Chapter A

Preliminaries of Real Analysis

A principal objective of this largely rudimentary chapter is to introduce the basic set-theoretical nomenclature that we adopt throughout the text. We start with an intuitive discussion of the notion of set, and then introduce the basic operations on sets, Cartesian products, and binary relations. After a quick excursion to order theory (in which the only relatively advanced topic that we cover is the completion of a partial order), functions are introduced as special cases of binary relations and sequences as special cases of functions. Our coverage of abstract set theory concludes with a brief discussion of the Axiom of Choice and the proof of Szpilrajn's Theorem on the completion of a partial order.

We assume here that the reader is familiar with the elementary properties of the real numbers and thus provide only a heuristic discussion of the basic number systems. No construction for the integers is given, in particular. After a short elaboration on ordered fields and the Completeness Axiom, we note without proof that the rational numbers form an ordered field and the real numbers form a complete ordered field. The related discussion is intended to be read more quickly than anywhere else in the text.

We next turn to real sequences. These we discuss relatively thoroughly because of the important role they play in real analysis. In particular, even though our coverage will serve only as a review for most readers, we study here the monotonic sequences and subsequential limits with some care, and prove a few useful results, such as the Bolzano-Weierstrass Theorem and Dirichlet's Rearrangement Theorem. These results will be used freely in the remainder of the book.

The final section of the chapter is nothing more than a swift refresher on the analysis of real functions. First we recall some basic definitions, and then, very quickly, we go over the concepts of limits and continuity of real functions defined on the real line. We then review the elementary theory of differentiation for single-variable functions, mostly through exercises. The primer we present on Riemann integration is a bit more leisurely.

In particular, we give a complete proof of the Fundamental Theorem of Calculus, which is used in the remainder of the book freely. We invoke our calculus review also to outline a basic analysis of exponential and logarithmic real functions. These maps are used in many examples throughout the book. The chapter concludes with a brief discussion of the theory of concave functions on the real line.

1 Elements of Set Theory

1.1 Sets

Intuitively speaking, a "set" is a collection of objects.[1] The distinguishing feature of a set is that whereas it may contain numerous objects, it is nevertheless conceived as a single entity. In the words of Georg Cantor, the great founder of abstract set theory, "a set is a Many which allows itself to be thought of as a One." It is amazing how much follows from this simple idea.

The objects that a set S contains are called the "elements" (or "members") of S. Clearly, to know S, it is necessary and sufficient to know all elements of S. The principal concept of set theory, then, is the relation of "being an element/member of." The universally accepted symbol for this relation is $\in$; that is, $x \in S$ (or $S \ni x$) means that "x is an element of S" (also read "x is a member of S," or "x is contained in S," or "x belongs to S," or "x is in S," or "S includes x," etc.). We often write $x, y \in S$ to denote that both $x \in S$ and $y \in S$ hold. For any natural number m, a statement like $x_1, \ldots, x_m \in S$ (or equivalently, $x_i \in S$, $i = 1, \ldots, m$) is understood analogously. If $x \in S$ is a false statement, then we write $x \notin S$, and read "x is not an element of S."

If the sets A and B have exactly the same elements, that is, $x \in A$ iff $x \in B$, then we say that A and B are identical, and write $A = B$; otherwise we write $A \neq B$.[2] (So, for instance, $\{x, y\} = \{y, x\}$, $\{x, x\} = \{x\}$, and $\{\{x\}\} \neq \{x\}$.) If every member of A is also a member of B, then we say that A is a **subset** of B (also read "A is a set in B" or "A is contained in B") and write $A \subseteq B$ (or $B \supseteq A$). Clearly, $A = B$ holds iff both $A \subseteq B$ and $B \subseteq A$ hold. If $A \subseteq B$

[1] The notion of "object" is left undefined, that is, it can be given any meaning. All I demand of our "objects" is that they be logically *distinguishable*. That is, if x and y are two objects, $x = y$ and $x \neq y$ cannot hold simultaneously, and the statement "either $x = y$ or $x \neq y$" is a tautology.

[2] *Reminder.* iff = if and only if.

but $A \neq B$, then A is said to be a **proper subset** of B, and we denote this situation by writing $A \subset B$ (or $B \supset A$).

For any set S that contains finitely many elements (in which case we say S is **finite**), we denote by $|S|$ the total number of elements that S contains, and refer to this number as the **cardinality** of S. We say that S is a **singleton** if $|S| = 1$. If S contains infinitely many elements (in which case we say S is **infinite**), then we write $|S| = \infty$. Obviously, we have $|A| \leq |B|$ whenever $A \subseteq B$, and if $A \subset B$ and $|A| < \infty$, then $|A| < |B|$.

We sometimes specify a set by enumerating its elements. For instance, $\{x, y, z\}$ is the set that consists of the objects x, y, and z. The contents of the sets $\{x_1, \ldots, x_m\}$ and $\{x_1, x_2, \ldots\}$ are similarly described. For example, the set $\mathbb{N}$ of positive integers can be written as $\{1, 2, \ldots\}$. Alternatively, one may describe a set S as a collection of all objects x that satisfy a given property P. If $P(x)$ stands for the (logical) statement "x satisfies the property P," then we can write $S = \{x : P(x) \text{ is a true statement}\}$, or simply $S = \{x : P(x)\}$. If A is a set and B is the set that contains all elements x of A such that $P(x)$ is true, we write $B = \{x \in A : P(x)\}$. For instance, where $\mathbb{R}$ is the set of all real numbers, the collection of all real numbers greater than or equal to 3 can be written as $\{x \in \mathbb{R} : x \geq 3\}$.

The symbol $\emptyset$ denotes the **empty set**, that is, the set that contains no elements (i.e., $|\emptyset| = 0$). Formally speaking, we can define $\emptyset$ as the set $\{x : x \neq x\}$, for this description entails that $x \in \emptyset$ is a false statement for any object x. Consequently, we write

$$\emptyset := \{x : x \neq x\},$$

meaning that the symbol on the left-hand side is *defined* by that on the right-hand side.[3] Clearly, we have $\emptyset \subseteq S$ for any set S, which in particular implies that $\emptyset$ is unique. (Why?) If $S \neq \emptyset$, we say that S is **nonempty**. For instance, $\{\emptyset\}$ is a nonempty set. Indeed, $\{\emptyset\} \neq \emptyset$—the former, after all, is a set of sets that contains the empty set, while $\emptyset$ contains nothing. (An empty box is not the same thing as nothing!)

We define the class of all subsets of a given set S as

$$2^S := \{T : T \subseteq S\},$$

[3] Recall my notational convention: For any symbols $\clubsuit$ and $\heartsuit$, either one of the expressions $\clubsuit := \heartsuit$ and $\heartsuit =: \clubsuit$ means that $\clubsuit$ is defined by $\heartsuit$.

which is called the **power set** of S. (The choice of notation is motivated by the fact that the power set of a set that contains m elements has exactly 2^m elements.) For instance, $2^{\emptyset} = \{\emptyset\}$, $2^{2^{\emptyset}} = \{\emptyset, \{\emptyset\}\}$, and $2^{2^{2^{\emptyset}}} = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}$, and so on.

Notation. Throughout this text, the class of all nonempty *finite* subsets of any given set S is denoted by $\mathcal{P}(S)$, that is,

$$\mathcal{P}(S) := \{T : T \subseteq S \text{ and } 0 < |T| < \infty\}.$$

Of course, if S is finite, then $\mathcal{P}(S) = 2^S \backslash \{\emptyset\}$.

Given any two sets A and B, by $A \cup B$ we mean the set $\{x : x \in A$ or $x \in B\}$, which is called the **union** of A and B. The **intersection** of A and B, denoted as $A \cap B$, is defined as the set $\{x : x \in A$ and $x \in B\}$. If $A \cap B = \emptyset$, we say that A and B are **disjoint**. Obviously, if $A \subseteq B$, then $A \cup B = B$ and $A \cap B = A$. In particular, $\emptyset \cup S = S$ and $\emptyset \cap S = \emptyset$ for any set S.

Taking unions and intersections are *commutative* operations in the sense that

$$A \cap B = B \cap A \quad \text{and} \quad A \cup B = B \cup A$$

for any sets A and B. They are also *associative*, that is,

$$A \cap (B \cap C) = (A \cap B) \cap C \quad \text{and} \quad A \cup (B \cup C) = (A \cup B) \cup C,$$

and *distributive*, that is,

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{and} \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$

for any sets A, B, and C.

Exercise 1 Prove the commutative, associative, and distributive laws of set theory stated above.

$\boxed{\text{Exercise 2}}$ Given any two sets A and B, by $A \backslash B$—the **difference** between A and B—we mean the set $\{x : x \in A$ and $x \notin B\}$.

(a) Show that $S \backslash \emptyset = S$, $S \backslash S = \emptyset$, and $\emptyset \backslash S = \emptyset$ for any set S.

(b) Show that $A \backslash B = B \backslash A$ iff $A = B$ for any sets A and B.

(c) (*De Morgan Laws*) Prove: For any sets A, B, and C,

$$A \backslash (B \cup C) = (A \backslash B) \cap (A \backslash C) \quad \text{and} \quad A \backslash (B \cap C) = (A \backslash B) \cup (A \backslash C).$$

Throughout this text we use the terms "class" or "family" only to refer to a nonempty collection of sets. So if $\mathcal{A}$ is a class, we understand that $\mathcal{A} \neq \emptyset$ and that any member $A \in \mathcal{A}$ is a set (which may itself be a collection of sets). The union of all members of this class, denoted as $\cup\mathcal{A}$, or $\cup\{A : A \in \mathcal{A}\}$, or $\cup_{A\in\mathcal{A}}A$, is defined as the set $\{x : x \in A \text{ for some } A \in \mathcal{A}\}$. Similarly, the intersection of all sets in $\mathcal{A}$, denoted as $\cap\mathcal{A}$, or $\cap\{A : A \in \mathcal{A}\}$, or $\cap_{A\in\mathcal{A}}A$, is defined as the set $\{x : x \in A \text{ for each } A \in \mathcal{A}\}$.

A common way of specifying a class $\mathcal{A}$ of sets is by designating a set I as a set of indices and by defining $\mathcal{A} := \{A_i : i \in I\}$. In this case, $\cup\mathcal{A}$ may be denoted as $\cup_{i\in I}A_i$. If $I = \{k, k+1, \ldots, K\}$ for some integers k and K with $k < K$, then we often write $\cup_{i=k}^{K}A_i$ (or $A_k \cup \cdots \cup A_K$) for $\cup_{i\in I}A_i$. Similarly, if $I = \{k, k+1, \ldots\}$ for some integer k, then we may write $\cup_{i=k}^{\infty}A_i$ (or $A_k \cup A_{k+1} \cup \cdots$) for $\cup_{i\in I}A_i$. Furthermore, for brevity, we frequently denote $\cup_{i=1}^{K}A_i$ as $\cup^{K}A_i$, and $\cup_{i=1}^{\infty}A_i$ as $\cup^{\infty}A_i$, throughout the text. Similar notational conventions apply to intersections of sets as well.

Warning. The symbols $\cup\emptyset$ and $\cap\emptyset$ are left *undefined* (in much the same way that the symbol $\frac{0}{0}$ is undefined in number theory).

Exercise 3 Let A be a set and $\mathcal{B}$ a class of sets. Prove that

$$A\cap\cup\mathcal{B} = \bigcup\{A\cap B : B \in \mathcal{B}\} \quad \text{and} \quad A\cup\cap\mathcal{B} = \bigcap\{A\cup B : B \in \mathcal{B}\},$$

while

$$A\backslash\cup\mathcal{B} = \bigcap\{A\backslash B : B \in \mathcal{B}\} \quad \text{and} \quad A\backslash\cap\mathcal{B} = \bigcup\{A\backslash B : B \in \mathcal{B}\}.$$

A word of caution may be in order before we proceed further. While duly intuitive, the "set theory" we have outlined so far provides us with no demarcation criterion for identifying what exactly constitutes a set. This may suggest that one is completely free in deeming any given collection of objects a set. But in fact, this would be a pretty bad idea that would entail serious foundational difficulties. The best known example of such difficulties was given by Bertrand Russell in 1902 when he asked if the set of all objects that are not members of *themselves* is a set: Is $S := \{x : x \notin x\}$ a set?[4] There is

[4] While a bit unorthodox, $x \in x$ may well be a statement that is true for some objects. For instance, the collection of all sets that I have mentioned in my life, say x, is a set that I have just mentioned, so $x \in x$. But the collection of all cheesecakes I have eaten in my life, say y, is not a cheesecake, so $y \notin y$.

nothing in our intuitive discussion above that forces us to conclude that S is not a set; it is a collection of objects (sets in this case) that is considered as a single entity. But we cannot accept S as a set, for if we do, we have to be able to answer the question, Is $S \in S$? If the answer is yes, then $S \in S$, but this implies $S \notin S$ by definition of S. If the answer is no, then $S \notin S$, but this implies $S \in S$ by definition of S. That is, we have a contradictory state of affairs no matter what! This is the so-called *Russell's paradox*, which started a severe foundational crisis for mathematics that eventually led to a complete axiomatization of set theory in the early twentieth century.[5]

Roughly speaking, this paradox would arise only if we allowed "unduly large" collections to be qualified as sets. In particular, it will not cause any harm for the mathematical analysis that will concern us here, precisely because in all of our discussions, we will *fix* a universal set of objects, say X, and consider sets like $\{x \in X : P(x)\}$, where $P(x)$ is an unambiguous logical statement in terms of x. (We will also have occasion to work with sets of such sets, and sets of sets of such sets, and so on.) Once such a domain X is fixed, Russell's paradox cannot arise. Why, you may ask, can't we have the same problem with the set $S := \{x \in X : x \notin x\}$? No, because now we can answer the question: Is $S \in S$? The answer is no! The statement $S \in S$ is false, simply because $S \notin X$. (For, if $S \in X$ was the case, then we would end up with the contradiction $S \in S$ iff $S \notin S$.)

So when the context is clear (that is, when a universe of objects is fixed), and when we define our sets as just explained, Russell's paradox will not be a threat against the resulting set theory. But can there be any other paradoxes? Well, there is really not an easy answer to this. To even discuss the matter unambiguously, we must leave our intuitive understanding of the notion of set and address the problem through a completely axiomatic approach (in which we would leave the expression "$x \in S$" undefined and give meaning to it *only* through axioms). This is, of course, not at all the place to do this. Moreover, the "intuitive" set theory that we covered here is more than enough for the mathematical analysis to come. We thus leave this topic by

[5] Russell's paradox is a classic example of the dangers of using *self-referential* statements carelessly. Another example of this form is the ancient *paradox of the liar*: "Everything I say is false." This statement can be declared neither true nor false! To get a sense of some other kinds of paradoxes and the way axiomatic set theory avoids them, you might want to read the popular account of Rucker (1995).

referring the reader who wishes to get a broader introduction to abstract set theory to Chapter 1 of Schechter (1997) or Marek and Mycielski (2001); both of these expositions provide nice introductory overviews of axiomatic set theory. If you want to dig deeper, then try the first three chapters of Enderton (1977).

1.2 Relations

An **ordered pair** is an ordered list (a, b) consisting of two objects a and b. This list is *ordered* in the sense that, as a defining feature of the notion of ordered pair, we assume the following: For any two ordered pairs (a, b) and (a', b'), we have $(a, b) = (a', b')$ iff $a = a'$ and $b = b'$.[6]

The **(Cartesian) product** of two nonempty sets A and B, denoted as $A \times B$, is defined as the set of all ordered pairs (a, b) where a comes from A and b comes from B. That is,

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

As a notational convention, we often write A^2 for $A \times A$. It is easily seen that taking the Cartesian product of two sets is not a commutative operation. Indeed, for any two *distinct* objects a and b, we have $\{a\} \times \{b\} = \{(a, b)\} \neq \{(b, a)\} = \{b\} \times \{a\}$. Formally speaking, it is not associative either, for $(a, (b, c))$ is not the same thing as $((a, b), c)$. Yet there is a natural correspondence between the elements of $A \times (B \times C)$ and $(A \times B) \times C$, so one can really think of these two sets as the same, thereby rendering the status of the set $A \times B \times C$ unambiguous.[7] This prompts us to define an ***n*-vector**

[6] This defines the notion of ordered pair as a new "primitive" for our set theory, but in fact, this is not really necessary. One can define an ordered pair by using only the concept of "set" as $(a, b) := \{\{a\}, \{a, b\}\}$. With this definition, which is due to Kazimierz Kuratowski, one can *prove* that, for any two ordered pairs (a, b) and (a', b'), we have $(a, b) = (a', b')$ iff $a = a'$ and $b = b'$. The "if" part of the claim is trivial. To prove the "only if" part, observe that $(a, b) = (a', b')$ entails that either $\{a\} = \{a'\}$ or $\{a\} = \{a', b'\}$. But the latter equality may hold only if $a = a' = b'$, so we have $a = a'$ in all contingencies. Therefore, $(a, b) = (a', b')$ entails that either $\{a, b\} = \{a\}$ or $\{a, b\} = \{a, b'\}$. The latter case is possible only if $b = b'$, while the former possibility arises only if $a = b$. But if $a = b$, then we have $\{\{a\}\} = (a, b) = (a, b') = \{\{a\}, \{a, b'\}\}$, which holds only if $\{a\} = \{a, b'\}$, that is, $b = a = b'$.

Quiz. (Wiener) Show that we would also have $(a, b) = (a', b')$ iff $a = a'$ and $b = b'$, if we instead defined (a, b) as $\{\{\emptyset, \{a\}\}, \{\{b\}\}\}$.

[7] What is this "natural" correspondence?

