that does not converge.

Example 0.11.2. Let $x_n = (-1)^n$ for $n \in \mathbb{N}$. The sequence $\bigl((-1)^n\bigr)_{n\in\mathbb{N}}$ does not converge. The limsup of this sequence is

$$\limsup_{n\to\infty} \,(-1)^n \;=\; \inf_{n\in\mathbb{N}} \biggl(\sup_{m\ge n}\,(-1)^m\biggr) \;=\; \inf_{n\in\mathbb{N}} 1 \;=\; 1,$$

and its liminf is

$$\liminf_{n\to\infty} \,(-1)^n \;=\; \sup_{n\in\mathbb{N}} \biggl(\inf_{m\ge n}\,(-1)^m\biggr) \;=\; \inf_{n\in\mathbb{N}} (-1) \;=\; -1. \qquad \diamondsuit$$

Now we modify the preceding example so that it is a little more interesting.

Example 0.11.3. Let $(x_n)_{n\in\mathbb{N}}$ be the sequence obtained by interleaving terms of the form $1 + 1/n$ and $-1 - 1/n$. Specifically, set

$$x_n \;=\; \begin{cases} 1+\frac{1}{n}, & \text{if } n \text{ is odd},\\ -1-\frac{1}{n}, & \text{if } n \text{ is even}. \end{cases}$$

That is,

$$(x_n)_{n\in\mathbb{N}} \;=\; \bigl(1+1,\, -1-\tfrac{1}{2},\, 1+\tfrac{1}{3},\, -1-\tfrac{1}{4},\, 1+\tfrac{1}{5},\dots\bigr).$$

For each $n \in \mathbb{N}$ let $v_n = \sup_{m\ge n} x_m$. Then we see that:

$$\begin{aligned}
v_1 &= \sup_{m\ge 1} x_m \;=\; \sup\{1+1,\, -1-\tfrac{1}{2},\, 1+\tfrac{1}{3},\, -1-\tfrac{1}{4},\, 1+\tfrac{1}{5},\dots\} \;=\; 1+1,\\
v_2 &= \sup_{m\ge 2} x_m \;=\; \sup\{-1-\tfrac{1}{2},\, 1+\tfrac{1}{3},\, -1-\tfrac{1}{4},\, 1+\tfrac{1}{5},\dots\} \;=\; 1+\tfrac{1}{3},\\
v_3 &= \sup_{m\ge 3} x_m \;=\; \sup\{1+\tfrac{1}{3},\, -1-\tfrac{1}{4},\, 1+\tfrac{1}{5},\dots\} \;=\; 1+\tfrac{1}{3},\\
v_4 &= \sup_{m\ge 4} x_m \;=\; \sup\{-1-\tfrac{1}{4},\, 1+\tfrac{1}{5},\dots\} \;=\; 1+\tfrac{1}{5},\\
v_5 &= \sup_{m\ge 5} x_m \;=\; \sup\{1+\tfrac{1}{5},\dots\} \;=\; 1+\tfrac{1}{5},
\end{aligned}$$

and so forth. Hence

$$\limsup_{n\to\infty} x_n \;=\; \inf_{n\in\mathbb{N}} v_n \;=\; \inf\{1+1,\, 1+\tfrac{1}{3},\, 1+\tfrac{1}{5},\dots\} \;=\; 1.$$

The reader should show that the liminf of this sequence is -1. ♢

0.11.2 A Characterization of Convergence

We have the following characterization of convergent sequences.

Theorem 0.11.4. *Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence of real numbers. Then*

$$(x_n)_{n\in\mathbb{N}} \text{ converges} \qquad \Longleftrightarrow \qquad \liminf_{n\to\infty} x_n \;=\; \limsup_{n\to\infty} x_n.$$

Furthermore, in this case we have

$$\lim_{n\to\infty} x_n \;=\; \liminf_{n\to\infty} x_n \;=\; \limsup_{n\to\infty} x_n.$$

Proof. $\Rightarrow$. Assume that $w = \lim x_n$ exists and is a finite real number. Since $(x_n)_{n\in\mathbb{N}}$ is a bounded sequence, both

$$u \;=\; \liminf_{n\to\infty} x_n \qquad \text{and} \qquad v \;=\; \limsup_{n\to\infty} x_n$$

are finite, and $u \le v$. For each n, set

$$y_n = \inf_{m \ge n} x_m \qquad \text{and} \qquad z_n = \sup_{m \ge n} x_m.$$

If we fix a number $\varepsilon > 0$, then there exists some $N > 0$ such that

$$w - \varepsilon \le x_m \le w + \varepsilon, \qquad \text{for all } m \ge N.$$

Hence for $n \ge N$,

$$w - \varepsilon \le \inf_{m \ge n} x_m = y_n \le z_n = \sup_{m \ge n} x_m \le w + \varepsilon.$$

Consequently, since the y_n are increasing,

$$u = \sup_{n \ge 1} y_n = \sup_{n \ge N} y_n \ge w - \varepsilon.$$

Likewise, since the z_n are decreasing,

$$v = \inf_{n \ge 1} z_n = \inf_{n \ge N} z_n \le w + \varepsilon.$$

This is true for every $\varepsilon > 0$, so $u \ge w$ and $v \le w$. Hence $w \le u \le v \le w$, and therefore $w = u = v$.

$\Leftarrow$. We assign the proof of this direction as Problem 0.11.7. $\square$

Thus, although the limit of a bounded sequence need not exist, its liminf and limsup will always exist and the sequence converges if and only if the limsup and liminf are equal (and Problem 0.11.7 extends this fact to sequences that need not be bounded). More properties of the limsup and liminf of sequences are given in the problems for this section. In particular, Problem 0.11.10 gives several equivalent reformulations of the definition of a limsup.

0.11.3 Real-Parameter Versions

On occasion we deal with real-parameter versions of liminf and limsup. Given a real-valued function f whose domain includes an interval centered at a point $x \in \mathbb{R}$, we define

$$\limsup_{t \to x} f(t) = \inf_{\delta > 0} \sup_{|t-x|<\delta} f(t) = \lim_{\delta \to 0} \sup_{|t-x|<\delta} f(t),$$

and $\liminf_{t \to x} f(t)$ is defined analogously. The properties of these real-parameter versions of liminf and limsup are similar to those of the sequence versions.

Problems

0.11.5. Let $(x_n)_{n\in\mathbb{N}}$ be any sequence of real numbers. Prove that

$$\limsup_{n\to\infty}\,(-x_n) \;=\; -\liminf_{n\to\infty}\, x_n.$$

0.11.6. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of strictly positive real numbers. Suppose that we have either

$$\limsup_{n\to\infty} \frac{x_{n+1}}{x_n} \;<\; 1 \qquad \text{or} \qquad \limsup_{n\to\infty} x_n^{1/n} \;<\; 1.$$

Prove that there exists constants $0 < r < 1$ and $C > 0$ such that $x_n \le Cr^n$ for every n. Conclude that $\lim x_n = 0$.

0.11.7. (a) Finish the proof of Theorem 0.11.4.

(b) Given any sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers, prove that

$$(x_n)_{n\in\mathbb{N}} \text{ diverges to } \infty \quad\iff\quad \liminf_{n\to\infty} x_n \;=\; \limsup_{n\to\infty} x_n \;=\; \infty.$$

0.11.8. Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be any two sequences of real numbers.

(a) As long as none of the following sums takes the indeterminate forms $\infty - \infty$ or $-\infty + \infty$, prove that

$$\begin{aligned}
\liminf_{n\to\infty} x_n \;+\; \liminf_{n\to\infty} y_n \;&\le\; \liminf_{n\to\infty}\,(x_n + y_n)\\
&\le\; \limsup_{n\to\infty} x_n \;+\; \liminf_{n\to\infty} y_n\\
&\le\; \limsup_{n\to\infty}\,(x_n + y_n)\\
&\le\; \limsup_{n\to\infty} x_n \;+\; \limsup_{n\to\infty} y_n,
\end{aligned}$$

Show by example that strict inequality can hold on any line above. (Consequently, in contrast to limits, neither limsup nor liminf is linear in general.)

(b) Prove that if the sequence $(x_n)_{n\in\mathbb{N}}$ converges, then

$$\liminf_{n\to\infty}\,(x_n + y_n) \;=\; \lim_{n\to\infty} x_n \;+\; \liminf_{n\to\infty} y_n$$

and a similar equality holds with liminf replaced by limsup.

0.11.9. Given any sequence of real numbers $(x_n)_{n\in\mathbb{N}}$, prove that there exist subsequences $(x_{n_k})_{k\in\mathbb{N}}$ and $(x_{m_j})_{j\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} x_{n_k} \;=\; \limsup_{n\to\infty} x_n \qquad \text{and} \qquad \lim_{j\to\infty} x_{m_j} \;=\; \liminf_{n\to\infty} x_n.$$

Remark: In fact, the next problem shows that if $(x_n)_{n\in\mathbb{N}}$ is bounded then $\limsup x_n$ is the largest possible limit of any subsequence $(x_{n_k})_{k\in\mathbb{N}}$, and $\liminf x_n$ is the smallest limit of any subsequence.

0.11.10. Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence of real numbers and let x be a real number. Prove that the following five statements are equivalent.

(a) $x = \limsup\limits_{n\to\infty} x_n$.

(b) $x = \lim\limits_{n\to\infty} \Bigl(\sup\limits_{m\ge n} x_m\Bigr)$.

(c) If $\varepsilon > 0$, then there are infinitely many x_n with $x_n > x - \varepsilon$, but only finitely many x_n such that $x_n > x + \varepsilon$.

(d) $x = \inf\bigl\{y \in \mathbb{R} : \text{there are only finitely many } x_n > y\bigr\}$.

(e) $x = \sup\bigl\{y \in \mathbb{R} : \text{there exists a subsequence } x_{n_k} \to y\bigr\}$.

0.12 Infinite Series of Numbers

Infinite series in the setting of normed spaces will be considered in detail in Section 1.2.3. Here we will review issues related to the convergence of series of real or complex numbers.

We say that a series $\sum_{n=1}^\infty c_n$ of real or complex numbers *converges* if there is a real or complex number s such that the *partial sums*

$$s_N = \sum_{n=1}^N c_n$$

converge to s as $N \to \infty$. In this case $\sum_{n=1}^\infty c_n$ is defined to be s:

$$\sum_{n=1}^\infty c_n = \lim_{N\to\infty} s_N = \lim_{N\to\infty} \sum_{n=1}^N c_n = s.$$

If the series $\sum_{n=1}^\infty c_n$ does not converge, then we say that it *diverges*.

We sometimes use the shorthands $\sum c_n$ or $\sum_n c_n$ to denote a series.

0.12.1 Examples

Here are two particular examples of infinite series (for proof, see texts on calculus).

Lemma 0.12.1. (a) *If z is a real or complex number with $|z| < 1$, then $\sum_{k=0}^{\infty} z^k$ converges and has the value*

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z}.$$

Conversely, if $|z| \geq 1$, then $\sum_{k=0}^{\infty} z^k$ does not converge.

(b) *If z is any real or complex number, then $\sum_{k=0}^{\infty} z^k/k!$ converges and has the value*

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z. \qquad \diamondsuit$$

0.12.2 Properties of Convergent Series

An important property of convergent series is given in the next exercise.

Exercise 0.12.2 (The nth Term Test). If $\sum_{n=1}^{\infty} c_n$ is a convergent series of real or complex numbers, then

$$\lim_{n\to\infty} c_n = 0. \qquad \diamondsuit$$

The converse of Exercise 0.12.2 is false in general. For example, consider the *harmonic series*

$$\sum_{n=1}^{\infty} \frac{1}{n}.$$

Although the scalars $1/n$ converge to zero as $n \to \infty$, we saw in Problem 0.10.15 that the partial sums $s_N = \sum_{n=1}^{N} \frac{1}{n}$ of this series are not bounded. Therefore these partial sums cannot converge, and so the harmonic series does not converge.

0.12.3 Convergence of Series of Nonnegative Numbers

We often deal with series where every term c_n is a *nonnegative real number*. In this case there are only the following two possibilities (see Problem 0.12.7):

- If $c_n \geq 0$ for every n and the sequence of partial sums $\{s_N\}_{N\in\mathbb{N}}$ is bounded above, then the series $\sum c_n$ converges to a nonnegative real number. In this case we write

$$\sum_{n=1}^{\infty} c_n < \infty.$$

- If $c_n \geq 0$ for every n and the sequence of partial sums $\{s_N\}_{N \in \mathbb{N}}$ is not bounded above, then s_N diverges to infinity. In this case, the series $\sum c_n$ diverges. We say that $\sum c_n$ *diverges to infinity*, and write

$$\sum_{n=1}^{\infty} c_n = \infty.$$

We introduce the following terminology for this situation.

Notation 0.12.3 (Existence of a Series of Nonnegative Scalars). Assume that $c_n \geq 0$ for every n. Then either the series $\sum c_n$ converges to a real number or it diverges to infinity. We therefore say that a series $\sum c_n$ with all nonnegative terms *exists* or that it *converges in the extended real sense*. ◇

Note that saying that a series converges in the extended real sense does not mean that the series converges. Instead, $\sum c_n$ *converges* if the partial sums converge *to a finite real scalar*.

The following exercise regarding the "tails of a nonnegative series" is often useful.

Exercise 0.12.4 (Tails of Convergent Series). If $c_n \geq 0$ for every n and $\sum c_n < \infty$, then

$$\lim_{N \to \infty} \left(\sum_{n=N}^{\infty} c_n \right) = 0. \qquad \diamondsuit$$

0.12.4 Absolutely Convergent Series of Scalars

We say that an infinite series $\sum c_n$ of real or complex numbers c_n *converges absolutely* if

$$\sum_{n=1}^{\infty} |c_n| < \infty.$$

Recall that every Cauchy sequence of real numbers converges (see Theorem 0.10.5). It follows from this that every absolutely convergent series of scalars converges. That is, if c_n is a real or complex number for each $n \in \mathbb{N}$, then

$$\sum_{n=1}^{\infty} |c_n| < \infty \quad \Longrightarrow \quad \sum_{n=1}^{\infty} c_n \text{ converges.}$$

However, the converse implication fails in general. For example, the *alternating harmonic series*

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

converges, but it does not converge absolutely (this is Problem 0.12.5).