(for any natural number n) as a list $(a_1, \ldots, a_n)$, with the understanding that $(a_1, \ldots, a_n) = (a'_1, \ldots, a'_n)$ iff $a_i = a'_i$ for each $i = 1, \ldots, n$. The **(Cartesian) product** of n sets $A_1, \ldots, A_n$, is then defined as

$$A_1 \times \ldots \times A_n := \{(a_1, \ldots, a_n) : a_i \in A_i,\ i = 1, \ldots, n\}.$$

We often write $\mathsf{X}^n A_i$ to denote $A_1 \times \cdots \times A_n$, and refer to $\mathsf{X}^n A_i$ as the **n-fold product** of $A_1, \ldots, A_n$. If $A_i = S$ for each i, we then write S^n for $A_1 \times \ldots \times A_n$, that is, $S^n := \mathsf{X}^n S$.

Exercise 4 For any sets A, B, and C, prove that

$$A \times (B \cap C) = (A \times B) \cap (A \times C) \quad \text{and} \quad A \times (B \cup C) = (A \times B) \cup (A \times C).$$

Let X and Y be two nonempty sets. A subset R of $X \times Y$ is called a (*binary*) **relation from X to Y**. If $X = Y$, that is, if R is a relation from X to X, we simply say that it is a **relation on X**. Put differently, R is a relation on X iff $R \subseteq X^2$. If $(x, y) \in R$, then we think of R as associating the object x with y, and if $\{(x, y), (y, x)\} \cap R = \emptyset$, we understand that there is no connection between x and y as envisaged by R. In concert with this interpretation, we adopt the convention of writing xRy instead of $(x, y) \in R$ throughout this text.

Definition
A relation R on a nonempty set X is said to be **reflexive** if xRx for each $x \in X$, and **complete** if either xRy or yRx holds for each $x, y \in X$. It is said to be **symmetric** if, for any $x, y \in X$, xRy implies yRx, and **antisymmetric** if, for any $x, y \in X$, xRy and yRx imply $x = y$. Finally, we say that R is **transitive** if xRy and yRz imply xRz for any $x, y, z \in X$.

The interpretations of these properties are straightforward, so we do not elaborate on them here. But note: While every complete relation is reflexive, there are no other logical implications between these properties.

Exercise 5 Let X be a nonempty set, and R a relation on X. The **inverse** of R is defined as the relation $R^{-1} := \{(x, y) \in X^2 : yRx\}$.

(a) If R is symmetric, does R^{-1} have to be also symmetric? Antisymmetric? Transitive?
(b) Show that R is symmetric iff $R = R^{-1}$.
(c) If R_1 and R_2 are two relations on X, the **composition** of R_1 and R_2 is the relation $R_2 \circ R_1 := \{(x, y) \in X^2 : xR_1z \text{ and } zR_2y \text{ for some } z \in X\}$. Show that R is transitive iff $R \circ R \subseteq R$.

Exercise 6 A relation R on a nonempty set X is called **circular** if xRz and zRy imply yRx for every $x, y, z \in X$. Prove that R is reflexive and circular iff it is reflexive, symmetric, and transitive.

$\boxed{\text{Exercise 7}}^{H}$ Let R be a reflexive relation on a nonempty set X. The **asymmetric part** of R is defined as the relation P_R on X as xP_Ry iff xRy but not yRx. The relation $I_R := R \backslash P_R$ on X is then called the **symmetric part** of R.
(a) Show that I_R is reflexive and symmetric.
(b) Show that P_R is neither reflexive nor symmetric.
(c) Show that if R is transitive, so are P_R and I_R.

$\boxed{\text{Exercise 8}}$ Let R be a relation on a nonempty set X. Let $R_0 = R$, and for each positive integer m, define the relation R_m on X by xR_my iff there exist $z_1, \ldots, z_m \in X$ such that xRz_1, $z_1Rz_2, \ldots,$ and z_mRy. The relation $tr(R) := R_0 \cup R_1 \cup \cdots$ is called the **transitive closure** of R. Show that $tr(R)$ is transitive, and if R' is a transitive relation with $R \subseteq R'$, then $tr(R) \subseteq R'$.

1.3 Equivalence Relations

In mathematical analysis, one often needs to "identify" two distinct objects when they possess a particular property of interest. Naturally, such an identification scheme should satisfy certain consistency conditions. For instance, if x is identified with y, then y must be identified with x. Similarly, if x and y are deemed identical, and so are y and z, then x and z should be identified. Such considerations lead us to the notion of *equivalence relation*.

Definition
A relation $\sim$ on a nonempty set X is called an **equivalence relation** if it is reflexive, symmetric, and transitive. For any $x \in X$, the **equivalence class** of x relative to $\sim$ is defined as the set

$$[x]_\sim := \{y \in X : y \sim x\}.$$

The class of all equivalence classes relative to $\sim$, denoted as $X/_\sim$, is called the **quotient set** of X relative to $\sim$, that is,

$$X/_\sim := \{[x]_\sim : x \in X\}.$$

Let X denote the set of all people in the world. "Being a sibling of" is an equivalence relation on X (provided that we adopt the convention of saying that any person is a sibling of himself). The equivalence class of a person relative to this relation is the set of all of his or her siblings. On the other hand, you would probably agree that "being in love with" is not an equivalence relation on X. Here are some more examples (that fit better with the "serious" tone of this course).

Example 1

[1] For any nonempty set X, the **diagonal relation** $D_X := \{(x, x) : x \in X\}$ is the smallest equivalence relation that can be defined on X (in the sense that if R is any other equivalence relation on X, we have $D_X \subseteq R$). Clearly, $[x]_{D_X} = \{x\}$ for each $x \in X$.[8] At the other extreme is X^2 which is the largest equivalence relation that can be defined on X. We have $[x]_{X^2} = X$ for each $x \in X$.

[2] By Exercise 7, the symmetric part of any reflexive and transitive relation on a nonempty set is an equivalence relation.

[3] Let $X := \{(a, b) : a, b \in \{1, 2, \ldots\}\}$, and define the relation $\sim$ on X by $(a, b) \sim (c, d)$ iff $ad = bc$. It is readily verified that $\sim$ is an equivalence relation on X, and that $[(a, b)]_\sim = \left\{(c, d) \in X : \frac{c}{d} = \frac{a}{b}\right\}$ for each $(a, b) \in X$.

[4] Let $X := \{\ldots, -1, 0, 1, \ldots\}$, and define the relation $\sim$ on X by $x \sim y$ iff $\frac{1}{2}(x - y) \in X$. It is easily checked that $\sim$ is an equivalence relation

[8] I say an equally suiting name for D_X is the "equality relation." What do you think?

on X. Moreover, for any integer x, we have $x \sim y$ iff $y = x - 2m$ for some $m \in X$, and hence $[x]_\sim$ equals the set of all even integers if x is even, and that of all odd integers if x is odd. □

One typically uses an equivalence relation to simplify a situation in a way that all things that are indistinguishable from a particular perspective are put together in a set and treated as if they were a single entity. For instance, suppose that for some reason we are interested in the signs of people. Then, any two individuals who are of the same sign can be thought of as "identical," so instead of the set of all people in the world, we would rather work with the set of all Capricorns, all Virgos, and so on. But the set of all Capricorns is of course none other than the equivalence class of any given Capricorn person relative to the equivalence relation of "being of the same sign." So when someone says "a Capricorn is. . . ," then one is really referring to a whole class of people. The equivalence relation of "being of the same sign" divides the world into twelve equivalence classes, and we can then talk "as if" there were only twelve individuals in our context of reference.

To take another example, ask yourself how you would define the set of positive rational numbers, given the set of natural numbers $\mathbb{N} := \{1, 2, \ldots\}$ and the operation of "multiplication." Well, you may say, a positive rational number is the ratio of two natural numbers. But wait, what is a "ratio"? Let us be a bit more careful about this. A better way of looking at things is to say that a positive rational number is an ordered pair $(a, b) \in \mathbb{N}^2$, although in daily practice, we write $\frac{a}{b}$ instead of (a, b). Yet we don't want to say that each ordered pair in $\mathbb{N}^2$ is a distinct rational number. (We would like to think of $\frac{1}{2}$ and $\frac{2}{4}$ as the same number, for instance.) So we "identify" all those ordered pairs that we wish to associate with a single rational number by using the equivalence relation $\sim$ introduced in Example 1.[3], and then define a rational number simply as an equivalence class $[(a, b)]_\sim$. Of course, when we talk about rational numbers in daily practice, we simply talk of a fraction like $\frac{1}{2}$, not $[(1, 2)]_\sim$, even though, formally speaking, what we really mean is $[(1, 2)]_\sim$. The equality $\frac{1}{2} = \frac{2}{4}$ is obvious, precisely because the rational numbers are constructed as equivalence classes such that $(2, 4) \in [(1, 2)]_\sim$.

This discussion suggests that an equivalence relation can be used to decompose a grand set of interest into subsets such that the members of the same subset are thought of as "identical," while the members of distinct subsets are viewed as "distinct." Let us now formalize this intuition. By a **partition** of a nonempty set X, we mean a class of pairwise disjoint, nonempty subsets of X whose union is X. That is, $\mathcal{A}$ is a partition of X iff $\mathcal{A} \subseteq 2^X \backslash \{\emptyset\}$, $\cup\mathcal{A} = X$ and $A \cap B = \emptyset$ for every distinct A and B in $\mathcal{A}$. The next result says that the set of equivalence classes induced by any equivalence relation on a set is a partition of that set.

PROPOSITION 1
For any equivalence relation $\sim$ on a nonempty set X, the quotient set $X/_\sim$ is a partition of X.

PROOF
Take any nonempty set X and an equivalence relation $\sim$ on X. Since $\sim$ is reflexive, we have $x \in [x]_\sim$ for each $x \in X$. Thus any member of $X/_\sim$ is nonempty, and $\cup\{[x]_\sim : x \in X\} = X$. Now suppose that $[x]_\sim \cap [y]_\sim \neq \emptyset$ for some $x, y \in X$. We wish to show that $[x]_\sim = [y]_\sim$. Observe first that $[x]_\sim \cap [y]_\sim \neq \emptyset$ implies $x \sim y$. (Indeed, if $z \in [x]_\sim \cap [y]_\sim$, then $x \sim z$ and $z \sim y$ by symmetry of $\sim$, so we get $x \sim y$ by transitivity of $\sim$.) This implies that $[x]_\sim \subseteq [y]_\sim$, because if $w \in [x]_\sim$, then $w \sim x$ (by symmetry of $\sim$), and hence $w \sim y$ by transitivity of $\sim$. The converse containment is proved analogously. ■

The following exercise shows that the converse of Proposition 1 also holds. Thus the notions of equivalence relation and partition are really two different ways of looking at the same thing.

EXERCISE 9 Let $\mathcal{A}$ be a partition of a nonempty set X, and consider the relation $\sim$ on X defined by $x \sim y$ iff $\{x, y\} \subseteq A$ for some $A \in \mathcal{A}$. Prove that $\sim$ is an equivalence relation on X.

1.4 Order Relations

Transitivity property is the defining feature of any *order* relation. Such relations are given various names depending on the properties they possess in addition to transitivity.

Definition

A relation $\succsim$ on a nonempty set X is called a **preorder** on X if it is transitive and reflexive. It is said to be a **partial order** on X if it is an antisymmetric preorder on X. Finally, $\succsim$ is called a **linear order** on X if it is a partial order on X that is complete.

By a **preordered set** we mean a list $(X, \succsim)$ where X is a nonempty set and $\succsim$ is a preorder on X. If $\succsim$ is a partial order on X, then $(X, \succsim)$ is called a **poset** (short for *partially ordered set*), and if $\succsim$ is a linear order on X, then $(X, \succsim)$ is called either a **chain** or a **loset** (short for *linearly ordered set*).

It is convenient to talk as if a preordered set $(X, \succsim)$ were indeed a set when referring to properties that apply only to X. For instance, by a "finite preordered set," we understand a preordered set $(X, \succsim)$ with $|X| < \infty$. Or, when we say that Y is a subset of the preordered set $(X, \succsim)$, we mean simply that $Y \subseteq X$. A similar convention applies to posets and losets as well.

Notation. Let $(X, \succsim)$ be a preordered set. Unless otherwise is stated explicitly, we denote by $\succ$ the asymmetric part of $\succsim$ and by $\sim$ the symmetric part of $\succsim$ (Exercise 7).

The main distinction between a preorder and a partial order is that the former may have a large symmetric part, while the symmetric part of the latter must equal the diagonal relation. As we shall see, however, in most applications this distinction is immaterial.

Example 2

[1] For any nonempty set X, the diagonal relation $D_X := \{(x, x) : x \in X\}$ is a partial order on X. In fact, this relation is the only partial order on X that is also an equivalence relation. (Why?) The relation X^2, on the other hand, is a complete preorder, which is not antisymmetric unless X is a singleton.

[2] For any nonempty set X, the equality relation $=$ and the subsethood relation $\supseteq$ are partial orders on 2^X. The equality relation is not linear, and $\supseteq$ is not linear unless X is a singleton.

[3] $(\mathbb{R}^n, \geq)$ is a poset for any positive integer n, where $\geq$ is defined coordinatewise, that is, $(x_1, \ldots, x_n) \geq (y_1, \ldots, y_n)$ iff $x_i \geq y_i$ for each

$i = 1, \ldots, n$. When we talk of $\mathbb{R}^n$ without specifying explicitly an alternative order, we always have in mind this partial order (which is sometimes called the **natural** (or **canonical**) **order** of $\mathbb{R}^n$). Of course, $(\mathbb{R}, \geq)$ is a loset.

[4] Take any positive integer n and preordered sets $(X_i, \succsim_i)$, $i = 1, \ldots, n$. The **product** of the preordered sets $(X_i, \succsim_i)$, denoted as $\boxtimes^n(X_i, \succsim_i)$, is the preordered set $(X, \succsim)$ with $X := \mathsf{X}^n X_i$ and

$$(x_1, \ldots, x_n) \succsim (y_1, \ldots, y_n) \quad \text{iff} \quad x_i \succsim_i y_i \quad \text{for all } i = 1, \ldots, n.$$

In particular, $(\mathbb{R}^n, \geq) = \boxtimes^n(\mathbb{R}, \geq)$. □

Example 3

In individual choice theory, a **preference relation** $\succsim$ on a nonempty alternative set X is defined as a preorder on X. Here the reflexivity is a trivial condition to require, and transitivity is viewed as a fundamental rationality postulate. (We will talk more about this in Section B.4.) The **strict preference relation** $\succ$ is defined as the asymmetric part of $\succsim$ (Exercise 7). This relation is transitive but not reflexive. The **indifference relation** $\sim$ is then defined as the symmetric part of $\succsim$, and is easily checked to be an equivalence relation on X. For any $x \in X$, the equivalence class $[x]_\sim$ is called in this context the **indifference class** of x, and is simply a generalization of the familiar concept of "*the indifference curve that passes through x*." In particular, Proposition 1 says that no two distinct indifference sets can have a point in common. (This is the gist of the fact that "distinct indifference curves cannot cross!")

In social choice theory, one often works with multiple (complete) preference relations on a given alternative set X. For instance, suppose that there are n individuals in the population, and $\succsim_i$ stands for the preference relation of the ith individual. The **Pareto dominance** relation $\succsim$ on X is defined as $x \succsim y$ iff $x \succsim_i y$ for each $i = 1, \ldots, n$. This relation is a preorder on X in general, and a partial order on X if each $\succsim_i$ is antisymmetric. □

Let $(X, \succsim)$ be a preordered set. By an **extension** of $\succsim$ we understand a preorder $\trianglerighteq$ on X such that $\succsim \subseteq \trianglerighteq$ and $\succ \subseteq \triangleright$, where $\triangleright$ is the asymmetric part of $\trianglerighteq$. Intuitively speaking, an extension of a preorder is "more complete" than the original relation in the sense that it allows one to compare more elements, but it certainly agrees exactly with the original relation when the latter applies. If $\trianglerighteq$ is a partial order, then it is an extension of $\succsim$ iff $\succsim \subseteq \trianglerighteq$. (Why?)

A fundamental result of order theory says that every partial order can be extended to a linear order, that is, for every poset $(X, \succsim)$ there is a loset $(X, \trianglerighteq)$ with $\succsim \subseteq \trianglerighteq$. Although it is possible to prove this by mathematical induction when X is finite, the proof in the general case is built on a relatively advanced method that we will cover later in the course. Relegating its proof to Section 1.7, we only state here the result for future reference.[9]

Szpilrajn's Theorem
Every partial order on a nonempty set X can be extended to a linear order on X.

A natural question is whether the same result holds for preorders as well. The answer is yes, and the proof follows easily from Szpilrajn's Theorem by means of a standard method.

Corollary 1
Let $(X, \succsim)$ be a preordered set. There exists a complete preorder on X that extends $\succsim$.

Proof
Let $\sim$ denote the symmetric part of $\succsim$, which is an equivalence relation. Then $(X/_\sim, \succsim^*)$ is a poset where $\succsim^*$ is defined on $X/_\sim$ by

$$[x]_\sim \succsim^* [y]_\sim \quad \text{if and only if} \quad x \succsim y.$$

By Szpilrajn's Theorem, there exists a linear order $\trianglerighteq^*$ on $X/_\sim$ such that $\succsim^* \subseteq \trianglerighteq^*$. We define $\trianglerighteq$ on X by

$$x \trianglerighteq y \quad \text{if and only if} \quad [x]_\sim \trianglerighteq^* [y]_\sim.$$

It is easily checked that $\trianglerighteq$ is a complete preorder on X with $\succsim \subseteq \trianglerighteq$ and $\succ \subseteq \triangleright$, where $\succ$ and $\triangleright$ are the asymmetric parts of $\succsim$ and $\trianglerighteq$, respectively. ■

[9] For an extensive introduction to the theory of linear extensions of posets, see Bonnet and Pouzet (1982).

EXERCISE 10 Let $(X, \succsim)$ be a preordered set, and define $\mathcal{L}(\succsim)$ as the set of all complete preorders that extend $\succsim$. Prove that $\succsim = \cap\mathcal{L}(\succsim)$. (Where do you use Szpilrajn's Theorem in the argument?)