Problems

0.12.5. Prove that the *alternating harmonic series* $\sum (-1)^n \frac{1}{n}$ converges, but the *harmonic series* $\sum \frac{1}{n}$ diverges to infinity.

0.12.6. (a) Let a_n and b_n be real or complex numbers. Show that if $\sum a_n$ and $\sum b_n$ each converge, then $\sum (a_n + b_n)$ converges and

$$\sum_{n=1}^{\infty} (a_n + b_n) \;=\; \sum_{n=1}^{\infty} a_n \;+\; \sum_{n=1}^{\infty} b_n.$$

(b) Exhibit real numbers a_n and b_n such that $\sum (a_n + b_n)$ converges but $\sum a_n$ and $\sum b_n$ do not converge.

0.12.7. Suppose that $a_n \ge 0$ for each $n \in \mathbb{N}$. Prove that either $\sum a_n$ converges or it diverges to infinity.

0.12.8. Prove *Fatou's Lemma for Series*: If $a_{kn} \ge 0$ for all $k, n \in \mathbb{N}$, then

$$\sum_{k=1}^{\infty} \Bigl(\liminf_{n\to\infty} a_{kn}\Bigr) \;\le\; \liminf_{n\to\infty} \Biggl(\sum_{k=1}^{\infty} a_{kn}\Biggr).$$

Show by example that strict inequality can hold.

0.12.9. Prove the *Monotone Convergence Theorem for Series*: If for each $k \in \mathbb{N}$ we have that $0 \le a_{kn} \nearrow b_k$ as $n \to \infty$, then

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} a_{kn} \;=\; \sum_{k=1}^{\infty} b_k.$$

0.12.10. Suppose that J is an uncountable index set, and $x_j > 0$ for each $j \in J$. Prove that

$$\sup \left\{ \sum_{j\in F} x_j : F \subseteq J,\ F \text{ is finite} \right\} \;=\; \infty.$$

Remark: As a consequence, we declare that if J is uncountable and $x_j > 0$ for every $j \in J$, then $\sum_{j\in J} x_j = \infty$.

Hint: For each $k \in \mathbb{N}$ let $J_k = \bigl\{ j \in J : x_j > \frac{1}{k} \bigr\}$, and first prove that at least one set J_k must be infinite.

0.13 Continuity, Differentiation, and The Riemann Integral

In this section we briefly review some facts and terminology connected with continuity, differentiation, and integration.

0.13.1 Continuity

There are several equivalent ways to define continuity. For scalar-valued functions on an interval, we will take the following as our definition of continuity. More generally, continuity for functions on metric spaces will be explored in detail in Section 1.1.4.

Definition 0.13.1 (Continuity). Let I be an interval in the real line, and let f be a real-valued or complex-valued function on I (that is, f has the form $f\colon I \to \mathbb{R}$ or $f\colon I \to \mathbb{C}$).

(a) We say that f is *continuous on* I if for each $x \in I$ we have

$$\lim_{\substack{y \to x, \\ y \in I}} f(y) \;=\; f(x).$$

Stated explicitly, this means that for each point $x \in I$ and for each $\varepsilon > 0$, there must exist a number $\delta > 0$ such that

$$y \in I, \quad |x-y| \,<\, \delta \quad \Longrightarrow \quad |f(x)-f(y)| \,<\, \varepsilon. \tag{0.5}$$

(b) We say that f is *uniformly continuous on* I if for each $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$x, y \in I,\; |x-y| \,<\, \delta \quad \Longrightarrow \quad |f(x)-f(y)| \,<\, \varepsilon. \qquad \diamondsuit \tag{0.6}$$

Note that the value of δ in equation (0.5) implicitly depends on the choice of x. That is, if we choose a different x then we may need a different value for δ in order to make equation (0.5) hold. In contrast, the value of δ in equation (0.6) must be *independent* of the choice of x. That is, in order for a function to be uniformly continuous there must be a *single* δ such that equation (0.6) holds.

If $I = [a,b]$ is a finite closed interval, then every continuous function on I is both uniformly continuous and bounded on $[a,b]$ (this is proved in Theorem 1.1.17). However, if I is any other type of interval in $\mathbb{R}$, then there are continuous functions on I that are not uniformly continuous. For example, $f(x) = x^2$ is not uniformly continuous on $I = \mathbb{R}$, and $f(x) = 1/x$ is not uniformly continuous on $I = (0,1]$. There are bounded functions that are not uniformly continuous, such as $f(x) = \sin x^2$ on $I = \mathbb{R}$.

Remark 0.13.2. Consider the function f on $[0,\infty)$ defined by $f(x) = 1/x$ for $x > 0$ and $f(0) = \infty$. Is this function continuous? We have only considered continuity for scalar-valued functions, so the definitions that we have introduced to this point cannot be applied to this extended real-valued function f. It is possible to extended the notion of continuity to extended real-valued functions, by defining an appropriate topology on the extended real line $[-\infty,\infty]$. If we do this in a way that appropriately extends the topology of the real line, then it turns out that the function f defined above is continuous in this extended real sense. However, we will not consider this extended notion of continuity in this text. Instead, when we are given a function $f\colon X \to \overline{\mathbf{F}}$, we will only apply terminology related to continuity if the function is real-valued or complex-valued. This should usually be clear from context, and in most cases the scalar-valued condition will be explicit. ♢

0.13.2 Derivatives and Everywhere Differentiability

Let f be a real-valued or complex-valued function on a domain $D \subseteq \mathbb{R}$. If $x \in D$ and there is some open interval I such that $x \in I \subseteq D$, then we say that f is *differentiable* at x if the limit

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

exists and is a scalar (in particular, f is not differentiable at x if this limit takes the form $\pm\infty$). In this case we call $f'(x)$ the *derivative* of f at x.

Let $[a,b]$ be a closed interval in the real line. We say that a function of the form $f\colon [a,b] \to \mathbb{R}$ or $f\colon [a,b] \to \mathbb{C}$ is *differentiable everywhere* on $[a,b]$ if it is differentiable at each point x in the interior (a,b) and if the appropriate one-sided derivatives exist at the endpoints a and b. In other words, f is differentiable everywhere on $[a,b]$ if

$$f'(x) \;=\; \lim_{\substack{y\to x,\\ y\in[a,b]}} \frac{f(y)-f(x)}{y-x}$$

exists and is a scalar for each $x \in [a,b]$. We use similar terminology if f is defined on other types of intervals in $\mathbb{R}$. For example, $x^{3/2}$ is differentiable everywhere on $[0,1]$ and $x^{1/2}$ is differentiable everywhere on $(0,1]$, but $x^{1/2}$ is not differentiable everywhere on $[0,1]$.

The Mean-Value Theorem is one of most important results of differential calculus. A proof can be found in undergraduate calculus texts, such as [HHW18]. Note that this result only holds for *real-valued functions* (see Problem 0.13.6).

Theorem 0.13.3 (Mean-Value Theorem). *Assume that* $f\colon [a,b] \to \mathbb{R}$ *is continuous, and* f *is differentiable at each point of* (a,b)*. Then there exists some point* $c \in (a,b)$ *such that*

$$f'(c) \;=\; \frac{f(b)-f(a)}{b-a}. \qquad \diamondsuit$$

0.13.3 The Riemann Integral

For proofs of the statements made here regarding the Riemann integral, we refer to calculus texts such as [HHW18].