EXERCISE 11 Let $(X, \succsim)$ be a finite preordered set. Taking $\mathcal{L}(\succsim)$ as in the previous exercise, we define $dim(X, \succsim)$ as the smallest positive integer k such that $\succsim = R_1 \cap \cdots \cap R_k$ for some $R_i \in \mathcal{L}(\succsim)$, $i = 1, \ldots, k$.

(a) Show that $dim(X, \succsim) \leq \left|X^2\right|$.

(b) What is $dim(X, D_X)$? What is $dim(X, X^2)$?

(c) For any positive integer n, show that $dim(\boxtimes^n(X_i, \succsim_i)) = n$, where $(X_i, \succsim_i)$ is a loset with $|X_i| \geq 2$ for each $i = 1, \ldots, n$.

(d) Prove or disprove: $dim(2^X, \supseteq) = |X|$.

DEFINITION

Let $(X, \succsim)$ be a preordered set, and $\emptyset \neq Y \subseteq X$. An element x of Y is said to be $\succsim$-**maximal** in Y if there is no $y \in Y$ with $y \succ x$, and $\succsim$-**minimal** in Y if there is no $y \in Y$ with $x \succ y$. If $x \succsim y$ for all $y \in Y$, then x is called the $\succsim$-**maximum** of Y, and if $y \succsim x$ for all $y \in Y$, then x is called the $\succsim$-**minimum** of Y.

Obviously, for any preordered set $(X, \succsim)$, every $\succsim$-maximum of a nonempty subset of X is $\succsim$-maximal in that set. Also note that if $(X, \succsim)$ is a poset, then there can be at most one $\succsim$-maximum of any $Y \in 2^X\backslash\{\emptyset\}$.

EXAMPLE 4

[1] Let X be any nonempty set, and $\emptyset \neq Y \subseteq X$. Every element of Y is both D_X-maximal and D_X-minimal in Y. Unless it is a singleton, Y has neither a D_X-maximum nor a D_X-minimum element. On the other hand, every element of Y is both X^2-maximum and X^2-minimum of Y.

[2] Given any nonempty set X, consider the poset $(2^X, \supseteq)$, and take any nonempty $\mathcal{A} \subseteq 2^X$. The class $\mathcal{A}$ has a $\supseteq$-maximum iff $\cup\mathcal{A} \in \mathcal{A}$, and it has a $\supseteq$-minimum iff $\cap\mathcal{A} \in \mathcal{A}$. In particular, the $\supseteq$-maximum of 2^X is X and the $\supseteq$-minimum of 2^X is $\emptyset$.

[3] (*Choice Correspondences*) Given a preference relation $\succsim$ on an alternative set X (Example 3) and a nonempty subset S of X, we define the "set of choices from S" for an individual whose preference relation is $\succsim$

as the set of all $\succsim$-maximal elements in S. That is, denoting this set as $C_{\succsim}(S)$, we have

$$C_{\succsim}(S) := \{x \in S : y \succ x \text{ for no } y \in S\}.$$

Evidently, if S is a finite set, then $C_{\succsim}(S)$ is nonempty. (Proof?) Moreover, if S is finite and $\succsim$ is complete, then there exists at least one $\succsim$-maximum element in S. The finiteness requirement cannot be omitted in this statement, but as we shall see throughout this book, there are various ways in which it can be substantially weakened. □

Exercise 12

(a) Which subsets of the set of positive integers have a $\geq$-minimum? Which ones have a $\geq$-maximum?
(b) If a set in a poset $(X, \succsim)$ has a unique $\succsim$-maximal element, does that element have to be a $\succsim$-maximum of the set?
(c) Which subsets of a poset $(X, \succsim)$ possess an element that is *both* $\succsim$-maximum *and* $\succsim$-minimum?
(d) Give an example of an infinite set in $\mathbb{R}^2$ that contains a unique $\geq$-maximal element that is also the unique $\geq$-minimal element of the set.

Exercise 13[H] Let $\succsim$ be a complete relation on a nonempty set X, and S a nonempty finite subset of X. Define

$$c_{\succsim}(S) := \{x \in S : x \succsim y \ \text{ for all } y \in S\}.$$

(a) Show that $c_{\succsim}(S) \neq \emptyset$ if $\succsim$ is transitive.
(b) We say that $\succsim$ is **acyclic** if there does not exist a positive integer k such that $x_1, \ldots, x_k \in X$ and $x_1 \succ \cdots \succ x_k \succ x_1$. Show that every transitive relation is acyclic, but not conversely.
(c) Show that $c_{\succsim}(S) \neq \emptyset$ if $\succsim$ is acyclic.
(d) Show that if $c_{\succsim}(T) \neq \emptyset$ for every finite $T \in 2^X \backslash \{\emptyset\}$, then $\succsim$ must be acyclic.

Exercise 14[H] Let $(X, \succsim)$ be a poset, and take any $Y \in 2^X \backslash \{\emptyset\}$ that has a $\succsim$-maximal element, say x^*. Prove that $\succsim$ can be extended to a linear order $\trianglerighteq$ on X such that x^* is $\trianglerighteq$-maximal in Y.

$\boxed{\text{Exercise 15}}$ Let $(X, \succsim)$ be a poset. For any $Y \subseteq X$, an element x in X is said to be an $\succsim$**-upper bound** for Y if $x \succsim y$ for all $y \in Y$; a $\succsim$**-lower bound** for Y is defined similarly. The **$\succsim$-supremum** of Y, denoted $\sup_{\succsim} Y$, is defined as the $\succsim$-minimum of the set of all $\succsim$-upper bounds for Y, that is, $\sup_{\succsim} Y$ is an $\succsim$-upper bound for Y and has the property that $z \succsim \sup_{\succsim} Y$ for any $\succsim$-upper bound z for Y. The **$\succsim$-infimum** of Y, denoted as $\inf_{\succsim} Y$, is defined analogously.

(a) Prove that there can be only one $\succsim$-supremum and only one $\succsim$-infimum of any subset of X.
(b) Show that $x \succsim y$ iff $\sup_{\succsim}\{x, y\} = x$ and $\inf_{\succsim}\{x, y\} = y$, for any $x, y \in X$.
(c) Show that if $\sup_{\succsim} X \in X$ (that is, if $\sup_{\succsim} X$ exists), then $\inf_{\succsim} \emptyset = \sup_{\succsim} X$.
(d) If $\succsim$ is the diagonal relation on X, and x and y are any two distinct members of X, does $\sup_{\succsim}\{x, y\}$ exist?
(e) If $X := \{x, y, z, w\}$ and $\succsim := \{(z, x), (z, y), (w, x), (w, y)\}$, does $\sup_{\succsim}\{x, y\}$ exist?

Exercise 16[H] Let $(X, \succsim)$ be a poset. If $\sup_{\succsim}\{x, y\}$ and $\inf_{\succsim}\{x, y\}$ exist for all $x, y \in X$, then we say that $(X, \succsim)$ is a **lattice**. If $\sup_{\succsim} Y$ and $\inf_{\succsim} Y$ exist for all $Y \in 2^X$, then $(X, \succsim)$ is called a **complete lattice**.

(a) Show that every complete lattice has an upper and a lower bound.
(b) Show that if X is finite and $(X, \succsim)$ is a lattice, then $(X, \succsim)$ is a complete lattice.
(c) Give an example of a lattice which is not complete.
(d) Prove that $(2^X, \supseteq)$ is a complete lattice.
(e) Let $\mathcal{X}$ be a nonempty subset of 2^X such that $X \in \mathcal{X}$ and $\cap\mathcal{A} \in \mathcal{X}$ for any (nonempty) class $\mathcal{A} \subseteq \mathcal{X}$. Prove that $(\mathcal{X}, \supseteq)$ is a complete lattice.

1.5 Functions

Intuitively, we think of a *function* as a rule that transforms the objects in a given set to those of another. Although this is not a formal definition—what is a "rule"?—we may now use the notion of binary relation to formalize the idea. Let X and Y be any two nonempty sets. By a **function f that maps X into Y**, denoted as $f : X \to Y$, we mean a relation $f \in X \times Y$ such that

(i) for every $x \in X$, there exists a $y \in Y$ such that $x\,f\,y$;

(ii) for every $y, z \in Y$ with $x\,f\,y$ and $x\,f\,z$, we have $y = z$.

Here X is called the **domain** of f and Y the **codomain** of f. The **range** of f is, on the other hand, defined as

$$f(X) := \{y \in Y : xfy \quad \text{for some } x \in X\}.$$

The set of all functions that map X into Y is denoted by Y^X. For instance, $\{0,1\}^X$ is the set of all functions on X whose values are either 0 or 1, and $\mathbb{R}^{[0,1]}$ is the set of all real-valued functions on $[0,1]$. The notation $f \in Y^X$ will be used interchangeably with the expression $f : X \to Y$ throughout this course. Similarly, the term **map** is used interchangeably with the term "function."

Although our definition of a function may look a bit strange at first, it is hardly anything other than a set-theoretic formulation of the concept we use in daily discourse. After all, we want a *function* f that maps X into Y to assign each member of X to a member of Y, right? Our definition says simply that one can think of f as a set of ordered pairs, so "$(x,y) \in f$" means "x is mapped to y by f." Put differently, all that f "does" is completely identified by the set $\{(x,f(x)) \in X \times Y : x \in X\}$, which is what f "is." The familiar notation $f(x) = y$ (which we shall also adopt in the rest of the exposition) is then nothing but an alternative way of expressing xfy. When $f(x) = y$, we refer to y as the **image** (or **value**) **of** x **under** f. Condition (i) says that every element in the domain X of f has an image under f in the codomain Y. In turn, condition (ii) states that no element in the domain of f can have more than one image under f.

Some authors adhere to the intuitive definition of a function as a "rule" that transforms one set into another and refer to the set of all ordered pairs $(x,f(x))$ as the **graph** of the function. Denoting this set by $Gr(f)$, then, we can write

$$Gr(f) := \{(x,f(x)) \in X \times Y : x \in X\}.$$

According to the formal definition of a function, f and $Gr(f)$ are the same thing. So long as we keep this connection in mind, there is no danger in thinking of a function as a "rule" in the intuitive way. In particular, we say that two functions f and g are equal if they have the same graph, or equivalently, if they have the same domain and codomain, and $f(x) = g(x)$ for all $x \in X$. In this case, we simply write $f = g$.

If its range equals its codomain, that is, if $f(X) = Y$, then one says that f maps X **onto** Y, and refers to it as a **surjection** (or as a **surjective** function/map). If f maps distinct points in its domain to distinct points in its codomain, that is, if $x \neq y$ implies $f(x) \neq f(y)$ for all $x, y \in X$, then we say that f is an **injection** (or a **one-to-one** or **injective** function/map). Finally, if f is both injective and surjective, then it is called a **bijection** (or a **bijective** function/map). For instance, if $X := \{1, \ldots, 10\}$, then $f := \{(1,2), (2,3), \ldots, (10,1)\}$ is a bijection in X^X, while $g \in X^X$, defined as $g(x) := 3$ for all $x \in X$, is neither an injection nor a surjection. When considered as a map in $(\{0\} \cup X)^X$, f is an injection but not a surjection.

WARNING. Every injective function can be viewed as a bijection, provided that one views the codomain of the function as its range. Indeed, if $f : X \to Y$ is an injection, then the map $f : X \to Z$ is a bijection, where $Z := f(X)$. This is usually expressed as saying that $f : X \to f(X)$ is a bijection.

Before we consider some examples, let us note that a common way of defining a particular function in a given context is to describe the domain and codomain of that function and the image of a generic point in the domain. So one would say something like, "let $f : X \to Y$ be defined by $f(x) := \ldots$" or "consider the function $f \in Y^X$ defined by $f(x) := \ldots$." For example, by the function $f : \mathbb{R} \to \mathbb{R}_+$ defined by $f(t) := t^2$, we mean the surjection that transforms every real number t to the nonnegative real number t^2. Since the domain of the function is understood from the expression $f : X \to Y$ (or $f \in Y^X$), it is redundant to add the phrase "for all $x \in X$" after the expression "$f(x) := \ldots$," although sometimes we may do so for clarity. Alternatively, when the codomain of the function is clear, a phrase like "the map $x \mapsto f(x)$ on X" is commonly used. For instance, one may refer to the quadratic function mentioned above unambiguously as "the map $t \mapsto t^2$ on $\mathbb{R}$."

EXAMPLE 5
In the following examples, X and Y stand for arbitrary nonempty sets.

[1] A **constant function** is one that assigns the same value to every element of its domain, that is, $f \in Y^X$ is constant iff there exists a $y \in Y$ such that $f(x) = y$ for all $x \in X$. (Formally speaking, this constant function is the set $X \times \{y\}$.) Obviously, $f(X) = \{y\}$ in this case, so a constant function is not surjective unless its codomain is a singleton, and it is not injective unless its domain is a singleton.

[2] A function whose domain and codomain are identical, that is, a function in X^X, is called a **self-map** on X. An important example of a self-map is the **identity function** on X. This function is denoted as id_X, and it is defined as $\mathrm{id}_X(x) := x$ for all $x \in X$. Obviously, id_X is a bijection, and formally speaking, it is none other than the diagonal relation D_X.

[3] Let $S \subseteq X$. The function that maps X into $\{0, 1\}$ such that every member of S is assigned to 1 and all the other elements of X are assigned to zero is called the **indicator function of S in X**. This function is denoted as 1_S (assuming that the domain X is understood from the context). By definition, we have

$$1_S(x) := \begin{cases} 1, & \text{if } x \in S \\ 0, & \text{if } x \in X \setminus S \end{cases}.$$

You can check that, for every $A, B \subseteq X$, we have $1_{A \cup B} + 1_{A \cap B} = 1_A + 1_B$ and $1_{A \cap B} = 1_A 1_B$. □

The following examples point to some commonly used methods of obtaining new functions from a given set of functions.

Example 6

In the following examples, X, Y, Z, and W stand for arbitrary nonempty sets.

[1] Let $Z \subseteq X \subseteq W$, and $f \in Y^X$. By the **restriction** of f to Z, denoted as $f|_Z$, we mean the function $f|_Z \in Y^Z$ defined by $f|_Z(z) := f(z)$. By an **extension** of f to W, on the other hand, we mean a function $f^* \in Y^W$ with $f^*|_X = f$, that is, $f^*(x) = f(x)$ for all $x \in X$. If f is injective, so must $f|_Z$, but surjectivity of f does not entail that of $f|_Z$. Of course, if f is not injective, $f|_Z$ may still turn out to be injective (e.g., $t \mapsto t^2$ is not injective on $\mathbb{R}$, but it is so on $\mathbb{R}_+$).

[2] Sometimes it is possible to extend a given function by *combining* it with another function. For instance, we can combine any $f \in Y^X$ and $g \in W^Z$ to obtain the function $h : X \cup Z \to Y \cup W$ defined by

$$h(t) := \begin{cases} f(t), & \text{if } t \in X \\ g(t), & \text{if } t \in Z \end{cases},$$

provided that $X \cap Z = \emptyset$, or $X \cap Z \neq \emptyset$ and $f|_{X\cap Z} = g|_{X\cap Z}$. Note that this method of combining functions does *not* work if $f(t) \neq g(t)$ for some $t \in X \cap Z$. For, in that case h would not be well-defined as a function. (What would be the image of t under h?)

[3] A function $f \in X^{X\times Y}$ defined by $f(x,y) := x$ is called the **projection from** $X \times Y$ **onto** X.[10] (The projection from $X \times Y$ onto Y is similarly defined.) Obviously, $f(X \times Y) = X$, that is, f is necessarily surjective. It is not injective unless Y is a singleton.

[4] Given functions $f : X \to Z$ and $g : Z \to Y$, we define the **composition** of f and g as the function $g \circ f : X \to Y$ by $g \circ f\,(x) := g(f(x))$. (For easier reading, we often write $(g \circ f)(x)$ instead of $g \circ f\,(x)$.) This definition accords with the way we defined the composition of two relations (Exercise 5). Indeed, we have $(g \circ f)(x) = \{(x,y) : x\,f\,z \text{ and } z\,g\,y \text{ for some } z \in Z\}$.

Obviously, $\mathrm{id}_Z \circ f = f = f \circ \mathrm{id}_X$. Even when $X = Y = Z$, the operation of taking compositions is not commutative. For instance, if the self-maps f and g on $\mathbb{R}$ are defined by $f(t) := 2$ and $g(t) := t^2$, respectively, then $(g \circ f)(t) = 4$ and $(f \circ g)(t) = 2$ for any real number t. The composition operation is, however, associative. That is, $h \circ (g \circ f) = (h \circ g) \circ f$ for all $f \in Y^X$, $g \in Z^Y$ and $h \in W^Z$. $\square$

Exercise 17 Let $\sim$ be an equivalence relation on a nonempty set X. Show that the map $x \mapsto [x]_\sim$ on X (called the **quotient map**) is a surjection on X which is injective iff $\sim = D_X$.

Exercise 18[H] (*A Factorization Theorem*) Let X and Y be two nonempty sets. Prove: For any function $f : X \to Y$, there exists a nonempty set Z, a surjection $g : X \to Z$, and an injection $h : Z \to Y$ such that $f = h \circ g$.

Exercise 19 Let X, Y, and Z be nonempty sets, and consider any $f, g \in Y^X$ and $u, v \in Z^Y$. Prove:

(a) If f is surjective and $u \circ f = v \circ f$, then $u = v$.

(b) If u is injective and $u \circ f = u \circ g$, then $f = g$.

(c) If f and u are injective (respectively, surjective), then so is $u \circ f$.

[10] Strictly speaking, I should write $f((x,y))$ instead of $f(x,y)$, but that's just splitting hairs.

Exercise 20[H] Show that there is no surjection of the form $f : X \to 2^X$ for any nonempty set X.

For any given nonempty sets X and Y, the (*direct*) **image** of a set $A \subseteq X$ under $f \in Y^X$, denoted $f(A)$, is defined as the collection of all elements y in Y with $y = f(x)$ for some $x \in A$. That is,

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

The range of f is thus the image of its entire domain: $f(X) = \{f(x) : x \in X\}$. (*Note*. If $f(A) = B$, then one says that "f maps A onto B.")