Let $f\colon [a,b] \to \mathbb{R}$ be a bounded, real-valued function on a finite, closed interval $[a,b]$. A *partition* of $[a,b]$ is a choice of finitely many points x_k in $[a,b]$ such that $a = x_0 < x_1 < \cdots < x_n = b$. If we wish to give this partition a name then we write:

$$\textit{Let } \Gamma = \bigl\{a = x_0 < \cdots < x_n = b\bigr\} \textit{ be a partition of } [a,b].$$

The *mesh size* of Γ is $|\Gamma| = \max\bigl\{x_j - x_{j-1} \,:\, j = 1,\dots,n\bigr\}$.

Given a partition $\Gamma = \bigl\{a = x_0 < \cdots < x_n = b\bigr\}$, for each $j = 1,\dots,n$ let m_j and M_j denote the infimum and supremum of f on the interval $[x_{j-1},x_j]$:

$$m_j \;=\; \inf_{x\in[x_{j-1},x_j]} f(x) \qquad \text{and} \qquad M_j \;=\; \sup_{x\in[x_{j-1},x_j]} f(x).$$

The numbers

$$L_\Gamma \;=\; \sum_{j=1}^n m_j\,(x_j - x_{j-1}) \qquad \text{and} \qquad U_\Gamma \;=\; \sum_{j=1}^n M_j\,(x_j - x_{j-1}),$$

are called *lower and upper Riemann sums for* f, respectively. We say that f *is Riemann integrable on* $[a,b]$ if there is a real number I such that

$$\sup_\Gamma L_\Gamma \;=\; \inf_\Gamma U_\Gamma \;=\; I,$$

where the supremum and infimum are taken over all partitions Γ of $[a,b]$. In this case, the number I is the *Riemann integral of* f *over* $[a,b]$.

Here is an equivalent definition of the Riemann integral. Given a partition $\Gamma = \{a = x_0 < \cdots < x_n = b\}$, choose any points $\xi_j \in [x_{j-1},x_j]$. We call

$$R_\Gamma \;=\; \sum_{j=1}^n f(\xi_j)\,(x_j - x_{j-1})$$

a *Riemann sum for* f (note that R_Γ implicitly depends on the choice of points ξ_j as well as the partition Γ). Then f is Riemann integrable if and only if there is a real number I such that

$$I \;=\; \lim_{|\Gamma|\to 0} R_\Gamma,$$

where this means that for every $\varepsilon > 0$, there is a $\delta > 0$ such that for any partition Γ with $|\Gamma| < \delta$ and any choice of points $\xi_j \in [x_{j-1}, x_j]$ we have $|I - R_\Gamma| < \varepsilon$. In this case, I is the Riemann integral of f over $[a,b]$.

We declare that a complex-valued function f on $[a,b]$ is Riemann integrable if its real and imaginary parts are both Riemann integrable.

Every continuous function $f\colon [a,b] \to \mathbb{C}$ is Riemann integrable, as is every piecewise continuous function on $[a,b]$. There exist functions that are not piecewise continuous but are Riemann integrable. We will characterize the Riemann integrable functions on $[a,b]$ in Section 4.5.5. The characteristic function of the rationals, $\chi_{\mathbb{Q}}$, is an example of a function that is not Riemann integrable on any interval $[a,b]$.

0.13.4 Indefinite Integrals

Suppose that $g\colon [a,b] \to \mathbb{C}$ is continuous on $[a,b]$. In this case g is Riemann integrable on every interval $[a,x]$ with $a \le x \le b$. The *indefinite integral* of g is the function

$$G(x) \;=\; \int_a^x g(t)\,dt, \qquad x \in [a,b].$$

The Fundamental Theorem of Calculus tells us that G is differentiable and its derivative is g.

Theorem 0.13.4 (Fundamental Theorem of Calculus). *If* $g\colon [a,b] \to \mathbb{C}$ *is continuous, then its indefinite integral* $G(x) = \int_a^x g(t)\,dt$ *is differentiable at each point* x *in* $[a,b]$, *and* $G'(x) = g(x)$ *for each* $x \in [a,b]$. $\diamondsuit$

Problems

0.13.5. Suppose that f and g are continuous functions whose domain is an interval I. Prove the following statements.

(a) cf is continuous for every real or complex number c.

(b) $f+g$ is continuous on I.

(c) fg is continuous on I.

(d) If $g(x) \ne 0$ for every x, then f/g is continuous on I.

0.13.6. This problem will show that the conclusion of the Mean-Value Theorem can fail for complex-valued functions. Set $f(t) = e^{it}$ for $t \in [0, 2\pi]$. This function is continuous on $[0, 2\pi]$ and is differentiable at every point of $(0, 2\pi)$. Prove that there is no point $c \in (0, 2\pi)$ such that

$$f'(c) \;=\; \frac{f(2\pi) - f(0)}{2\pi - 0}.$$

0.14 Vector Spaces

0.14.1 Euclidean Space

We will give the definition of a general vector space below. First, however, we discuss the most familiar vector space, $\mathbb{R}^d$, which is the set of all ordered d-tuples of real numbers. The complex analogue of $\mathbb{R}^d$ is also very important. This is $\mathbb{C}^d$, which is the set of all ordered d-tuples of complex numbers. We refer to either $\mathbb{R}^d$ or $\mathbb{C}^d$ as a *Euclidean space.* A *scalar* means an element of our chosen field. That is, a scalar is a real number if we are working with $\mathbb{R}^d$, while it means a complex number if we are working with $\mathbb{C}^d$.

If x is an element of $\mathbb{R}^d$ or $\mathbb{C}^d$, then x is a d-tuple of real or complex numbers. We usually write x as

$$x \;=\; (x_1, \dots, x_d),$$

and refer to x_k as the *kth component* of x. However, on occasion it is more convenient to write the components of x in the form

$$x \;=\; \bigl(x(1), \dots, x(d)\bigr).$$

Here are some important notions that apply to vectors in the Euclidean spaces $\mathbb{R}^d$ and $\mathbb{C}^d$.

- The *sum* of $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$ is the vector $x + y = (x_1 + y_1, \dots, x_d + y_d)$.
- The *product* (or *scalar product*) of a scalar c with a vector $x = (x_1, \dots, x_d)$ is $cx = (cx_1, \dots, cx_d)$.
- The *dot product* of two vectors $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$ is the scalar
$$x \cdot y \;=\; x_1 \overline{y_1} + \cdots + x_d \overline{y_d}. \tag{0.7}$$
- The *Euclidean norm* of a vector $x = (x_1, \dots, x_d)$ is
$$\|x\|_2 \;=\; (x \cdot x)^{1/2} \;=\; \bigl(|x_1|^2 + \cdots + |x_d|^2\bigr)^{1/2}. \tag{0.8}$$

If we are dealing with $\mathbb{R}^d$ then the complex conjugate in the definition of the dot product is superfluous and can be ignored. That is,

$$x, y \in \mathbb{R}^d \implies x \cdot y = x_1 y_1 + \cdots + x_d y_d.$$

However, if we are dealing with $\mathbb{C}^d$ then the complex conjugate is essential to the definition of the dot product. This is because $|z|^2 = z\overline{z}$ need not equal z^2 when z is complex. Including the complex conjugate, the final equality in equation (0.8) follows from the calculation that

$$x \cdot x = x_1 \overline{x_1} + \cdots + x_d \overline{x_d} = |x_1|^2 + \cdots + |x_d|^2.$$

0.14.2 Abstract Scalars

The definition of a vector space involves two sets and two operations that tell us how to combine elements of those sets. One of the two sets is the vector space itself (whose elements we call "vectors"), but we must also have a second set, called the associated *field of scalars* or simply the *scalar field*. There exist many different sets that are fields, but in this volume the only two scalar fields that we will ever consider are the real line $\mathbb{R}$ and the complex plane $\mathbb{C}$.