The **inverse image** of a set B in Y under f, denoted as $f^{-1}(B)$, is defined as the set of all x in X whose images under f belong to B, that is,

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

By convention, we write $f^{-1}(y)$ for $f^{-1}(\{y\})$, that is,

$$f^{-1}(y) := \{x \in X : f(x) = y\} \qquad \text{for any } y \in Y.$$

Obviously, $f^{-1}(y)$ is a singleton for each $y \in Y$ iff f is an injection. For instance, if f stands for the map $t \mapsto t^2$ on $\mathbb{R}$, then $f^{-1}(1) = \{-1, 1\}$, whereas $f|_{\mathbb{R}_+}^{-1}(1) = \{1\}$.

The issue of whether or not one can express the image (or the inverse image) of a union/intersection of a collection of sets as the union/intersection of the images (inverse images) of each set in the collection arises quite often in mathematical analysis. The following exercise summarizes the situation in this regard.

$\boxed{\text{Exercise 21}}$ Let X and Y be nonempty sets and $f \in Y^X$. Prove that, for any (nonempty) classes $\mathcal{A} \subseteq 2^X$ and $\mathcal{B} \subseteq 2^Y$, we have

$$f(\cup\mathcal{A}) = \bigcup\{f(A) : A \in \mathcal{A}\} \quad \text{and} \quad f(\cap\mathcal{A}) \subseteq \bigcap\{f(A) : A \in \mathcal{A}\},$$

whereas

$$f^{-1}(\cup\mathcal{B}) = \bigcup\{f^{-1}(B) : B \in \mathcal{B}\} \quad \text{and} \quad f^{-1}(\cap\mathcal{B}) = \bigcap\{f^{-1}(B) : B \in \mathcal{B}\}.$$

A general rule that surfaces from this exercise is that inverse images are quite well-behaved with respect to the operations of taking unions and intersections, while the same cannot be said for direct images in the case of taking intersections. Indeed, for any $f \in Y^X$, we have $f(A \cap B) \supseteq f(A) \cap f(B)$ for all $A, B \subseteq X$ if, *and only if*, f is injective.[11] The "if" part of this assertion is trivial. The "only if" part follows from the observation that, if the claim was not true, then, for any distinct $x, y \in X$ with $f(x) = f(y)$, we would find $\emptyset = f(\emptyset) = f(\{x\} \cap \{y\}) = f(\{x\}) \cap f(\{y\}) = \{f(x)\}$, which is absurd.

Finally, we turn to the problem of *inverting* a function. For any function $f \in Y^X$, let us define the set

$$f^{-1} := \{(y, x) \in Y \times X : x f y\}$$

which is none other than the inverse of f viewed as a relation (Exercise 5). This relation simply *reverses* the map f in the sense that if x is mapped to y by f, then f^{-1} maps y back to x. Now, f^{-1} may or may not be a function. If it is, we say that f is **invertible** and f^{-1} is the **inverse** of f. For instance, $f : \mathbb{R} \to \mathbb{R}_+$ defined by $f(t) := t^2$ is not invertible (since $(1, 1) \in f^{-1}$ *and* $(1, -1) \in f^{-1}$, that is, 1 does not have a unique image under f^{-1}), whereas $f|_{\mathbb{R}_+}$ is invertible and $f|_{\mathbb{R}_+}^{-1}(t) = \sqrt{t}$ for all $t \in \mathbb{R}$.

The following result gives a simple characterization of invertible functions.

Proposition 2
Let X and Y be two nonempty sets. A function $f \in Y^X$ is invertible if, and only if, it is a bijection.

Exercise 22 Prove Proposition 2.

By using the composition operation defined in Example 6.[4], we can give another useful characterization of invertible functions.

Proposition 3
Let X and Y be two nonempty sets. A function $f \in Y^X$ is invertible if, and only if, there exists a function $g \in X^Y$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$.

[11] Of course, this does not mean that $f(A \cap B) = f(A) \cap f(B)$ can never hold for a function that is not one-to-one. It only means that, for any such function f, we can always find nonempty sets A and B in the domain of f such that $f(A \cap B) \supseteq f(A) \cap f(B)$ is false.

Proof

The "only if" part is readily obtained upon choosing $g := f^{-1}$. To prove the "if" part, suppose there exists a $g \in X^Y$ with $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$, and note that, by Proposition 2, it is enough to show that f is a bijection. To verify the injectivity of f, pick any $x, y \in X$ with $f(x) = f(y)$, and observe that

$$x = \mathrm{id}_X(x) = (g \circ f)(x) = g(f(x)) = g(f(y)) = (g \circ f)(y) = \mathrm{id}_X(y) = y.$$

To see the surjectivity of f, take any $y \in Y$ and define $x := g(y)$. Then we have

$$f(x) = f(g(y)) = (f \circ g)(y) = \mathrm{id}_Y(y) = y,$$

which proves $Y \subseteq f(X)$. Since the converse containment is trivial, we are done. ■

1.6 Sequences, Vectors, and Matrices

By a *sequence* in a given nonempty set X, we intuitively mean an ordered array of the form $(x_1, x_2, \ldots)$ where each term x_i of the sequence is a member of X. (Throughout this text we denote such a sequence by (x_m), but note that some books prefer instead the notation $(x_m)_{m=1}^{\infty}$.) As in the case of ordered pairs, one could introduce the notion of a sequence as a new object to our set theory, but again there is really no need to do so. Intuitively, we understand from the notation $(x_1, x_2, \ldots)$ that the ith term in the array is x_i. But then we can think of this array as a function that maps the set $\mathbb{N}$ of positive integers into X in the sense that it tells us that "the ith term in the array is x_i" by mapping i to x_i. With this definition, our intuitive understanding of the ordered array $(x_1, x_2, \ldots)$ is formally captured by the function $\{(i, x_i) : i = 1, 2, \ldots\} = f$. Thus, we define a **sequence** in a nonempty set X as any function $f : \mathbb{N} \to X$, and *represent* this function as $(x_1, x_2, \ldots)$ where $x_i := f(i)$ for each $i \in \mathbb{N}$. Consequently, the set of all sequences in X is equal to $X^{\mathbb{N}}$. As is common, however, we denote this set as X^{∞} throughout the text.

By a *subsequence* of a sequence $(x_m) \in X^{\infty}$, we mean a sequence that is made up of the terms of (x_m) that appear in the subsequence in the same order they appear in (x_m). That is, a subsequence of (x_m) is of the form $(x_{m_1}, x_{m_2}, \ldots)$, where (m_k) is a sequence in $\mathbb{N}$ such that $m_1 < m_2 < \cdots$. (We denote this subsequence as (x_{m_k}).) Once again, we use the notion of function to formalize this definition. Strictly speaking, a **subsequence** of a sequence $f \in X^{\mathbb{N}}$ is a function of the form $f \circ \sigma$, where $\sigma : \mathbb{N} \to \mathbb{N}$ is strictly increasing (that is, $\sigma(k) < \sigma(l)$ for any $k, l \in \mathbb{N}$ with $k < l$). We *represent* this function as the array $(x_{m_1}, x_{m_2}, \ldots)$ with the understanding that $m_k = \sigma(k)$ and $x_{m_k} = f(m_k)$ for each $k = 1, 2, \ldots$. For instance, $(x_{m_k}) := (1, \frac{1}{3}, \frac{1}{5}, \ldots)$ is a subsequence of $(x_m) := (\frac{1}{m}) \in \mathbb{R}^{\infty}$. Here (x_m) is a representation for the function $f \in \mathbb{R}^{\mathbb{N}}$, which is defined by $f(i) := \frac{1}{i}$, and (x_{m_k}) is a representation of the map $f \circ \sigma$, where $\sigma(k) := 2k - 1$ for each $k \in \mathbb{N}$.

By a *double sequence* in X, we mean an infinite matrix each term of which is a member of X. Formally, a **double sequence** is a function $f \in X^{\mathbb{N} \times \mathbb{N}}$. As in the case of sequences, we *represent* this function as (x_{kl}), with the understanding that $x_{kl} := f(k, l)$. The set of all double sequences in X equals $X^{\mathbb{N} \times \mathbb{N}}$, but it is customary to denote this set as $X^{\infty \times \infty}$. We note that one can always view (in more than one way) a double sequence in X as a sequence of sequences in X, that is, as a sequence in X^{∞}. For instance, we can think of (x_{kl}) as $((x_{1l}), (x_{2l}), \ldots)$ or as $((x_{k1}), (x_{k2}), \ldots)$.

The basic idea of viewing a string of objects as a particular function also applies to *finite* strings, of course. For instance, how about $X^{\{1,\ldots,n\}}$, where X is a nonempty set and n is some positive integer? The preceding discussion shows that this function space is none other than the set $\{(x_1, \ldots, x_n) : x_i \in X, i = 1, \ldots, n\}$. Thus we may define an ***n*-vector** in X as a function $f : \{1, \ldots, n\} \to X$, and *represent* this function as $(x_1, \ldots, x_n)$ where $x_i := f(i)$ for each $i = 1, \ldots, n$. (Check that $(x_1, \ldots, x_n) = (x'_1, \ldots, x'_n)$ iff $x_i = x'_i$ for each $i = 1, \ldots, n$, so everything is in concert with the way we defined n-vectors in Section 1.2.) The ***n*-fold product** of X is then defined as $X^{\{1,\ldots,n\}}$, but is denoted as X^n. (So $\mathbb{R}^n = \mathbb{R}^{\{1,\ldots,n\}}$. This makes sense, no?) The main lesson is that everything that is said about arbitrary functions also applies to sequences and vectors.

Finally, for any positive integers m and n, by an $m \times n$ **matrix** (read "m by n matrix") in a nonempty set X, we mean a function $f : \{1, \ldots, m\} \times \{1, \ldots, n\} \to X$. We *represent* this function as $[a_{ij}]_{m \times n}$, with the understanding that $a_{ij} := f(i, j)$ for each $i = 1, \ldots, m$ and $j = 1, \ldots, n$. (As you know, one often views a matrix like $[a_{ij}]_{m \times n}$ as a rectangular array with m rows and n columns in which a_{ij} appears in the ith row and jth column.)

The set of all $m \times n$ matrices in X is $X^{\{1,\ldots,m\} \times \{1,\ldots,n\}}$, but it is much better to denote this set as $X^{m \times n}$. Needless to say, both $X^{1 \times n}$ and $X^{n \times 1}$ can be identified with X^n. (Wait, what does this mean?)

1.7* A Glimpse of Advanced Set Theory: The Axiom of Choice

We now turn to a problem that we have so far conveniently avoided: How do we define the Cartesian product of *infinitely* many nonempty sets? Intuitively speaking, the Cartesian product of all members of a class $\mathcal{A}$ of sets is the set of all collections each of which contains one and only one element of each member of $\mathcal{A}$. That is, a member of this product is really a function on $\mathcal{A}$ that selects a single element from each set in $\mathcal{A}$. The question is simple to state: Does there exist such a function?

If $|\mathcal{A}| < \infty$, then the answer would obviously be yes, because we can *construct* such a function by choosing an element from each set in $\mathcal{A}$ one by one. But when $\mathcal{A}$ contains infinitely many sets, then this method does not readily work, so we need to *prove* that such a function exists.

To get a sense of this, suppose $\mathcal{A} := \{A_1, A_2, \ldots\}$, where $\emptyset \neq A_i \subseteq \mathbb{N}$ for each $i = 1, 2, \ldots$. Then we're okay. We can define $f : \mathcal{A} \to \cup\mathcal{A}$ by $f(A) :=$ the smallest element of A – this well-defines f as a map that selects one element from each member of $\mathcal{A}$ *simultaneously*. Or, if each A_i is a bounded interval in $\mathbb{R}$, then again we're fine. This time we can define f, say, as follows: $f(A) :=$ the midpoint of A. But what if all we knew was that each A_i consists of real numbers? Or worse, what if we were not told anything about the contents of $\mathcal{A}$? You see, in general, we can't write down a formula, or an algorithm, the application of which yields such a function. Then how do you know that such a thing exists in the first place?[12]

[12] But, how about the following algorithm? Start with A_1, and pick any a_1 in A_1. Now move to A_2 and pick any $a_2 \in A_2$. Continue this way, and define $g : \mathcal{A} \to \cup\mathcal{A}$ by $g(A_i) = a_i$, $i = 1, 2, \ldots$. Aren't we done? No, we are not! The function at hand is not well-defined—its definition does not tell me *exactly* which member of A_{27} is assigned to $g(A_{27})$—this is very much unlike how I defined f above in the case where each A_i was contained in $\mathbb{N}$ (or was a bounded interval).

Perhaps you are still not quite comfortable about this. You might think that f is well-defined, it's just that it is defined *recursively*. Let me try to illustrate the problem by means of a concrete example. Take any infinite set S, and ask yourself if you can define an injection f from $\mathbb{N}$ into S. Sure, you might say, "recursion" is again the name of the game. Let $f(1)$ be any member a_1 of S. Then let $f(2)$ be any member of $S\backslash\{a_1\}$, $f(3)$ any member $S\backslash\{a_1, a_2\}$, and so on. Since $S\backslash T \neq \emptyset$ for any finite $T \subset S$, this well-defines f, recursively, as an injection from $\mathbb{N}$ into S. Wrong! If this were the case, on the basis of the knowledge of $f(1), \ldots, f(26)$, I would know the value of f at 27. The "definition" of f doesn't do that—it just points to some *arbitrary* member of A_{27}—so it is not a proper definition at all.

(*Note*. As "obvious" as it might seem, the proposition "for any infinite set S, there is an injection in $S^{\mathbb{N}}$," cannot be proved within the standard realm of set theory.)

In fact, it turns out that the problem of "finding an $f : \mathcal{A} \to \cup\mathcal{A}$ for *any* given class $\mathcal{A}$ of sets" cannot be settled in one way or another by means of the standard axioms of set theory.[13] The status of our question is thus a bit odd, it is *undecidable*.

To make things a bit more precise, let us state formally the property that we are after.

The Axiom of Choice. *For any (nonempty) class $\mathcal{A}$ of sets, there exists a function $f : \mathcal{A} \to \cup\mathcal{A}$ such that $f(A) \in A$ for each $A \in \mathcal{A}$.*

One can reword this in a few other ways.

Exercise 23 Prove that the Axiom of Choice is equivalent to the following statements.

(i) For any nonempty set S, there exists a function $f : 2^S \backslash \{\emptyset\} \to S$ such that $f(A) \in A$ for each $\emptyset \neq A \subseteq S$.

(ii) (*Zermelo's Postulate*) If $\mathcal{A}$ is a (nonempty) class of sets such that $A \cap B = \emptyset$ for each distinct $A, B \in \mathcal{A}$, then there exists a set S such that $|S \cap A| = 1$ for every $A \in \mathcal{A}$.

(iii) For any nonempty sets X and Y, and any relation R from X into Y, there is a function $f : Z \to Y$ with $\emptyset \neq Z \subseteq X$ and $f \subseteq R$. (That is: *Every relation contains a function.*)

The first thing to note about the Axiom of Choice is that it cannot be disproved by using the standard axioms of set theory. Provided that these axioms are consistent (that is, no contradiction may be logically deduced from them), adjoining the Axiom of Choice to these axioms yields again a consistent set of axioms. This raises the possibility that perhaps the Axiom of Choice can be deduced as a "theorem" from the standard axioms. The second thing to know about the Axiom of Choice is that this is false, that is, the Axiom of Choice is not provable from the standard axioms of set theory.[14]

[13] For brevity, I am again being imprecise about this *standard* set of axioms (called the *Zermelo-Fraenkel-Skolem* axioms). For the present discussion, nothing will be lost if you just think of these as the formal properties needed to "construct" the set theory we outlined intuitively earlier. It is fair to say that these axioms have an unproblematic standing in mathematics.

[14] These results are of extreme importance for the foundations of the entire field of mathematics. The first one was proved by Kurt Gödel in 1939 and the second one by Paul Cohen in 1963.

We are then at a crossroads. We must either reject the validity of the Axiom of Choice and confine ourselves to the conclusions that can be reached only on the basis of the standard axioms of set theory, or alternatively, adjoin the Axiom of Choice to the standard axioms to obtain a richer set theory that is able to yield certain results that could not have been proved within the confines of the standard axioms. Most analysts follow the second route. However, it is fair to say that the status of the Axiom of Choice is in general viewed as less appealing than the standard axioms, so one often makes it explicit if this axiom is a prerequisite for a particular theorem to be proved. Given our applied interests, we will be more relaxed about this matter and mention the (implicit) use of the Axiom of Choice in our arguments only rarely.

As an immediate application of the Axiom of Choice, we now define the **Cartesian product** of an arbitrary (nonempty) class $\mathcal{A}$ of sets as the set of all $f : \mathcal{A} \to \cup\mathcal{A}$ with $f(A) \in A$ for each $A \in \mathcal{A}$. We denote this set by $\mathsf{X}\mathcal{A}$, and note that $\mathsf{X}\mathcal{A} \neq \emptyset$ because of the Axiom of Choice. If $\mathcal{A} = \{A_i : i \in I\}$, where I is an index set, then we write $\mathsf{X}_{i\in I}A_i$ for $\mathsf{X}\mathcal{A}$. Clearly, $\mathsf{X}_{i\in I}A_i$ is the set of all maps $f : I \to \cup\{A_i : i \in I\}$ with $f(i) \in A_i$ for each $i \in I$. It is easily checked that this definition is consistent with the definition of the Cartesian product of finitely many sets given earlier.

There are a few equivalent versions of the Axiom of Choice that are often more convenient to use in applications than the original statement of the axiom. To state the most widely used version, let us first agree on some terminology. For any poset $(X, \succsim)$, by a "poset in $(X, \succsim)$," we mean a poset like $(Y, \succsim \cap\, Y^2)$ with $Y \subseteq X$, but we denote this poset more succinctly as $(Y, \succsim)$. By an *upper bound* for such a poset, we mean an element x of X with $x \succsim y$ for all $y \in Y$ (Exercise 15).