0.14.3 Abstract Vector Spaces

A vector space is a set V that is associated with a scalar field (always either $\mathbb{R}$ or $\mathbb{C}$ in this volume), and two operations that allow us to add vectors together and to multiply a vector by a scalar. Here is the precise definition (where we refer to an element of V as a "vector," and an element of the scalar field as a "scalar").

Definition 0.14.1 (Vector Space). A *vector space* is a set V, together with a scalar field, that satisfies the following conditions.

Closure Axioms

(1) Vector addition: For each pair of vectors $x, y \in V$, there is a unique vector $x + y$ in V, which we call the *sum* of x and y.

(2) Scalar multiplication: For each vector $x \in V$ and each scalar c, there exists a unique vector cx in V, which we call the *product* of c and x.

Addition Axioms

(3) Commutativity: $x + y = y + x$ for all $x, y \in V$.

(4) Associativity: $(x + y) + z = x + (y + z)$ for all $x, y, z \in V$.

(5) Additive Identity: There exists an element $0 \in V$ that satisfies $x+0=x$ for all $x \in V$. We call this element 0 the *zero vector* of V.

(6) Additive Inverses: For each vector $x \in V$, there exists a vector $(-x) \in V$ that satisfies $x+(-x)=0$. We call $-x$ the *additive inverse* of x, and we declare that $x-y=x+(-y)$.

Multiplication Axioms

(7) Associativity: $(ab)x = a(bx)$ for all scalars a, b and all vectors $x \in V$.

(8) Multiplicative Identity: Scalar multiplication by the number 1 satisfies $1x = x$ for every $x \in V$.

Distributive Axioms

(9) $c(x+y) = cx+cy$ for all vectors $x, y \in V$ and all scalars c.

(10) $(a+b)x = ax+bx$ for all vectors $x \in V$ and scalars a, b. ♢

Another name for a vector space is *linear space*. We call the elements of a vector space *vectors* (regardless of whether they are numbers, sequences, functions, operators, tensors, or other types of objects), and we call the elements of the scalar field *scalars*. The *trivial vector space* is $V = \{0\}$. If V contains more than just the zero vector, then it is a *nontrivial* vector space.

If S is a subset of a vector space V and S is itself a vector space (using the same operations of vector addition and scalar multiplication as V), then we call S a *subspace* of V. A *proper subspace* of V is a subspace S that satisfies $S \neq V$. A *nontrivial subspace* of V is a subspace S such that $S \neq \{0\}$. Thus, a *proper nontrivial subspace* is a subspace S that satisfies $\{0\} \subsetneq S \subsetneq V$.

Once we know that a given set V is a vector space, we can easily check whether a subset Y is a vector space by applying the following exercise. In the statement of this lemma, we implicitly assume that the vector space operations on Y are the same operations that are used in V.

Exercise 0.14.2. Let Y be a nonempty subset of a vector space V. Prove that if

(a) Y is closed under vector addition, i.e.,

$$x, y \in Y \implies x+y \in Y,$$

(b) Y is closed under scalar multiplication, i.e.,

$$x \in Y, \quad c \text{ is a scalar} \implies cx \in Y,$$

then Y is itself a vector space with respect to the operations of vector addition and scalar multiplication that are defined on V. ♢

A subset Y of V that satisfies the conditions of Exercise 0.14.2 is called a *subspace* of V.

0.14.4 Examples

We will give some examples of vector spaces whose elements are functions.

Example 0.14.3. Let X be a nonempty set, and let $\mathcal{F}(X)$ denote the set of all scalar-valued functions whose domain is X. That is, if our field of scalars is $\mathbb{R}$ then $\mathcal{F}(X)$ is the set of all real-valued functions $f\colon X \to \mathbb{R}$, while if our field of scalars is $\mathbb{C}$ then $\mathcal{F}(X)$ is the set of complex-valued functions $f\colon X \to \mathbb{C}$. Every real-valued function is complex-valued. Therefore, for example, if $X = \mathbb{R}$ then the function f whose rule is $f(t) = \sin t$ is a vector in $\mathcal{F}(\mathbb{R})$ regardless of whether the scalar field is $\mathbb{R}$ or $\mathbb{C}$. Similarly, if $g(t) = e^t$ and $h(t) = t^2$, then g and h are examples of vectors in $\mathcal{F}(\mathbb{R})$.

If f and g are two elements of $\mathcal{F}(X)$, then $f+g$ is the function defined by

$$(f+g)(t) \;=\; f(t) + g(t), \qquad \text{for } t \in X.$$

If $f \in \mathcal{F}(X)$ and c is a scalar, then the scalar product cf is the function whose rule is

$$(cf)(t) \;=\; cf(t), \qquad \text{for } t \in X.$$

The set $\mathcal{F}(X)$ is a nontrivial vector space with regard to the two operations of addition of functions and multiplication of a function by a scalar (this is Problem 0.14.7).

The zero element of the vector space $\mathcal{F}(X)$ is the function that maps every element of X to zero. We denote this function by the symbol 0, which is the same symbol that we use to represent the number zero. That is, the zero function is the function 0 whose rule is $0(t) = 0$ for every $t \in X$. It will usually be clear from context whether the symbol 0 is to be interpreted as the zero function or the number zero. ♢

Here are some examples of subspaces.

Example 0.14.4. (a) Let I be an interval in the real line, and let $C(I)$ be the set of all *continuous* scalar-valued functions whose domain is I, i.e.,

$$C(I) \;=\; \bigl\{f \in \mathcal{F}(I) : f \text{ is continuous}\bigr\}.$$

The zero function is continuous, so we have $0 \in C(I)$. If f and g are continuous then so are $f+g$ and cf, so $f+g \in C(I)$ and $cf \in C(I)$. Therefore $C(I)$ is nonempty and is closed under both addition and multiplication by scalars. Therefore Exercise 0.14.2 tells us that $C(I)$ is a subspace of $\mathcal{F}(I)$.

(b) Now let $\mathcal{P}$ be the set of all polynomial functions on I. That is, $\mathcal{P}$ consists of all functions p of the form

$$p(t) \;=\; \sum_{k=0}^{N} c_k t^k \;=\; c_0 + c_1 t + \cdots + c_N t^N, \qquad \text{for } t \in I,$$

where $N \geq 0$ and $c_0, c_1, \dots, c_N$ are scalars. Since $\mathcal{P}$ is a nonempty subset of $C(I)$ and $\mathcal{P}$ is closed under both addition and multiplication by scalars, we conclude that $\mathcal{P}$ is a subspace of $C(I)$, and therefore it is a subspace of $\mathcal{F}(I)$ as well. In fact, we have the inclusions $\{0\} \subsetneq \mathcal{P} \subsetneq C(I) \subsetneq \mathcal{F}(I)$. $\diamondsuit$

To avoid multiplicities of brackets and parentheses, if $I = (a,b)$ then we usually write $C(a,b)$ instead of $C((a,b))$, if $I = [a,b)$ then we usually write $C[a,b)$ instead of $C([a,b))$, and so forth.

If we restrict our attention to open intervals, then we can create further subspaces consisting of differentiable functions.

Example 0.14.5. Let I be an open interval in the real line, and let $C^1(I)$ be the set of all scalar-valued functions f that are differentiable on I and whose derivative f' is continuous on I:

$$C^1(I) \;=\; \bigl\{f \in C(I) \,:\, f \text{ is differentiable and } f' \text{ is continuous on } I\bigr\}.$$

If the scalar field is $\mathbb{R}$ then functions in $C^1(I)$ are real-valued, while if the scalar field is $\mathbb{C}$ then $C^1(I)$ includes both real-valued and complex-valued functions. Every differentiable function is continuous, so $C^1(I)$ is a subset of $C(I)$, but the examples below will show that it is a *proper* subset, i.e., $C^1(I) \subsetneq C(I)$. $\diamondsuit$

To illustrate, take $I = \mathbb{R}$ and let f, g, h, k be the functions defined by

$$f(t) = |t|, \qquad g(t) = t^2, \qquad h(t) = e^{-|t|}, \qquad k(t) = e^{-t^2}. \tag{0.9}$$

Each of these functions is real-valued, and hence is also complex-valued since every real number is a complex number. These functions have the following properties.

- f is continuous but not differentiable on $\mathbb{R}$, so f belongs to $C(\mathbb{R})$ but does not belong to $C^1(\mathbb{R})$.
- g is differentiable and $g'(t) = 2t$ is continuous on $\mathbb{R}$ (in fact, g' is differentiable on $\mathbb{R}$), so $g \in C^1(\mathbb{R})$.
- h is continuous but not differentiable on $\mathbb{R}$, so $h \in C(\mathbb{R}) \setminus C^1(\mathbb{R})$.
- k is differentiable and $k'(t) = -2te^{-t^2}$ is continuous on $\mathbb{R}$ (in fact, k' is differentiable on $\mathbb{R}$), so $k \in C^1(\mathbb{R})$.

The reader should use Exercise 0.14.2 to prove that $C^1(I)$ is a subspace of $C(I)$. Since $C^1(I)$ is contained in but not equal to $C(I)$, we see that $C^1(I)$ is a *proper subspace* of $C(I)$, which is itself a proper subspace of $\mathcal{F}(I)$.

The two functions g and k defined in equation (0.9) are actually *infinitely differentiable* on $\mathbb{R}$, which means that $g', g'', g''', \dots$ and $k', k'', k''', \dots$ all exist and are differentiable at every point. Are there any functions that are differentiable on $\mathbb{R}$ but are not infinitely differentiable?

Example 0.14.6. Let $I = \mathbb{R}$ and define

$$w(t) \;=\; \begin{cases} t^2 \sin \frac{1}{t}, & t \neq 0, \\ 0, & x = 0. \end{cases} \tag{0.10}$$

The reader should check (this is Problem 0.14.9) that:

- w is continuous on $\mathbb{R}$, so $w \in C(\mathbb{R})$,
- w is differentiable *at every point of* $\mathbb{R}$, i.e., $w'(t)$ exists and is a scalar for each $t \in \mathbb{R}$, but
- w' is *not* continuous at every point of $\mathbb{R}$.

Therefore, although w is differentiable on $\mathbb{R}$, its derivative is not continuous. We say that w is *once differentiable* because $w'(t)$ exists for every t. However, w is not *twice-differentiable* because $w''(t)$ does not exist for every t. Because w is continuous it is an element of $C(\mathbb{R})$, but $w \notin C^1(\mathbb{R})$ because even though $w'(t)$ exists at every point, w' is not continuous on $\mathbb{R}$. ◊

So far we have considered domains I that are open intervals. If I is an interval in $\mathbb{R}$ that is not open, then we define $C^1(I)$ by considering one-sided differentiability at any endpoint of I, just as we did in the discussion in Section 0.13.2. So, for example, $C^1[0,1]$ consists of all functions f that are differentiable everywhere on $[0,1]$ and whose derivative f' is continuous on $[0,1]$.

Thus $C^1(I)$ can be defined for any type of interval I. We can keep going and define $C^2(I)$ to be the space of all functions f such that both f and f' exist and are differentiable on I and f'' exists and is continuous on I. This is a proper subspace of $C^1(I)$. Then we continue further and define $C^3(I)$ and so forth, obtaining a nested decreasing sequence of spaces. The space $C^\infty(I)$ that consists of all infinitely differentiable functions is itself a proper subspace of each of these. Moreover, the set of all polynomial functions,

$$\mathcal{P} \;=\; \left\{ \sum_{k=0}^{N} c_k t^k : N \geq 0,\, c_k \text{ are scalars} \right\},$$

is a proper subspace of $C^\infty(I)$. Thus we have the infinitely many distinct vector spaces

$$\mathcal{F}(I) \;\supsetneq\; C(I) \;\supsetneq\; C^1(I) \;\supsetneq\; C^2(I) \;\supsetneq\; \cdots \;\supsetneq\; C^\infty(I) \;\supsetneq\; \mathcal{P}.$$

Problems

0.14.7. Let X be any set, and let $\mathcal{F}(X)$ be the set of all scalar-valued functions on the domain X. That is, if our scalar field is $\mathbb{R}$ then $\mathcal{F}(X)$ consists of all functions $f: X \to \mathbb{R}$, while if the scalar field is $\mathbb{C}$ then $\mathcal{F}(X)$ is the set of all functions $f: X \to \mathbb{C}$.

(a) Prove that $\mathcal{F}(X)$ is a vector space.

(b) For this part we take $X = \{1, \dots, d\}$. Assuming the scalar field is $\mathbb{R}$, explain why the Euclidean vector space $\mathbb{R}^d$ and the vector space $\mathcal{F}(\{1, \dots, d\})$ are really the "same space," in the sense that each vector in $\mathbb{R}^d$ naturally corresponds to a function in $\mathcal{F}(\{1, \dots, d\})$, with the operations in $\mathbb{R}^d$ being the "same" as the operations in $\mathcal{F}(\{1, \dots, d\})$. Then, assuming the scalar field is $\mathbb{C}$, formulate an analogous statement for $\mathbb{C}^d$.

(c) Now let $X = \mathbb{N} = \{1, 2, 3, \dots\}$, and let $\mathcal{S}$ be the set of all infinite sequences $x = (x_1, x_2, \dots)$. In what sense are $\mathcal{F}(\mathbb{N})$ and $\mathcal{S}$ the "same" vector space?

0.14.8. Let I be an interval in the real line. A scalar-valued function f on I is *bounded* if $\sup_{t \in I} |f(t)| < \infty$.

(a) Let $\mathcal{F}_b(I)$ be the set of all bounded functions on I. Prove that $\mathcal{F}_b(I)$ is a proper subspace of $\mathcal{F}(I)$.

(b) Let $C_b(I)$ be the set of all bounded continuous functions on I. Prove that $C_b(I)$ is a subspace of $C(I)$. Show that if I is any type of interval *other than* a bounded closed interval $[a, b]$, then $C_b(I) \neq C(I)$.