Zorn's Lemma
If every loset in a given poset has an upper bound, then that poset must have a maximal element.

Although this is a less intuitive statement than the Axiom of Choice (no?), it can in fact be shown to be equivalent to the Axiom of Choice.[15] (That is, we can deduce Zorn's Lemma from the standard axioms *and* the Axiom of Choice, and we can prove the Axiom of Choice by using the standard axioms

[15] For a proof, see Enderton (1977, pp. 151–153) or Kelley (1955, pp. 32–35).

and Zorn's Lemma.) Since we take the Axiom of Choice as "true" in this text, therefore, we must also accept the validity of Zorn's Lemma.

We conclude this discussion by means of two quick applications that illustrate how Zorn's Lemma is used in practice. We will see some other applications in later chapters.

Let us first prove the following fact:

The Hausdorff Maximal Principle
There exists a $\supseteq$-maximal loset in every poset.

Proof
Let $(X, \succsim)$ be a poset, and

$$\mathcal{L}(X, \succsim) := \left\{ Z \subseteq X : (Z, \succsim) \text{ is a loset} \right\}.$$

(Observe that $\mathcal{L}(X, \succsim) \neq \emptyset$ by reflexivity of $\succsim$.) We wish to show that there is a $\supseteq$-maximal element of $\mathcal{L}(X, \succsim)$. This will follow from Zorn's Lemma, if we can show that every loset in the poset $(\mathcal{L}(X, \succsim), \supseteq)$ has an upper bound, that is, for any $\mathcal{A} \subseteq \mathcal{L}(X, \succsim)$ such that $(\mathcal{A}, \supseteq)$ is a loset, there is a member of $\mathcal{L}(X, \succsim)$ that contains $\mathcal{A}$. To establish that this is indeed the case, take any such $\mathcal{A}$, and let $Y := \cup\mathcal{A}$. Then $\succsim$ is a complete relation on Y, because, since $\supseteq$ linearly orders $\mathcal{A}$, for any $x, y \in Y$ we must have $x, y \in A$ for some $A \in \mathcal{A}$ (why?), and hence, given that $(A, \succsim)$ is a loset, we have either $x \succsim y$ or $y \succsim x$. Therefore, $(Y, \succsim)$ is a loset, that is, $Y \in \mathcal{L}(X, \succsim)$. But it is obvious that $Y \supseteq A$ for any $A \in \mathcal{A}$. ■

In fact, the Hausdorff Maximal Principle is equivalent to the Axiom of Choice.

Exercise 24 Prove Zorn's Lemma assuming the validity of the Hausdorff Maximal Principle.

As another application of Zorn's Lemma, we prove Szpilrajn's Theorem.[16] Our proof uses the Hausdorff Maximal Principle, but you now know that this is equivalent to invoking Zorn's Lemma or the Axiom of Choice.

[16] In case you are wondering, Szpilrajn's Theorem is *not* equivalent to the Axiom of Choice.

Proof of Szpilrajn's Theorem

Let $\succsim$ be a partial order on a nonempty set X. Let $\mathcal{T}_X$ be the set of all partial orders on X that extend $\succsim$. Clearly, $(\mathcal{T}_X, \supseteq)$ is a poset, so by the Hausdorff Maximal Principle, it has a maximal loset, say, $(\mathcal{A}, \supseteq)$. Define $\succsim^* := \cup\mathcal{A}$. Since $(\mathcal{A}, \supseteq)$ is a loset, $\succsim^*$ is a partial order on X that extends $\succsim$. (Why?) $\succsim^*$ is in fact complete. To see this, suppose we can find some $x, y \in X$ with neither $x \succsim^* y$ nor $y \succsim^* x$. Then the transitive closure of $\succsim^* \cup \{(x, y)\}$ is a member of $\mathcal{T}_X$ that contains $\succsim^*$ as a proper subset (Exercise 8). (Why exactly?) This contradicts the fact that $(\mathcal{A}, \supseteq)$ is a maximal loset within $(\mathcal{T}_X, \supseteq)$. (Why?) Thus $\succsim^*$ is a linear order, and we are done. ■

2 Real Numbers

This course assumes that the reader has a basic understanding of the real numbers, so our discussion here will be brief and duly heuristic. In particular, we will not even attempt to give a *construction* of the set $\mathbb{R}$ of real numbers. Instead we will mention some axioms that $\mathbb{R}$ satisfies, and focus on certain properties that $\mathbb{R}$ possesses. Some books on real analysis give a fuller view of the construction of $\mathbb{R}$, some talk about it even less than we do. If you are really curious about this, it's best if you consult a book that specializes in this sort of a thing. (Try, for instance, Chapters 4 and 5 of Enderton (1977).)

2.1 Ordered Fields

In this subsection we talk briefly about a few topics in abstract algebra that will facilitate our discussion of real numbers.

Definition

Let X be any nonempty set. We refer to a function of the form $\bullet : X \times X \to X$ as a **binary operation** on X, and write $x \bullet y$ instead of $\bullet(x, y)$ for any $x, y \in X$.

For instance, the usual addition and multiplication operations $+$ and $\cdot$ are binary operations on the set $\mathbb{N}$ of natural numbers. The subtraction operation is, on the other hand, not a binary operation on $\mathbb{N}$ (e.g., $1+(-2) \notin \mathbb{N}$), but it is a binary operation on the set of all integers.

DEFINITION

Let X be any nonempty set, let $+$ and $\cdot$ be two binary operations on X, and let us agree to write xy for $x \cdot y$ for simplicity. The list $(X, +, \cdot)$ is called a **field** if the following properties are satisfied:

(i) (*Commutativity*) $x + y = y + x$ and $xy = yx$ for all $x, y \in X$;

(ii) (*Associativity*) $(x + y) + z = x + (y + z)$ and $(xy)z = x(yz)$ for all $x, y, z \in X$;[17]

(iii) (*Distributivity*) $x(y + z) = xy + xz$ for all $x, y, z \in X$;

(iv) (*Existence of Identity Elements*) There exist elements 0 and 1 in X such that $0 + x = x = x + 0$ and $1x = x = x1$ for all $x \in X$;

(v) (*Existence of Inverse Elements*) For each $x \in X$ there exists an element $-x$ in X (the *additive inverse* of x) such that $x + -x = 0 = -x + x$, and for each $x \in X\backslash\{0\}$ there exists an element x^{-1} in X (the *multiplicative inverse* of x) such that $xx^{-1} = 1 = x^{-1}x$.

A field $(X, +, \cdot)$ is an algebraic structure that envisions two binary operations, $+$ and $\cdot$, on the set X in a way that makes a satisfactory *arithmetic* possible. In particular, given the $+$ and $\cdot$ operations, we can define the two other (inverse) operations $-$ and $/$ by $x - y := x + -y$ and $x/y := xy^{-1}$, the latter provided that $y \neq 0$. (Strictly speaking, the division operation $/$ is not a binary operation; for instance, $1/0$ is not defined in X.)

Pretty much the entire arithmetic that we are familiar with in the context of $\mathbb{R}$ can be performed within an arbitrary field. To illustrate this, let us establish a few arithmetic laws that you may recall from high school algebra. In particular, let us show that

$$x+y = x+z \quad \text{iff} \quad y = z, \qquad -(-x) = x \quad \text{and} \quad -(x+y) = -x+-y \quad (1)$$

in any field $(X, +, \cdot)$. The first claim is a *cancellation law*, which is readily proved by observing that, for any $w \in X$, we have $w = 0+w = (-x+x)+w = -x + (x + w)$. Thus, $x + y = x + z$ implies $y = -x + (x + y) = z$, and we're done. As an immediate corollary of this cancellation law, we find that

[17] Throughout this exposition, (w) is the same thing as w, for any $w \in X$. For instance, $(x + y)$ corresponds to $x + y$, and $(-x)$ corresponds to $-x$. The brackets are used at times only for clarity.

the additive inverse of each element in X is unique. (The same holds for the multiplicative inverses as well. *Quiz*. Prove!) On the other hand, the second claim in (1) is true because

$$x = x+0 = x+(-x+-(-x)) = (x+-x)+-(-x) = 0+-(-x) = -(-x).$$

Finally, given that the additive inverse of $x+y$ is unique, the last claim in (1) follows from the following argument:

$$\begin{aligned}(x+y)+(-x+-y) &= (x+y)+(-y+-x)\\ &= x+(y+(-y+-x))\\ &= x+((y+-y)+-x)\\ &= x+(0+-x)\\ &= x+-x\\ &= 0.\end{aligned}$$

(*Quiz*. Prove that $-1x = -x$ in any field. *Hint*. There is something to be proved here!)

Exercise 25 (*Rules of Exponentiation*) Let $(X, +, \cdot)$ be a field. For any $x \in X$, we define $x^0 := 1$, and for any positive integer k, we let $x^k := x^{k-1}x$ and $x^{-k} := (x^k)^{-1}$. For any integers i and j, prove that $x^i x^j = x^{i+j}$ and $(x^i)^j = x^{ij}$ for any $x \in X$, and $x^i/x^j = x^{i-j}$ and $(y/x)^i = y^i/x^i$ for any $x \in X\backslash\{0\}$.

Although a field provides a rich environment for doing arithmetic, it lacks structure for *ordering* things. We introduce such a structure next.

Definition

The list $(X, +, \cdot, \geq)$ is called an **ordered field** if $(X, +, \cdot)$ is a field, and if $\geq$ is a partial order on X that is compatible with the operations $+$ and $\cdot$ in the sense that $x \geq y$ implies $x+z \geq y+z$ for any $x, y, z \in X$, and $xz \geq yz$ for any $x, y, z \in X$ with $z \geq 0$. We note that the expressions $x \geq y$ and $y \leq x$ are identical. The same goes also for the expressions $x > y$ and

$y < x$.[18] We also adopt the following notation:

$$X_+ := \{x \in X : x \geq 0\} \quad \text{and} \quad X_{++} := \{x \in X : x > 0\},$$

and

$$X_- := \{x \in X : x \leq 0\} \quad \text{and} \quad X_{--} := \{x \in X : x < 0\}.$$

An ordered field is a rich algebraic system within which many algebraic properties of real numbers can be established. This is of course not the place to get into a thorough algebraic analysis, but we should consider at least one example to give you an idea about how this can be done.

Example 7

(*The Triangle Inequality*) Let $(X, +, \cdot, \geq)$ be an ordered field. The function $|\cdot| : X \to X$ defined by

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

is called the **absolute value** function.[19] The following is called the **triangle inequality**:

$$|x + y| \leq |x| + |y| \qquad \text{for all } x, y \in X.$$

You have surely seen this inequality in the case of real numbers. The point is that it is valid within *any* ordered field, so the only properties responsible for it are the ordered field axioms.

We divide the argument into five easy steps. All x and y that appear in these steps are arbitrary elements of X.

(a) $|x| \geq x$. *Proof.* If $x \geq 0$, then $|x| = x$ by definition. If $0 > x$, on the other hand, we have

$$|x| = -x = 0 + -x \geq x + -x = 0 \geq x.$$

(b) $x \geq 0$ implies $-x \leq 0$, and $x \leq 0$ implies $-x \geq 0$. *Proof.* If $x \geq 0$, then

$$0 = x + -x \geq 0 + -x = -x.$$

[18] Naturally, $x > y$ means that x and y are distinct members of X with $x \geq y$. That is, $>$ is the asymmetric part of $\geq$.

[19] We owe the notation $|x|$ to Karl Weierstrass. Before Weierstrass's famous 1858 lectures, there was apparently no unity on denoting the absolute value function. For instance, Bernhard Bolzano would write $\pm x$!

The second claim is proved analogously.

(c) $x \geq -|x|$. *Proof.* If $x \geq 0$, then $x \geq 0 \geq -x = -|x|$ where the second inequality follows from (b). If $0 > x$, then $-|x| = -(-x) = x$ by (1).

(d) $x \geq y$ implies $-y \geq -x$. *Proof.* Exercise.

(e) $|x + y| \leq |x| + |y|$. *Proof.* Applying (a) twice,

$$|x| + |y| \geq x + |y| = |y| + x \geq y + x = x + y.$$

Similarly, by using (c) twice,

$$x + y \geq -|x| + -|y| = (|x| + |y|)$$

where we used the third claim in (1) to get the final equality. By (d), therefore, $|x| + |y| \geq -(x + y)$, and we are done. $\square$

Exercise 26 Let $(X, +, \cdot, \geq)$ be an ordered field. Prove:

$$|xy| = |x|\,|y| \qquad \text{and} \qquad |x - y| \geq \big||x| - |y|\big| \qquad \text{for all } x, y \in X.$$

2.2 Natural Numbers, Integers, and Rationals

As you already know, we denote the set of all *natural numbers* by $\mathbb{N}$, that is, $\mathbb{N} := \{1, 2, 3, \ldots\}$. Among the properties that this system satisfies, a particularly interesting one that we wish to mention is the following:

> The Principle of Mathematical Induction
> *If S is a subset of $\mathbb{N}$ such that $1 \in S$, and $i + 1 \in S$ whenever $i \in S$, then $S = \mathbb{N}$.*

This property is actually one of the main *axioms* that are commonly used to construct the natural numbers.[20] It is frequently employed when giving

[20] Roughly speaking, the standard construction goes as follows. One postulates that $\mathbb{N}$ is a set with a linear order, called the *successor relation*, which specifies an immediate successor for each member of $\mathbb{N}$. If $i \in \mathbb{N}$, then the immediate successor of i is denoted as $i + 1$. Then, $\mathbb{N}$ is the set that is characterized by the Principle of Mathematical Induction and the following three axioms: (i) there is an element 1 in $\mathbb{N}$ that is not a successor of any other element in $\mathbb{N}$; (ii) if $i \in \mathbb{N}$, then $i + 1 \in \mathbb{N}$; and (iii) if i and j have the same successor, then $i = j$. Along with the Principle of Mathematical Induction, these properties are known as the *Peano axioms* (in honor of Giuseppe Peano (1858–1932), who first formulated these postulates and laid out an axiomatic foundation for the integers). The binary operations $+$ and $\cdot$ are defined via the successor relation, and behave "well" because of these axioms.

a recursive definition (as in Exercise 25), or when proving infinitely many propositions by recursion. Suppose $P_1, P_2, \ldots$ are logical statements. If we can prove that P_1 is true, and then show that the validity of P_{i+1} would in fact follow from the validity of P_i (i being arbitrarily fixed in $\mathbb{N}$), then we may invoke the Principle of Mathematical Induction to conclude that each proposition in the string $P_1, P_2, \ldots$ is true. For instance, suppose we wish to prove that

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^i} = 2 - \frac{1}{2^i} \quad \text{for each } i \in \mathbb{N}. \tag{2}$$

Then we first check if the claim holds for $i = 1$. Since $1 + \frac{1}{2} = 2 - \frac{1}{2}$, this is indeed the case. On the other hand, if we assume that the claim is true for an arbitrarily fixed $i \in \mathbb{N}$ (the *induction hypothesis*), then we see that the claim is true for $i + 1$, because

$$\begin{aligned} 1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^{i+1}} &= \left(1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^i}\right) + \frac{1}{2^{i+1}} \\ &= 2 - \frac{1}{2^i} + \frac{1}{2^{i+1}} \quad \text{(by the induction hypothesis)} \\ &= 2 - \frac{1}{2^{i+1}}. \end{aligned}$$

Thus, by the Principle of Mathematical Induction, we conclude that (2) holds. We shall use this principle numerous times throughout the text. Here is another example.

Exercise 27 Let $(X, +, \cdot, \geq)$ be an ordered field. Use the Principle of Mathematical Induction to prove the following generalization of the triangle inequality: For any $m \in \mathbb{N}$,

$$|x_1 + \cdots + x_m| \leq |x_1| + \cdots + |x_m| \qquad \text{for all } x_1, \ldots, x_m \in X.$$

Adjoining to $\mathbb{N}$ an element to serve as the additive identity, namely the *zero*, we obtain the set of all *nonnegative integers*, which is denoted as $\mathbb{Z}_+$. In turn, adjoining to $\mathbb{Z}_+$ the set $\{-1, -2, \ldots\}$ of all *negative integers* (whose construction would mimic that of $\mathbb{N}$), we obtain the set $\mathbb{Z}$ of all *integers*. In the process, the binary operations $+$ and $\cdot$ are suitably extended from $\mathbb{N}$ to $\mathbb{Z}$ so that they become binary operations on $\mathbb{Z}$ that satisfy all of the field axioms except the existence of multiplicative inverse elements.

Unfortunately, the nonexistence of multiplicative inverses is a serious problem. For instance, while an equation like $2x = 1$ makes sense in $\mathbb{Z}$, it cannot possibly be solved in $\mathbb{Z}$. To be able to solve such linear equations, we need to extend $\mathbb{Z}$ to a *field*. Doing this (in the minimal way) leads us to the set $\mathbb{Q}$ of all *rational numbers*, which can be thought of as the collection of all fractions $\frac{m}{n}$ with $m, n \in \mathbb{Z}$ and $n \neq 0$. The operations $+$ and $\cdot$ are extended to $\mathbb{Q}$ in the natural way (so that, for instance, the additive and multiplicative inverses of $\frac{m}{n}$ are $-\frac{m}{n}$ and $\frac{n}{m}$, respectively, provided that $m, n \neq 0$). Moreover, the standard order $\geq$ on $\mathbb{Z}$ (which is deduced from the *successor relation* that leads to the construction of $\mathbb{N}$) is also extended to $\mathbb{Q}$ in the straightforward manner.[21] The resulting algebraic system, which we denote simply as $\mathbb{Q}$ instead of the fastidious $(\mathbb{Q}, +, \cdot, \geq)$, is significantly richer than $\mathbb{Z}$. In particular, the following is true.

Proposition 4
$\mathbb{Q}$ is an ordered field.

Since we did not give a formal construction of $\mathbb{Q}$, we cannot prove this fact here.[22] But it is certainly good to know that all algebraic properties of an ordered field are possessed by $\mathbb{Q}$. For instance, thanks to Proposition 4, Example 7, and Exercise 25, the triangle inequality and the standard rules of exponentiation are valid in $\mathbb{Q}$.

2.3 Real Numbers

Although it is far superior to that of $\mathbb{Z}$, the structure of $\mathbb{Q}$ is nevertheless not strong enough to deal with many worldly matters. For instance, if we take a square with sides having length one, and attempt to compute the length

[21] [*Only for the formalists*] These definitions are meaningful only insofar as one knows the operation of "division" (and we don't, since the binary operation / is not defined on $\mathbb{Z}$). As noted in Section 1.3, the proper approach is to define $\mathbb{Q}$ as the set of equivalence classes $[(m, n)]_\sim$ where the equivalence relation $\sim$ is defined on $\mathbb{Z} \times (\mathbb{Z}\backslash\{0\})$ by $(m, n) \sim (k, l)$ iff $ml = nk$. The addition and multiplication operations on $\mathbb{Q}$ are then defined as $[(m, n)]_\sim + [(k, l)]_\sim = [(ml + nk, nl)]_\sim$ and $[(m, n)]_\sim[(k, l)]_\sim = [(mk, nl)]_\sim$. Finally, the linear order $\geq$ on $\mathbb{Q}$ is defined via the ordering of integers as follows: $[(m, n)]_\sim \geq [(k, l)]_\sim$ iff $ml \geq nk$.

[22] If you followed the previous footnote, you should be able to supply a proof, *assuming* the usual properties of $\mathbb{Z}$.

r of its diagonal, we would be in trouble if we were to use only the rational numbers. After all, we know from planar geometry (from the Pythagorean Theorem, to be exact) that r must satisfy the equation $r^2 = 2$. The trouble is that no rational number is equal to the task. Suppose that $r^2 = 2$ holds for some $r \in \mathbb{Q}$. We may then write $r = \frac{m}{n}$ for some integers $m, n \in \mathbb{Z}$ with $n \neq 0$. Moreover, we can assume that m and n do not have a common factor. (Right?) Then $m^2 = 2n^2$, from which we conclude that m^2 is an even integer. But this is possible only if m is an even integer itself. (Why?) Hence we may write $m = 2k$ for some $k \in \mathbb{Z}$. Then we have $2n^2 = m^2 = 4k^2$ so that $n^2 = 2k^2$, that is, n^2 is an even integer. But then n is even, which means 2 is a common factor of both m and n, a contradiction.

This observation is easily generalized:

Exercise 28 Prove: If a is a positive integer such that $a \neq b^2$ for any $b \in \mathbb{Z}$, then there is no rational number r such that $r^2 = a$.[23]

Here is another way of looking at the problem above. There are certainly two rational numbers p and q such that $p^2 > 2 > q^2$, but now we know that there is no $r \in \mathbb{Q}$ with $r^2 = 2$. It is as if there were a "hole" in the set of rational numbers. Intuitively speaking, then, we wish to *complete* $\mathbb{Q}$ by filling up its holes with "new" numbers. And, lo and behold, doing this leads us to the set $\mathbb{R}$ of real numbers. (*Note.* Any member of the set $\mathbb{R}\backslash\mathbb{Q}$ is said to be an **irrational number**.)

This is not the place to get into the formal details of how such a completion would be carried out, so we will leave things at this fairy tale level. However, we remark that, during this completion, the operations of addition and multiplication are extended to $\mathbb{R}$ in such a way as to make it a field. Similarly, the order $\geq$ is extended from $\mathbb{Q}$ to $\mathbb{R}$ nicely, so a great many algebraic properties of $\mathbb{Q}$ are inherited by $\mathbb{R}$.

[23] This fact provides us with lots of real numbers that are not rational, e.g., $\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{6}, \ldots$, etc. There are many other *irrational* numbers. (Indeed, there is a sense in which there are more of such numbers than of rational numbers.) However, it is often difficult to prove the irrationality of a number. For instance, while the problem of incommensurability of the circumference and the diameter of a circle was studied since the time of Aristotle, it was not until 1766 that a complete proof of the irrationality of π was given. Fortunately, elementary proofs of the fact that $\pi \notin \mathbb{Q}$ are since then formulated. If you are curious about this issue, you might want to take a look at Chapter 6 of Aigner and Ziegler (1999), where a brief and self-contained treatment of several such results (e.g., $\pi^2 \notin \mathbb{Q}$ and $e \notin \mathbb{Q}$) is given.

Proposition 5
$\mathbb{R}$ is an ordered field.

Notation. Given Propositions 4 and 5, it is natural to adopt the notations $\mathbb{Q}_+, \mathbb{Q}_{++}, \mathbb{Q}_-$, and $\mathbb{Q}_{--}$ to denote, respectively, the nonnegative, positive, nonpositive, and negative subsets of $\mathbb{Q}$, and similarly for $\mathbb{R}_+, \mathbb{R}_{++}, \mathbb{R}_-$, and $\mathbb{R}_{--}$.

There are, of course, many properties that $\mathbb{R}$ satisfies but $\mathbb{Q}$ does not. To make this point clearly, let us restate the order-theoretic properties given in Exercise 15 for the special case of $\mathbb{R}$. A set $S \subseteq \mathbb{R}$ is said to be **bounded from above** if it has an $\geq$-upper bound, that is, if there is a real number a such that $a \geq s$ for all $s \in S$. In what follows, we shall refer to an $\geq$-upper bound (or the $\geq$-maximum, etc.) of a set in $\mathbb{R}$ simply as an upper bound (or the maximum, etc.) of that set. Moreover, we will denote the $\geq$-supremum of a set $S \subseteq \mathbb{R}$ by $\sup S$. That is, $s^* = \sup S$ iff s^* is an upper bound of S, and $a \geq s^*$ holds for all upper bounds a of S. (The number $\sup S$ is often called the **least upper bound** of S.) The lower bounds of S and $\inf S$ are defined dually. (The number $\inf S$ is called the **greatest lower bound** of S.)

The main difference between $\mathbb{Q}$ and $\mathbb{R}$ is captured by the following property:

The Completeness Axiom
Every nonempty subset S of $\mathbb{R}$ that is bounded from above has a supremum in $\mathbb{R}$. That is, if $\emptyset \neq S \subseteq \mathbb{R}$ is bounded from above, then there exists a real number s^* such that $s^* = \sup S$.

It is indeed this property that distinguishes $\mathbb{R}$ from $\mathbb{Q}$. For instance, $S := \{q \in \mathbb{Q} : q^2 < 2\}$ is obviously a set in $\mathbb{Q}$ that is bounded from above. Yet $\sup S$ does *not* exist in $\mathbb{Q}$, as we will prove shortly. But $\sup S$ exists in $\mathbb{R}$ by the Completeness Axiom (or, as is usually said, by the *completeness of the reals*), and of course, $\sup S = \sqrt{2}$. (This is not entirely trivial; we will prove it shortly.) In an intuitive sense, therefore, $\mathbb{R}$ is obtained from $\mathbb{Q}$ by filling the "holes" in $\mathbb{Q}$ to obtain an ordered field that satisfies the Completeness Axiom. We thus say that $\mathbb{R}$ is a **complete ordered field**.[24]

In the rest of this section, we explore some important consequences of the completeness of the reals. Let us first warm up with an elementary exercise that tells us why we did not need to assume anything about the *greatest lower bound* of a set when stating the Completeness Axiom.

$\boxed{\text{Exercise 29}}^{\text{H}}$ Prove: If $\emptyset \neq S \subseteq \mathbb{R}$ and there exists an $a \in \mathbb{R}$ with $a \leq s$ for all $s \in S$, then $\inf S \in \mathbb{R}$.

Here is a result that shows how powerful the Completeness Axiom really is.

Proposition 6

(a) (*The Archimedean Property*) *For any* $(a, b) \in \mathbb{R}_{++} \times \mathbb{R}$, *there exists an* $m \in \mathbb{N}$ *such that* $b < ma$.

(b) *For any* $a, b \in \mathbb{R}$ *such that* $a < b$, *there exists a* $q \in \mathbb{Q}$ *such that* $a < q < b$.[25]

Proof

(a) This is an immediate consequence of the completeness of $\mathbb{R}$. Indeed, if the claim was not true, then there would exist a real number $a > 0$ such that $\{ma : m \in \mathbb{N}\}$ is bounded from above. But then $s = \sup\{ma : m \in \mathbb{N}\}$ would be a real number, and hence $a > 0$ would imply that $s - a$ is not an upper bound of $\{ma : m \in \mathbb{N}\}$, that is, there exists an $m^* \in \mathbb{N}$ such that $s < (m^* + 1)a$, which is not possible in view of the choice of s.

[24] Actually, one can say a bit more in this junction. $\mathbb{R}$ is not only "a" complete ordered field, it is in fact "the" complete ordered field. To say this properly, let us agree to call an ordered field $(X, \oplus, \odot, \succcurlyeq)$ *complete* if $\sup_{\succcurlyeq} S \in X$ for any $S \in 2^X \backslash \{\emptyset\}$ that has an $\succcurlyeq$-upper bound in X. It turns out that any such ordered field is equivalent to $\mathbb{R}$ up to relabeling. That is, for any complete ordered field $(X, \oplus, \odot, \succcurlyeq)$, there exists a bijection $f : X \to \mathbb{R}$ such that $f(x \oplus y) = f(x) + f(y)$, $f(x \odot y) = f(x)f(y)$, and $x \succcurlyeq y$ iff $f(x) \geq f(y)$. (This is *the Isomorphism Theorem*. McShane and Botts (1959) prove this as Theorem 6.1 (of Chapter 1) in their classic treatment of real analysis (reprinted by Dover in 2005).)

[25] We thus say that the rationals are *order-dense* in the reals.

(b) Take any $a, b \in \mathbb{R}$ with $b - a > 0$. By the Archimedean Property, there exists an $m \in \mathbb{N}$ such that $m(b - a) > 1$, that is, $mb > ma + 1$. Define $n := \min\{k \in \mathbb{Z} : k > ma\}$.[26] Then $ma < n \leq 1 + ma < mb$ (why?), so letting $q := \frac{n}{m}$ completes the proof. ■

Exercise 30[H] Show that, for any $a, b \in \mathbb{R}$ with $a < b$, there exists a $c \in \mathbb{R}\backslash\mathbb{Q}$ such that $a < c < b$.

We will make use of Proposition 6.(b) (and hence the Archimedean Property, and hence the Completeness Axiom) on many occasions. Here is a quick illustration. Let $S := \{q \in \mathbb{Q} : q < 1\}$. What is $\sup S$? The natural guess is, of course, that it is 1. Let us prove this formally. First of all, note that S is bounded from above (by 1, in particular), so by the Completeness Axiom, we know that $\sup S$ is a real number. Thus, if $1 \neq \sup S$, then by definition of $\sup S$, we must have $1 > \sup S$. But then by Proposition 6.(b), there exists a $q \in \mathbb{Q}$ such that $1 > q > \sup S$. Yet the latter inequality is impossible, since $q \in S$ and $\sup S$ is an upper bound of S. Hence, $1 = \sup S$.

One can similarly compute the sup and inf of other sets, although the calculations are bound to be a bit tedious at this primitive stage of the development. For instance, let us show that

$$\sup\{q \in \mathbb{Q} : q^2 < 2\} = \sqrt{2}.$$

That is, where $S := \{q \in \mathbb{Q} : q^2 < 2\}$, we wish to show that $\sup S$ is a real number the square of which equals 2. Notice first that S is a nonempty set that is bounded from above, so the Completeness Axiom ensures that $s := \sup S$ is real number. Suppose we have $s^2 > 2$. Then $s^2 - 2 > 0$, so by the Archimedean Property there exists an $m \in \mathbb{N}$ such that $m(s^2 - 2) > 2s$. Then

$$\left(s - \frac{1}{m}\right)^2 = s^2 - \frac{2s}{m} + \frac{1}{m^2} > s^2 - (s^2 - 2) = 2,$$

which means that $\left(s - \frac{1}{m}\right)^2 > q^2$ for all $q \in S$. But then $s - \frac{1}{m}$ is an upper bound for S, contradicting that s is the *smallest* upper bound for S. It follows that we have $s^2 \leq 2$. Good, let us now look at what happens if we have $s^2 < 2$.

[26] By the Archimedean Property, there must exist a $k \in \mathbb{N}$ such that $k > ma$, so n is well-defined.

In that case we use again the Archimedean Property to find an $m \in \mathbb{N}$ such that $m(2-s^2) > 4s$ and $m > \frac{1}{2s}$. Then

$$\left(s+\frac{1}{m}\right)^2 = s^2 + \frac{2s}{m} + \frac{1}{m^2} < s^2 + \frac{2s}{m} + \frac{2s}{m} < s^2 + (2-s^2) = 2.$$

But, by Proposition 6.(b), there exists a $q \in \mathbb{Q}$ with $s < q < s + \frac{1}{m}$. It follows that $s < q \in S$, which is impossible, since s is an upper bound for S. Conclusion: $s^2 = 2$. Put differently, the equation $x^2 = 2$ has a solution in $\mathbb{R}$, thanks to the Completeness Axiom, while it does not have a solution in $\mathbb{Q}$.

Exercise 31 Let S be a nonempty subset of $\mathbb{R}$ that is bounded from above. Show that $s^* = \sup S$ iff both of the following two conditions hold:

(i) $s^* \geq s$ for all $s \in S$;
(ii) for any $\varepsilon > 0$, there exists an $s \in S$ such that $s > s^* - \varepsilon$.

Exercise 32 Let A and B be two nonempty subsets of $\mathbb{R}$ that are bounded from above. Show that $A \subseteq B$ implies $\sup A \leq \sup B$, and that

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

Moreover, if $c \geq a$ for all $a \in A$, then $c \geq \sup A$.

Exercise 33 Let $S \subseteq \mathbb{R}$ be a nonempty set that is bounded from below. Prove that $\inf S = -\sup(-S)$, where $-S := \{-s \in \mathbb{R} : s \in S\}$. Use this result to state and prove the versions of the results reported in Exercises 31 and 32 for nonempty subsets of $\mathbb{R}$ that are bounded from below.

2.4 Intervals and $\overline{\mathbb{R}}$

For any real numbers a and b with $a < b$, the open interval (a,b) is defined as $(a,b) := \{t \in \mathbb{R} : a < t < b\}$, and the semiopen intervals $(a,b]$ and $[a,b)$ are defined as $(a,b] := (a,b) \cup \{b\}$ and $[a,b) := \{a\} \cup (a,b)$, respectively.[27] Finally, the closed interval $[a,b]$ is defined as $[a,b] := \{t \in \mathbb{R} : a \leq t \leq b\}$. Any one of these intervals is said to be **bounded** and of **length** $b-a$. Any

[27] The French tradition is to denote these sets as $]a,b[$, $]a,b]$ and $[a,b[$, respectively. While this convention has the advantage of avoiding use of the same notation for ordered pairs and open intervals, it is not commonly adopted in the literature.

one of them is called **nondegenerate** if $b - a > 0$. In this book, when we write (a, b) or $(a, b]$ or $[a, b)$, we always mean that these intervals are nondegenerate. (We allow for $a = b$ when we write $[a, b]$, however.) We also adopt the following standard notation for **unbounded intervals**: $(a, \infty) := \{t \in \mathbb{R} : t > a\}$ and $[a, \infty) := \{a\} \cup (a, \infty)$. The unbounded intervals $(-\infty, b)$ and $(-\infty, b]$ are defined similarly. By an **open interval**, we mean an interval of the form (a, b), (a, ∞), $(-\infty, b)$, or $\mathbb{R}$; the **closed intervals** are defined similarly.

We have $\sup(-\infty, b) = \sup(a, b) = \sup(a, b] = b$ and $\inf(a, \infty) = \inf(a, b) = \inf[a, \infty) = a$. The Completeness Axiom says that every nonempty subset S of $\mathbb{R}$ that fits in an interval of finite length has both an inf and a sup. Conversely, if S does not fit in any interval of the form $(-\infty, b)$, then $\sup S$ does not exist (i.e., $\sup S \notin \mathbb{R}$). We sometimes indicate that this is the case by writing $\sup S = \infty$, but this is only a notational convention since ∞ is not a real number. (The statement $\inf S = -\infty$ is interpreted similarly.)

It will be convenient on occasion to work with a trivial extension of $\mathbb{R}$ that is obtained by adjoining to $\mathbb{R}$ the symbols $-\infty$ and ∞. The resulting set is called the set of **extended real numbers** and is denoted by $\overline{\mathbb{R}}$. By definition, $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. We *extend* the linear order $\geq$ of $\mathbb{R}$ to $\overline{\mathbb{R}}$ by letting

$$\infty > -\infty \quad \text{and} \quad \infty > t > -\infty \quad \text{for all } t \in \mathbb{R}, \tag{3}$$

and hence view $\overline{\mathbb{R}}$ itself as a loset. Interestingly, $\overline{\mathbb{R}}$ satisfies the Completeness Axiom. In fact, a major advantage of $\overline{\mathbb{R}}$ is that *every set S in $\overline{\mathbb{R}}$ has a $\geq$-infimum and a $\geq$-supremum.* (Just as in $\mathbb{R}$, we denote these extended real numbers as $\inf S$ and $\sup S$, respectively.) For, if $S \subseteq \overline{\mathbb{R}}$ and $\sup S \notin \mathbb{R}$, then (3) implies that $\sup S = \infty$, and similarly for $\inf S$.[28] In this sense, the supremum (infimum) of a set is quite a different notion than the maximum (minimum) of a set. Recall that, for any set S in $\overline{\mathbb{R}}$, the **maximum** of S, denoted as $\max S$, is defined to be the number $s^* \in S$, with $s^* \geq s$ for all $s \in S$. (The **minimum** of S, denoted as $\min S$, is defined dually.) Clearly, $\sup(0, 1) = 1$ but $\max(0, 1)$ does not exist. Of course, if S is finite, then both $\max S$ and $\min S$ exist. In general, we have $\sup S = \max S$ and $\inf S = \min S$, provided that $\max S$ and $\min S$ exist.