Remark: If $I = [a, b]$ is a bounded closed interval, then every continuous function on I is bounded (for one proof, see [Heil18, Lem. 2.9.6]). Therefore $C_b[a, b] = C[a, b]$.

0.14.9. Let w be the function defined in equation (0.10).

(a) Use the product rule to prove that w is differentiable at every point $x \neq 0$, and use the definition of the derivative to prove that w is differentiable at $x = 0$.

(b) Show that even though $w'(t)$ exists for every t, the derivative w' is not continuous at the origin.

0.15 Span and Independence

We will discuss some vector space concepts that are based on the notion of *finite linear combinations*.

0.15.1 Span

A *finite linear combination* (or simply a *linear combination*, for short) of vectors $x_1, \dots, x_N$ in a vector space V is any vector that has the form

$$x = \sum_{k=1}^{N} c_k x_k = c_1 x_1 + \cdots + c_N x_N,$$

where N is a positive integer and $c_1, \dots, c_N$ are scalars. We collect all of the linear combinations together to form the following set.

Definition 0.15.1 (Span). If A is a nonempty subset of a vector space V, then the *finite linear span* of A, denoted by $\mathrm{span}(A)$, is the set of all finite linear combinations of elements of A:

$$\mathrm{span}(A) = \left\{ \sum_{n=1}^{N} c_n x_n : N > 0,\, x_n \in A,\, c_n \text{ are scalars} \right\}. \tag{0.11}$$

We say that A *spans* V if $\mathrm{span}(A) = V$.

We *declare* that the span of the empty set is $\mathrm{span}(\varnothing) = \{0\}$. ♢

We also refer to $\mathrm{span}(A)$ as the *finite span*, the *linear span*, or simply the *span* of A.

If $A = \{x_1, \dots, x_n\}$ is a finite set, then we usually write $\mathrm{span}\{x_1, \dots, x_n\}$ instead of $\mathrm{span}\big(\{x_1, \dots, x_n\}\big)$. In this case, equation (0.11) simplifies to

$$\mathrm{span}\{x_1, \dots, x_n\} = \left\{ c_1 x_1 + \cdots + c_n x_n : c_1, \dots, c_n \text{ are scalars} \right\}.$$

Similarly, if $A = \{x_n\}_{n \in \mathbb{N}}$ is a sequence, then we usually write $\mathrm{span}\{x_n\}_{n \in \mathbb{N}}$ instead of $\mathrm{span}\big(\{x_n\}_{n \in \mathbb{N}}\big)$, and in this case equation (0.11) simplifies to

$$\mathrm{span}\{x_n\}_{n \in \mathbb{N}} = \left\{ \sum_{n=1}^{N} c_n x_n : N > 0,\, c_n \text{ are scalars} \right\}.$$

Example 0.15.2. In this example we implicitly assume that the domain of our functions is some fixed interval I in the real line. For each integer $k \ge 0$ let p_k be the function defined by the rule

$$p_k(t) = t^k, \qquad \text{for } t \in I.$$

A finite linear combination of the p_k is a function of the form

$$p = \sum_{k=0}^{n} c_k p_k = c_0 p_0 + c_1 p_1 + \cdots + c_n p_n,$$

where $n \ge 0$ and $c_0, c_1, \dots, c_n$ are scalars. Such a function p is given by the rule

$$p(t) = \sum_{k=0}^{n} c_k p_k(t) = \sum_{k=0}^{n} c_k t^k, \qquad \text{for } t \in I.$$

That is, p is a *polynomial*. Earlier we declared that $\mathcal{P}$ denotes the set of all polynomials, so if we let $\mathcal{M} = \{p_k\}_{k \ge 0}$ then we have shown that

$$\operatorname{span}(\mathcal{M}) \;=\; \mathcal{P}.$$

That is, the subset $\mathcal{M}$ spans $\mathcal{P}$. ♢

We were excessively careful in Example 0.15.2 to distinguish between the function, p_k, and its rule, $p_k(t) = t^k$ for $t \in I$. Technically, p_k denotes a function while $p_k(t)$ denotes the evaluation of p_k at the point t and hence is a number. Therefore it is not literally correct to write "the function $p_k(t)$" or to say that "t^k is a vector in $\mathcal{P}$." However, since the meaning is clear, we typically abuse notation and simply write "the function t^k" or "the vector t^k" instead of "the function p_k given by the rule $p_k(t) = t^k$ for $t \in I$." We will abuse notation in this way many times below, beginning with the following remark.

Remark 0.15.3. A *monomial* is a polynomial that has only one nonzero term, i.e., it is a polynomial of the form ct^k where $c \neq 0$ and $k \ge 0$. Thus the collection $\mathcal{M}$ that is discussed in Example 0.15.2 is the set of all monomials t^k with $k \ge 0$. We often summarize Example 0.15.2 by saying that *the monomials t^k span $\mathcal{P}$*. ♢

0.15.2 Linear Independence

By definition, if $x \in \operatorname{span}(A)$, then x is some finite linear combination of elements of A. In general there could be many different linear combinations that equal the vector x. Often we wish to ensure that x is a unique linear combination of elements of A. This issue is related to the following notion.

Definition 0.15.4 (Linear Independence). A nonempty subset A of a vector space V is *finitely linearly independent* if for each integer $N > 0$, any choice of finitely many distinct vectors $x_1, \dots, x_N \in A$, and any scalars $c_1, \dots, c_N$, we have

$$\sum_{n=1}^{N} c_n x_n \;=\; 0 \quad\Longleftrightarrow\quad c_1 \;=\; \cdots \;=\; c_N \;=\; 0.$$

We *declare* that the empty set $\varnothing$ is a linearly independent set. ♢

Instead of saying that a set A is *finitely linearly independent*, we sometimes abbreviate this to A *is linearly independent*, or even just A *is independent*.

Often the set A is a finite set or a countably infinite sequence. Rewriting the definition for these cases, we see that a finite set $\{x_1, \dots, x_n\}$ is independent if and only if $\sum_{k=1}^{n} c_k x_k = 0$ only for $c_1 = \cdots = c_n = 0$. A countable sequence $\{x_n\}_{n\in\mathbb{N}}$ is independent if and only if for *every integer* $N \in \mathbb{N}$ we have $\sum_{n=1}^{N} c_n x_n = 0$ only when $c_1 = \cdots = c_N = 0$.

Example 0.15.5. Let I be an interval in the real line. As in Example 0.15.2, let $\mathcal{P}$ be the set of all polynomials on I, and let $\mathcal{M} = \{t^k\}_{k=0}^{\infty}$. We will show that $\mathcal{M}$ is a finitely linearly independent subset of $\mathcal{P}$. To do this, choose any integer $N \geq 0$, and suppose that $c_0, c_1, \dots, c_N$ are scalars such that

$$\sum_{k=0}^{N} c_k t^k \;=\; 0. \tag{0.12}$$

Note that the vector on the left-hand side of equation (0.12) is a function, the polynomial $p(t) = \sum_{k=0}^{N} c_k t^k$. The vector 0 on the right-hand side of equation (0.12) is the zero element of the vector space $\mathcal{P}$, which is the zero function. Hence what we are assuming in equation (0.12) is that the polynomial p equals the zero function, i.e., $p(t) = 0$ for *every* $t \in I$ (not just for *some* t).

We wish to show that each scalar c_k in equation (0.12) must be zero. There are many ways to do this. We ask for a *direct* proof in Problem 0.15.12 and spell out an *indirect* proof here. The *Fundamental Theorem of Algebra* tells us that a polynomial of degree N can have at most N roots, i.e., there can be at most N values of t for which $p(t) = 0$. If the scalar c_N is nonzero, then $p(t) = \sum_{k=0}^{N} c_k t^k$ has degree N, and therefore can have at most N roots. But we know that $p(t) = 0$ for every $t \in I$, so p has infinitely many roots. This is a contradiction, so we must have $c_N = 0$.

Since $c_N = 0$, if c_{N-1} is nonzero then p has degree $N-1$, which again leads to a contradiction since $p(t) = 0$ for every t. Therefore we must have $c_{N-1} = 0$.

Continuing in this way, we see that $c_{N-2} = 0$, and so forth, down to $c_1 = 0$. So we are left with $p(t) = c_0$, i.e., p is a constant polynomial. The only way a constant polynomial can equal the zero function is if $c_0 = 0$. Therefore every c_k is zero, so we have shown that $\mathcal{M}$ is linearly independent. $\diamondsuit$

0.15.3 Hamel Bases

A set that is both linearly independent and spans a vector space V is usually called a *basis* for V, but to avoid confusion with other types of bases that arise in analysis (such as the *Schauder bases* discussed in Problem 7.4.11), we will use the following terminology for a set that both spans and is linearly independent.

Definition 0.15.6 (Hamel Basis). Let V be a nontrivial vector space. A set of vectors $\mathcal{B}$ is a *Hamel basis*, *vector space basis*, or simply a *basis* for V if

$$\mathcal{B} \text{ is linearly independent} \qquad \text{and} \qquad \operatorname{span}(\mathcal{B}) = V. \qquad \diamondsuit$$

For example, since Example 0.15.2 shows that the set of monomials $\mathcal{M} = \{t^k\}_{k=0}^{\infty}$ spans the set of polynomials $\mathcal{P}$, and since Example 0.15.5 shows that $\mathcal{M}$ is linearly independent, it follows that $\mathcal{M}$ is a Hamel basis for $\mathcal{P}$.

It can be shown that any two Hamel bases for a given vector space V have the same cardinality. In particular, if V has a Hamel basis that consists of finitely many vectors, say $\mathcal{B} = \{x_1, \dots, x_d\}$, then any other Hamel basis for V must also contain exactly d vectors. We call this number d the *dimension* of V and set

$$\dim(V) \;=\; d.$$

On the other hand, if V has a Hamel basis that consists of infinitely many vectors, then any other Hamel basis must also be infinite. In this case we say that V is *infinite-dimensional* and we set

$$\dim(V) \;=\; \infty.$$

Remark 0.15.7. Instead of simply setting $\dim(V) = \infty$ for all infinite-dimensional vector spaces, we could let $\dim(V)$ denote the actual cardinality of a basis for V. This would allow us to distinguish between vector spaces whose dimension is countably infinite and those whose dimension is uncountable, or even to distinguish further among the uncountable-dimensional spaces according to cardinality. However, that will not be necessary in this text. A more important way to distinguish between different types of “large” spaces will be the notions of *separability* and *nonseparability* introduced in Definition 1.1.5 and studied in detail in Chapters 7 and 8. $\diamondsuit$

For finite-dimensional vector spaces, we have the following characterization of Hamel bases.

Exercise 0.15.8. A set of vectors $\mathcal{B} = \{x_1, \dots, x_d\}$ is a Hamel basis for a nontrivial finite-dimensional vector space V if and only if each vector $x \in V$ can be written as

$$x \;=\; \sum_{n=1}^{d} c_n(x)\, x_n$$

for a *unique choice* of scalars $c_1(x), \dots, c_d(x)$. $\diamondsuit$

The trivial vector space $\{0\}$ is a bit of an anomaly in this discussion, since it does not contain any linearly independent subsets and therefore does not contain a Hamel basis. To handle this case we *declare* that the empty set $\varnothing$ is a Hamel basis for the trivial vector space $\{0\}$, and we set

$$\dim(\{0\}) \;=\; 0.$$

Problems

0.15.9. Let $e_k = (0, \dots, 0, 1, 0, \dots, 0)$ be the vector that has a 1 in the kth component and zeros elsewhere. Prove that $\{e_1, \dots, e_d\}$ is a Hamel basis for $\mathbb{R}^d$ (if the scalar field is $\mathbb{R}$) or $\mathbb{C}^d$ (if the scalar field is $\mathbb{C}$). This is called the *standard basis* for $\mathbb{R}^d$ or $\mathbb{C}^d$.

0.15.10. Let V be a vector space.

(a) Show that if $A \subseteq B \subseteq V$, then $\operatorname{span}(A) \subseteq \operatorname{span}(B)$.

(b) Show by example that $\operatorname{span}(A) = \operatorname{span}(B)$ is possible even if $A \neq B$. Can you find an example where A and B are both infinite?

0.15.11. Given a linearly independent set A in a vector space V, prove the following statements.

(a) If $B \subseteq A$, then B is linearly independent.

(b) If $x \in V$ but $x \notin \operatorname{span}(A)$, then $A \cup \{x\}$ is linearly independent.

(c) If there is no vector $x \in V$ such that $A \cup \{x\}$ is linearly independent, then A is a basis for V.

0.15.12. Let I be an interval in the real line. Without invoking the Fundamental Theorem of Algebra, give a direct proof that the set of monomials $\{t^k\}_{k=0}^{\infty}$ is a linearly independent set of functions in $C(I)$.

0.15.13. For each $k \in \mathbb{N}$, define $e_k(t) = e^{kt}$ for $t \in \mathbb{R}$.

(a) Prove that $\{e_k\}_{k\in\mathbb{N}}$ is a linearly independent set in $C(\mathbb{R})$.

(b) Let 1 be the constant function, i.e., the function that takes the value 1 at every t. Prove that 1 does not belong to $\operatorname{span}\{e_k\}_{k\in\mathbb{N}}$.

(c) Find a function f that does not belong to $\operatorname{span}\{e_k\}_{k\geq 0}$, where $e_0 = 1$.

0.15.14. Let $\mathcal{B} = \{x_j\}_{j\in J}$ be a subset of a vector space V, where J is some fixed index set. Prove that $\mathcal{B}$ is a Hamel basis for V if and only if every nonzero vector $x \in V$ can be written as

$$x \;=\; \sum_{k=1}^{N} c_k(x)\, x_{j_k}$$

for a unique integer $N \in \mathbb{N}$, indices $j_1, \dots, j_N \in J$, and nonzero scalars $c_1(x), \dots, c_N(x)$.

0.15.15. For each $r > 0$, let g_r be the function $g_r(x) = e^{rx}$. Let A be the set of all such functions, i.e., $A = \{g_r : r > 0\}$.

(a) Prove that A is a uncountable subset of $C(\mathbb{R})$ that is linearly independent.

(b) Let 1 be the constant function, i.e., $1(x) = 1$ for all x. Prove that 1 does not belong to $\text{span}(A)$, and therefore A is not a basis for $C(\mathbb{R})$.