[28] Even $\sup \emptyset$ is well-defined in $\overline{\mathbb{R}}$. *Quiz.* $\sup \emptyset =$? (*Hint.* $\inf \emptyset > \sup \emptyset$!)

The interval notation introduced above extends readily to $\overline{\mathbb{R}}$. For instance, for any extended real $a > -\infty$, the semiopen interval $[-\infty, a)$ stands for the set $\{t \in \overline{\mathbb{R}} : -\infty \leq t < a\}$. Other types of intervals in $\overline{\mathbb{R}}$ are defined similarly. Clearly, $\min[-\infty, a) = \inf[-\infty, a) = -\infty$ and $\max[-\infty, a) = \sup[-\infty, a) = a$.

Finally, we extend the standard operations of addition and multiplication to $\overline{\mathbb{R}}$ by means of the following definitions: For any $t \in \mathbb{R}$,

$$t + \infty := \infty + t := \infty, \quad t + -\infty := -\infty + t := -\infty, \quad \infty + \infty := \infty,$$

$$-\infty + -\infty := -\infty, \quad t.\infty := \infty.t := \begin{cases} \infty, & \text{if } 0 < t \leq \infty \\ -\infty, & \text{if } -\infty \leq t < 0 \end{cases},$$

and

$$t(-\infty) := (-\infty)t := \begin{cases} -\infty, & \text{if } 0 < t \leq \infty \\ \infty, & \text{if } -\infty \leq t < 0 \end{cases}.$$

WARNING. The expressions $\infty + (-\infty)$, $-\infty + \infty$, $\infty \cdot 0$, and $0 \cdot \infty$ are left undefined, so $\overline{\mathbb{R}}$ cannot be considered a field.

EXERCISE 34 Letting $|t| := t$ for all $t \in [0, \infty]$, and $|t| := -t$ for all $t \in [-\infty, 0)$, show that $|a + b| \leq |a| + |b|$ for all $a, b \in \overline{\mathbb{R}}$ with $a + b \in \overline{\mathbb{R}}$. Also show that $|ab| = |a|\,|b|$ for all $a, b \in \overline{\mathbb{R}}\backslash\{0\}$.

3 Real Sequences

3.1 Convergent Sequences

By a **real sequence**, we mean a sequence in $\mathbb{R}$. The set of all real sequences is thus $\mathbb{R}^{\mathbb{N}}$, but recall that we denote this set instead by $\mathbb{R}^{\infty}$. We think of a sequence $(x_m) \in \mathbb{R}^{\infty}$ as *convergent* if there is a real number x such that the later terms of the sequence get arbitrarily close to x. Put precisely, (x_m) is said to **converge** to x if, for each $\varepsilon > 0$, there exists a real number M (that may depend on ε) such that $|x_m - x| < \varepsilon$ for all $m \in \mathbb{N}$ with $m \geq M$.[29]

[29] By the Archimedean Property, we can always choose M to be a natural number, and write "for all $m = M, M + 1, \ldots$" instead of "for all $m \in \mathbb{N}$ with $m \geq M$" in this definition. Since the fact that each m must be a natural number is clear from the context, one often writes simply "for all $m \geq M$" instead of either of these expressions (whether or not $M \in \mathbb{N}$).

In this case, we say that (x_m) is **convergent**, and x is the **limit** of (x_m). We describe this situation by writing $\lim_{m\to\infty} x_m = x$, or $\lim x_m = x$, or simply, $x_m \to x$ (as $m \to \infty$). In words, $x_m \to x$ means that, no matter how small $\varepsilon > 0$ is, all but finitely many terms of the sequence (x_m) are contained in the open interval $(x - \varepsilon, x + \varepsilon)$.[30]

A sequence that does not converge to a real number is called **divergent**. If, for every real number y, there exists an $M \in \mathbb{R}$ with $x_m \geq y$ for each $m \geq M$, then we say that (x_m) **diverges** (or **converges**) **to** ∞, or that "the limit of (x_m) is ∞," and write either $x_m \to \infty$ or $\lim x_m = \infty$. We say that (x_m) **diverges (**or **converges) to** $-\infty$, or that "the limit of (x_m) is $-\infty$," and write $x_m \to -\infty$ or $\lim x_m = -\infty$, if $-x_m \to \infty$. (See Figure 1.)

The idea is that the *tail* of a convergent real sequence approximates the limit of the sequence to any desired degree of accuracy. Some initial (finitely many) terms of the sequence may be quite apart from its limit point, but *eventually* all terms of the sequence accumulate around this limit. For instance, the real sequence $(\frac{1}{m})$ and $(y_m) := (1, 2, \ldots, 100, 1, \frac{1}{2}, \frac{1}{3}, \ldots)$ have the same long-run behavior—they both converge to 0—even though their first few terms are quite different from each other. The initial terms of the sequence have no say on the behavior of the tail of the sequence.

To see this more clearly, let us show formally that $\frac{1}{m} \to 0$. To this end, pick an arbitrary $\varepsilon > 0$, and ask if there is an $M \in \mathbb{R}$ large enough to guarantee that $\left|\frac{1}{m} - 0\right| = \left|\frac{1}{m}\right| < \varepsilon$ for all $m \geq M$. In this simple example, the choice is clear. By choosing M to be a number strictly greater than $\frac{1}{\varepsilon}$, we get the desired inequality straightaway. The point is that we can prove that $y_m \to 0$ analogously, except that we need to choose our threshold M larger in this case, meaning that we need to wait a bit longer (in fact, for 100 more "periods") for the terms of (y_m) to enter and never leave the interval $(0, \varepsilon)$.

For another example, note that $((-1)^m)$ and (m) are divergent real sequences. While there is no real number a such that all but finitely many terms of $((-1)^m)$ belong to $(a - \frac{1}{2}, a + \frac{1}{2})$, we have $\lim m = \infty$ by the

[30] While the "idea" of convergence of a sequence was around for some time, we owe this precise formulation to Augustin-Louis Cauchy (1789–1857). It would not be an exaggeration to say that Cauchy is responsible for the emergence of what is called real analysis today. (The same goes for complex analysis too, as a matter of fact.) Just to give you an idea, let me note that it was Cauchy who proved the Fundamental Theorem of Calculus (in 1822) as we know it today (although for uniformly continuous functions). Cauchy published 789 mathematical articles in his lifetime.

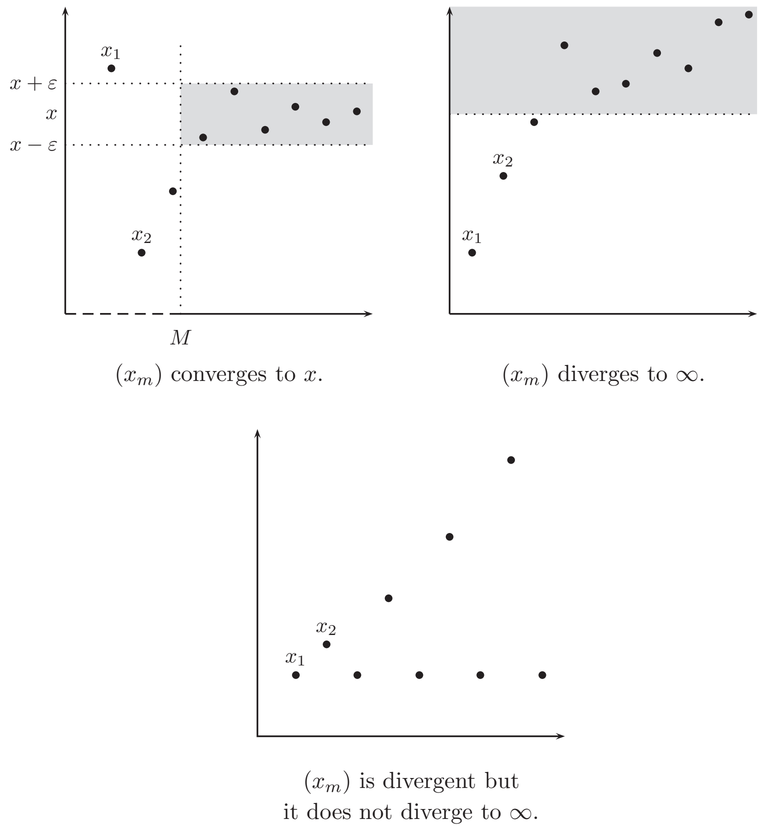


Figure 1

Archimedean Property. Also note that $\lim a^m = 0$ for any real number a with $|a| < 1$.[31] The following example is also very useful.

Lemma 1

For any real number $x \in \mathbb{R}$, there exists a sequence (q_m) of rational numbers and (p_m) of irrational numbers such that $q_m \to x$ and $p_m \to x$.

---

[31] *Quiz.* Prove this! *Hint.* Use the Principle of Mathematical Induction to obtain first the *Bernoulli Inequality*: $(1+t)^m \geq 1 + mt$ for any $(t, m) \in \mathbb{R} \times \mathbb{N}$. This inequality will make the proof very easy.

Proof

Take any $x \in \mathbb{R}$, and use Proposition 6.(b) to choose a $q_m \in (x, x + \frac{1}{m})$ for each $m \in \mathbb{N}$. For any $\varepsilon > 0$, by choosing any real number $M > \frac{1}{\varepsilon}$, we find that $\left|q_m - x\right| < \frac{1}{m} < \varepsilon$ for all $m \geq M$. Recalling Exercise 30, the second assertion is proved analogously. ■

A real sequence cannot have more than one limit. For, if (x_m) is a convergent real sequence such that $x_m \to x$ and $x_m \to y$ with $x \neq y$, then by choosing $\varepsilon := \frac{1}{2}\left|x - y\right|$, we can find an $M > 0$ large enough to guarantee that $|x_m - x| < \frac{\varepsilon}{2}$ and $\left|x_m - y\right| < \frac{\varepsilon}{2}$ for all $m \geq M$. Thanks to the triangle inequality, this yields the following contradiction:

$$\left|x - y\right| \leq |x - x_m| + \left|x_m - y\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon = \frac{1}{2}\left|x - y\right|.$$

Here is another simple illustration of how one works with convergent real sequences in practice. Suppose we are given a sequence $(x_m) \in \mathbb{R}^\infty$ with $x_m \to x \in \mathbb{R}$. We wish to show that if b is real number with $x_m \leq b$ for all m, then we have $x \leq b$. The idea is that if $x > b$ were the case, then, since the terms of (x_m) get eventually very close to x, we would have $x_m > b$ for m large enough. To say this formally, let $\varepsilon := x - b > 0$, and note that there exists an $M \in \mathbb{R}$ such that $|x_m - x| < \varepsilon$ for all $m \geq M$, so $x_M > x - \varepsilon = b$, which contradicts our main hypothesis. Amending this argument only slightly, we can state a more general fact: *For any* $-\infty \leq a < b \leq \infty$, *and convergent* $(x_m) \in [a, b]^\infty$, *we have* $\lim x_m \in [a, b]$.[32]

The following exercises may help you recall some other common tricks that come up when playing with convergent sequences.

Exercise 35 Let (x_m) and (y_m) be two real sequences such that $x_m \to x$ and $y_m \to y$ for some real numbers x and y. Prove:

(a) $|x_m| \to |x|$;

(b) $x_m + y_m \to x + y$;

(c) $x_m y_m \to xy$;

(d) $\frac{1}{x_m} \to \frac{1}{x}$, provided that $x, x_m \neq 0$ for each m.

[32] *Reminder.* For any nonempty subset S of $\mathbb{R}$, "$(x_m) \in S^\infty$" means that (x_m) is a real sequence such that $x_m \in S$ for each m. (Recall Section 1.6.)

Exercise 36[H] Let (x_m), (y_m) and (z_m) be real sequences such that $x_m \leq y_m \leq z_m$ for each m. Show that if $\lim x_m = \lim z_m = a$, then $y_m \to a$.

3.2 Monotonic Sequences

We say that a real sequence (x_m) is **bounded from above** if $\{x_1, x_2, \ldots\}$ is bounded from above, that is, if there exists a real number K with $x_m \leq K$ for all $m = 1, 2, \ldots$. By the Completeness Axiom, this is equivalent to saying that

$$\sup\{x_m : m \in \mathbb{N}\} < \infty.$$

Dually, (x_m) is said to be **bounded from below** if $\{x_1, x_2, \ldots\}$ is bounded from below, that is, if $\inf\{x_m : m \in \mathbb{N}\} > -\infty$. Finally, (x_m) is called **bounded** if it is bounded from both above and below, that is,

$$\sup\{|x_m| : m \in \mathbb{N}\} < \infty.$$

Boundedness is a property all convergent real sequences share. For, if all but finitely (say, M) many terms of a sequence are at most some $\varepsilon > 0$ away from a fixed number x, then this sequence is bounded either by $|x| + \varepsilon$ or by the largest of the first M terms (in absolute value). This is almost a proof, but let us write things out precisely anyway.

Proposition 7
Every convergent real sequence is bounded.

Proof
Take any $(x_m) \in \mathbb{R}^\infty$ with $x_m \to x$ for some real number x. Then there must exist a natural number M such that $|x_m - x| < 1$, and hence $|x_m| < |x| + 1$, for all $m \geq M$. But then $|x_m| \leq \max\{|x| + 1, |x_1|, \ldots, |x_M|\}$ for all $m \in \mathbb{N}$. ∎

The converse of Proposition 7 does not hold, of course. (Think of the sequence $((-1)^m)$, for instance.) However, there is one very important class of bounded sequences that always converge.

Definition

A real sequence (x_m) is said to be **increasing** if $x_m \leq x_{m+1}$ for each $m \in \mathbb{N}$, and **strictly increasing** if $x_m < x_{m+1}$ for each $m \in \mathbb{N}$. It is said to be **(strictly) decreasing** if $(-x_m)$ is (strictly) increasing. Finally, a real sequence which is either increasing or decreasing is referred to as a **monotonic** sequence.[33] If (x_m) is increasing and converges to $x \in \overline{\mathbb{R}}$, then we write $x_m \nearrow x$, and if it is decreasing and converges to $x \in \overline{\mathbb{R}}$, we write $x_m \searrow x$.

The following fact attests to the importance of monotonic sequences. We owe it to the Completeness Axiom.

Proposition 8

Every increasing (decreasing) real sequence that is bounded from above (below) converges.

Proof

Let $(x_m) \in \mathbb{R}^\infty$ be an increasing sequence which is bounded from above, and let $S := \{x_1, x_2, \ldots\}$. By the Completeness Axiom, $x := \sup S \in \mathbb{R}$. We claim that $x_m \nearrow x$. To show this, pick an arbitrary $\varepsilon > 0$. Since x is the least upper bound of S, $x - \varepsilon$ cannot be an upper bound of S, so $x_M > x - \varepsilon$ for some $M \in \mathbb{N}$. Since (x_m) is increasing, we must then have $x \geq x_m \geq x_M > x - \varepsilon$, so $|x_m - x| < \varepsilon$, for all $m \geq M$. The proof of the second claim is analogous. ■

Proposition 8 is an extremely useful observation. For one thing, monotonic sequences are not terribly hard to come by. In fact, within every real sequence there is one!

Proposition 9

Every real sequence has a monotonic subsequence.

[33] That is, an increasing (decreasing) real sequence is an increasing (decreasing) real function on $\mathbb{N}$. Never forget that a real sequence is just a special kind of a real function.

Proof

(Thurston) Take any $(x_m) \in \mathbb{R}^\infty$ and define $S_m := \{x_m, x_{m+1}, \ldots\}$ for each $m \in \mathbb{N}$. If there is no maximum element in S_1, then it is easy to see that (x_m) has a monotonic subsequence. (Let $x_{m_1} := x_1$, let x_{m_2} be the first term in the sequence $(x_2, x_3, \ldots)$ greater than x_1, let x_{m_3} be the first term in the sequence $(x_{m_2+1}, x_{m_2+2}, \ldots)$ greater than x_{m_2}, and so on.) By the same logic, if for any $m \in \mathbb{N}$ there is no maximum element in S_m, then we are done. Assume, then, max S_m exists for each $m \in \mathbb{N}$. Now define the subsequence (x_{m_k}) recursively as follows:

$$x_{m_1} := \max S_1, \qquad x_{m_2} := \max S_{m_1+1}, \qquad x_{m_3} := \max S_{m_2+1}, \qquad \ldots.$$

Clearly, (x_{m_k}) is decreasing. ■

Putting the last two observations together, we get the following famous result as an immediate corollary.

The Bolzano-Weierstrass Theorem.[34]
Every bounded real sequence has a convergent subsequence.

Exercise 37 Show that every unbounded real sequence has a subsequence that diverges to either ∞ or $-\infty$.

Exercise 38[H] Let S be a nonempty bounded subset of $\mathbb{R}$. Show that there is an increasing sequence $(x_m) \in S^\infty$ such that $x_m \nearrow \sup S$, and a decreasing sequence $(y_m) \in S^\infty$ such that $y_m \searrow \inf S$.

Exercise 39 For any real number x and $(x_m) \in \mathbb{R}^\infty$, show that $x_m \to x$ iff every subsequence of (x_m)has itself a subsequence that converges to x.

[34] Bernhard Bolzano (1781–1848) was one of the early founders of real analysis. Much of his work was found too unorthodox by his contemporaries and so was ignored. The depth of his discoveries was understood only after his death, after a good number of them were rediscovered and brought to light by Karl Weierstrass (1815–1897). The Bolzano-Weierstrass Theorem is perhaps best viewed as an outcome of an intertemporal (in fact, intergenerational) collaboration.

Exercise 40[H] (*The Cauchy Criterion*) We say that an $(x_m) \in \mathbb{R}^\infty$ is a **real Cauchy sequence** if, for any $\varepsilon > 0$, there exists an $M \in \mathbb{R}$ such that $|x_k - x_l| < \varepsilon$ for all $k, l \geq M$.

(a) Show that every real Cauchy sequence is bounded.

(b) Show that every real Cauchy sequence converges.

A double real sequence $(x_{kl}) \in \mathbb{R}^{\infty\times\infty}$ is said to **converge** to $x \in \mathbb{R}$, denoted as $x_{kl} \to x$, if, for each $\varepsilon > 0$, there exists a real number M (that may depend on ε) such that $|x_{kl} - x| < \varepsilon$ for all $k, l \geq M$. The following exercise tells us when one can conclude that (x_{kl}) converges by looking at the behavior of (x_{kl}) first as $k \to \infty$ and then as $l \to \infty$ (or vice versa).

Exercise 41[H] (*The Moore-Osgood Theorem*) Take any $(x_{kl}) \in \mathbb{R}^{\infty\times\infty}$ such that there exist $(y_k) \in \mathbb{R}^\infty$ and $(z_l) \in \mathbb{R}^\infty$ such that

(i) for any $\varepsilon > 0$, there exists an $L \in \mathbb{N}$ such that $\left|x_{kl} - y_k\right| < \varepsilon$ for all $k \geq 1$ and $l \geq L$; and

(ii) for any $\varepsilon > 0$ and $l \in \mathbb{N}$, there exists a $K_l \in \mathbb{N}$ such that $|x_{kl} - z_l| < \varepsilon$ for all $k \geq K_l$.

(a) Prove that there exists an $x \in \mathbb{R}$ such that $x_{kl} \to x$ and

$$\lim_{k\to\infty} \lim_{l\to\infty} x_{kl} = x = \lim_{l\to\infty} \lim_{k\to\infty} x_{kl}. \tag{4}$$

(b) Check if (4) holds for the double sequence $\left(\frac{kl}{k^2+l^2}\right)$. What goes wrong?

3.3 Subsequential Limits

Any subsequence of a convergent real sequence converges to the limit of the mother sequence. (Why?) What is more, even if the mother sequence is divergent, it may still possess a convergent subsequence (as in the Bolzano-Weierstrass Theorem). This suggests that we can get at least some information about the long-run behavior of a sequence by studying those points to which at least one subsequence of the sequence converges. Given any $(x_m) \in \mathbb{R}^\infty$, we say that $x \in \overline{\mathbb{R}}$ is a **subsequential limit** of (x_m) if there exists a subsequence (x_{m_k}) with $x_{m_k} \to x$ (as $k \to \infty$). For instance, -1 and 1 are the only subsequential limits of $((-1)^m)$, and -1, 1 and ∞ are the only subsequential limits of the sequence (x_m) where $x_m = -1$ for each odd m not divisible by 3, $x_m = 1$ for each even m, and $x_m = m$ for each odd m divisible by 3.

If x is a subsequential limit of (x_m), then we understand that (x_m) visits the interval $(x-\varepsilon, x+\varepsilon)$ *infinitely often*, no matter how small $\varepsilon > 0$ is. It is in this sense that subsequential limits give us asymptotic information about the long-run behavior of a real sequence. Of particular interest in this regard are the largest and smallest subsequential limits of a real sequence. These are called the **limit superior** (abbreviated as lim sup) and **limit inferior** (abbreviated as lim inf) of a real sequence.

Definition

For any $x \in \mathbb{R}$ and $(x_m) \in \mathbb{R}^\infty$, we write $\limsup x_m = x$ if

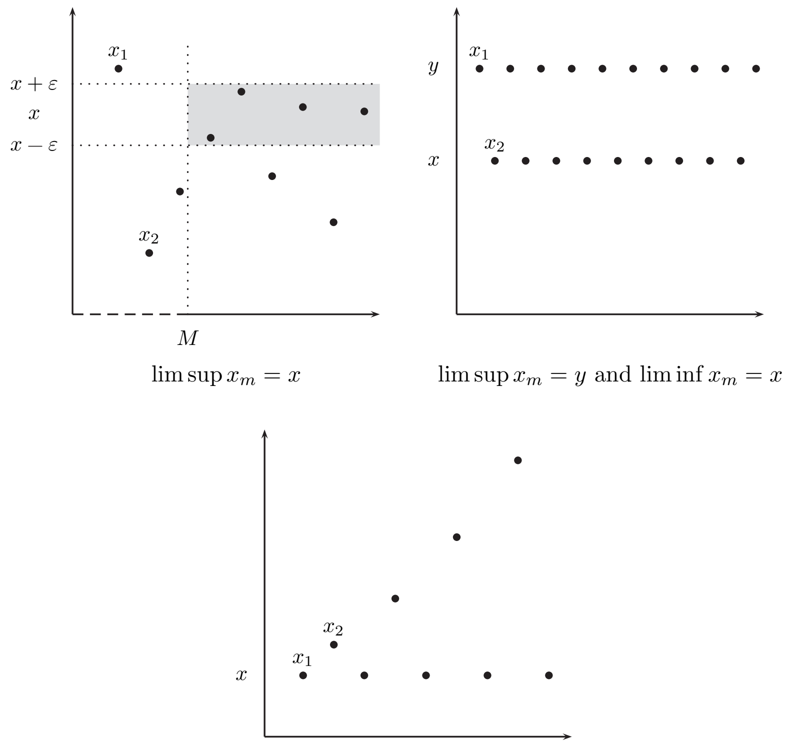
(i) for any $\varepsilon > 0$, there exists an $M > 0$ such that $x_m < x + \varepsilon$ for all $m \geq M$,

(ii) for any $\varepsilon > 0$ and $m \in \mathbb{N}$, there exists an integer $k > m$ such that $x_k > x - \varepsilon$.

We write $\limsup x_m = \infty$ if ∞ is a subsequential limit of (x_m); and $\limsup x_m = -\infty$ if $x_m \to -\infty$. The expression $\liminf x_m$ is defined dually (or by letting $\liminf x_m := -\limsup(-x_m)$).

If $\limsup x_m = x \in \mathbb{R}$, we understand that all but finitely many terms of the sequence are smaller than $x + \varepsilon$, no matter how small $\varepsilon > 0$ is. (Such a sequence is thus bounded from above, but it need not be bounded from below.) If $x = \lim x_m$ was the case, we could say in addition to this that all but finitely many terms of (x_m) are also larger than $x - \varepsilon$, no matter how small $\varepsilon > 0$ is. When $x = \limsup x_m$, however, all we can say in this regard is that infinitely many terms of (x_m) are larger than $x - \varepsilon$, no matter how small $\varepsilon > 0$ is. That is, if $x = \limsup x_m$, then the terms of the sequence (x_m) need not accumulate around x; it is just that all but finitely many of them are in $(-\infty, x+\varepsilon)$, and infinitely many of them are in $(x-\varepsilon, x+\varepsilon)$, no matter how small $\varepsilon > 0$ is. (See Figure 2.) The expression $\liminf x_m = x$ is similarly interpreted. For instance, $\lim(-1)^m$ does not exist, but $\limsup(-1)^m = 1$ and $\liminf(-1)^m = -1$.

It is easy to see that any real sequence (x_m) has a monotonic subsequence (x_{m_k}) such that $x_{m_k} \to \limsup x_m$. (For, $\limsup x_m$ is a subsequential limit of (x_m) (why?), so the claim obtains upon applying Proposition 9 to a



$\limsup x_m = x$ $\qquad$ $\limsup x_m = y$ and $\liminf x_m = x$

$\limsup x_m = \infty$ and $\liminf x_m = x$

Figure 2

subsequence of (x_m) that converges to $\limsup x_m$.) Of course, the analogous claim is true for $\liminf\, x_m$ as well. It also follows readily from the definitions that, for any $(x_m) \in \mathbb{R}^\infty$,

$$\liminf x_m \leq \limsup x_m,$$

and

$$(x_m) \text{ is convergent} \quad \text{iff} \quad \liminf x_m = \limsup x_m.$$

(Right?) Thus, to prove that a real sequence (x_m) converges, it is enough to show that $\liminf x_m \geq \limsup x_m$, which is sometimes easier than adopting the direct approach. The following exercises outline some other facts concerning the lim sup and lim inf of real sequences. If you're not already familiar with these concepts, it is advisable that you work through these before proceeding further.

Exercise 42 Let (x_m) be a real sequence and $x \in \overline{\mathbb{R}}$. Show that the following statements are equivalent:
(i) $\limsup x_m = x$.
(ii) x is the largest subsequential limit of (x_m).
(iii) $x = \inf\{\sup\{x_m, x_{m+1}, \ldots\} : m = 1, 2, \ldots\}$.
State and prove the analogous result for the $\liminf$ of (x_m).

A Corollary of Exercise 42. The lim sup and lim inf of any real sequence exist in $\overline{\mathbb{R}}$.

Exercise 43[H] Prove: For any bounded real sequences (x_m) and (y_m), we have

$$\begin{aligned}\liminf x_m + \liminf y_m &\leq \liminf (x_m + y_m)\\ &\leq \limsup (x_m + y_m)\\ &\leq \limsup x_m + \limsup y_m.\end{aligned}$$

Also, give an example for which *all* of these inequalities hold strictly.

Exercise 44 Prove: For any $x \geq 0$ and $(x_m), (y_m) \in \mathbb{R}^\infty$ with $x_m \to x$, we have $\limsup x_m y_m = x \limsup y_m$.

3.4 Infinite Series

Let (x_m) be a real sequence. We define

$$\sum_{i=1}^{m} x_i := x_1 + \cdots + x_m \quad \text{and} \quad \sum_{i=k}^{m} x_i := \sum_{i=1}^{m-k+1} x_{i+k-1}$$

for any $m \in \mathbb{N}$ and $k \in \{1, \ldots, m\}$.[35] For simplicity, however, we often write $\sum^m x_i$ for $\sum_{i=1}^m x_i$ within the text. For any nonempty finite subset S of $\mathbb{N}$, we write $\sum_{i\in S} x_i$ to denote the sum of all terms of (x_m) the indices of which belong to S.[36]

[35] There is no ambiguity in the definition of $\sum_{i=1}^m x_i$, precisely because the addition operation on $\mathbb{R}$ is associative.

[36] Formally speaking, $\sum_{i\in S} x_i := \sum_{i=1}^{|S|} x_{\sigma(i)}$, where σ is *any* bijection from $\{1, \ldots, |S|\}$ onto S.

Convention. For any $(x_m) \in \mathbb{R}^\infty$, we let $\sum_{i \in \emptyset} x_i := 0$ in this text. This is nothing more than a notational convention.

By an **infinite series**, we mean a real sequence of the form $(\sum^m x_i)$ for some $(x_m) \in \mathbb{R}^\infty$. When the limit of this sequence exists in $\overline{\mathbb{R}}$, we denote it as $\sum_{i=1}^\infty x_i$, but, again, we write $\sum^\infty x_i$ for $\sum_{i=1}^\infty x_i$ within the text. That is,

$$\sum_{i=1}^{\infty} x_i = \lim_{m\to\infty} \sum_{i=1}^{m} x_i,$$

provided that $(\sum^m x_i)$ converges in $\overline{\mathbb{R}}$. Similarly,

$$\sum_{i=k}^{\infty} x_i = \sum_{i=1}^{\infty} x_{i+k-1}, \quad k = 1, 2, \ldots,$$

provided that the right-hand side is well-defined. We say that an infinite series $(\sum^m x_i)$ is **convergent** if it has a finite limit (i.e., $\sum^\infty x_i \in \mathbb{R}$). In this case, with a standard abuse of terminology, we say that "the series $\sum^\infty x_i$ is convergent." If $(\sum^m x_i)$ diverges to ∞ or $-\infty$, that is, $\sum^\infty x_i \in \{-\infty, \infty\}$, then we say that the series is **divergent**. With the same abuse of terminology, we say then that "the series $\sum^\infty x_i$ is divergent."

Warning. In the present terminology, $\sum^\infty x_i$ may not be well-defined. For instance, the infinite series $(\sum^m (-1)^i)$ does not have a limit, so the notation $\sum^\infty (-1)^i$ is meaningless. Before dealing with an object like $\sum^\infty x_i$ in practice, you should first make sure that it is well-defined.

It is useful to note that the convergence of $\sum^\infty x_i$ implies $\lim x_m = 0$, but not conversely. For,

$$\lim_{m\to\infty} x_m = \lim_{m\to\infty} \left(\sum_{i=1}^{m+1} x_i - \sum_{i=1}^{m} x_i \right) = \lim_{m\to\infty} \sum_{i=1}^{m+1} x_i - \lim_{m\to\infty} \sum_{i=1}^{m} x_i = 0$$

where, given that $(\sum^m x_i)$ is convergent, the second equality follows from Exercise 35.(b). (What about the third equality?) The series $\sum^\infty \frac{1}{i}$, on the other hand, diverges to ∞, so the converse of this observation does not hold in general.[37]

Here are a few examples that come up frequently in applications.

[37] Consider the sequence $(y_m) := \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \ldots\right)$, and check that $\sum^\infty \frac{1}{i} \geq \sum^\infty y_i = \infty$.

Example 8

[1] $\sum^\infty \frac{1}{i^\alpha}$ *converges iff* $\alpha > 1$. This is easily proved by calculus: For any $m \in \mathbb{N}$ and $\alpha > 1$,

$$\sum_{i=1}^{m} \frac{1}{i^\alpha} \leq 1 + \int_1^m \frac{1}{t^\alpha} dt = 1 + \frac{1}{\alpha-1}\left(1 - \frac{1}{m^{\alpha-1}}\right) < 1 + \frac{1}{\alpha-1} = \frac{\alpha}{\alpha-1}.$$

(Draw a picture to see why the first inequality is true.) Thus, $\sum^\infty \frac{1}{i^\alpha} \leq \frac{\alpha}{\alpha-1}$ whenever $\alpha > 1$. Conversely, $\sum^\infty \frac{1}{i^\alpha} \geq \sum^\infty \frac{1}{i} = \infty$ for any $\alpha \leq 1$.

[2] $\sum^\infty \frac{1}{2^i} = 1$. For, by using (2), we have $\lim \sum^m \frac{1}{2^i} = \lim\left(1 - \frac{1}{2^i}\right) = 1$. The next example generalizes this useful observation.

[3] $\sum^\infty \delta^i = \frac{\delta}{1-\delta}$ for any $-1 < \delta < 1$. To prove this, observe that

$$(1 + \delta + \cdots + \delta^m)(1 - \delta) = 1 - \delta^{m+1}, \qquad m = 1, 2, \ldots$$

so that, for any $\delta \neq 1$, we have

$$\sum_{i=1}^{\infty} \delta^i = \lim_{m\to\infty} \sum_{i=1}^{m} \delta^i = \lim_{m\to\infty} \frac{1-\delta^{m+1}}{1-\delta} - 1 = \frac{\delta - \lim \delta^{m+1}}{1-\delta}.$$

But when $|\delta| < 1$, we have $\lim \delta^{m+1} = 0$ (as you were asked to prove about ten pages ago), and hence the claim. □

$\boxed{\text{Exercise 45}}$ [H] For any infinite series $(\sum^m x_i)$, prove:
(a) If $\sum^\infty x_i$ converges, then $\lim_{k\to\infty} \sum_{i=k}^\infty x_i = 0$;
(b) If $\sum^\infty x_i$ converges, $\left|\sum_{i=k}^\infty x_i\right| \leq \sum_{i=k}^\infty |x_i|$ for any $k \in \mathbb{N}$.

Exercise 46 Prove: If $(x_m) \in \mathbb{R}^\infty$ is a decreasing sequence such that $\sum^\infty x_i$ converges, then $mx_m \to 0$.

Exercise 47 Let $0! := 1$, and define $m! := ((m-1)!)\,m$ for any $m \in \mathbb{N}$. Prove that $\lim\left(1 + \frac{1}{m}\right)^m = 1 + \sum^\infty \frac{1}{i!}$. (*Note*. The common value of these expressions equals the real number $e = 2.71\ldots$. Can you show that e is irrational, by the way?)

*Exercise 48 [H] Let (x_m) be a real sequence, and $s_j := \sum^j x_i$, $j = 1, 2, \ldots$.
(a) Give an example to show that $(\frac{1}{m} \sum^m s_i)$ may converge even if $\sum^\infty x_i$ is not convergent.
(b) Show that if $\sum^\infty x_i$ is convergent, then $\lim \frac{1}{m} \sum^m s_i = \sum^\infty x_i$.

*Exercise 49 Prove *Tannery's Theorem*: Take any $(x_{kl}) \in \mathbb{R}^{\infty\times\infty}$ such that $\sum_{j=1}^{\infty} x_{kj}$ converges for each k and $(x_{1l}, x_{2l}, \ldots)$ converges for each l. If there exists a real sequence $(K_1, K_2, \ldots)$ such that $|x_{kl}| \leq K_l$ for each l, and $\sum^{\infty} K_i$ converges, then

$$\lim_{k\to\infty} \sum_{j=1}^{\infty} x_{kj} = \sum_{j=1}^{\infty} \lim_{k\to\infty} x_{kj}.$$

3.5 Rearrangement of Infinite Series

An issue that arises frequently in practice concerns the *rearrangement* of an infinite series. The question is if, and when, one can sum the terms of a given real sequence in different orders and obtain the same number in result (as it would be the case for any n-vector). Let's consider an example that points to the fact that the problem is not trivial. Fix any $\alpha \geq 1$. It is easily checked that $\sum^{\infty} \frac{1}{2i-1} = \infty$, so there must exist a smallest natural number $m_1 \geq 2$ with

$$\sum_{i=1}^{m_1} \frac{1}{2i-1} > \alpha.$$

Due to the choice of m_1, we have

$$\sum_{i=1}^{m_1} \frac{1}{2i-1} - \frac{1}{3} \leq \sum_{i=1}^{m_1} \frac{1}{2i-1} - \frac{1}{2m_1-1} \leq \sum_{i=1}^{m_1-1} \frac{1}{2i-1} \leq \alpha.$$

(Why?) Now let m_2 be the smallest number in $\{m_1+1, m_1+2, \ldots\}$ such that

$$\sum_{i=1}^{m_1} \frac{1}{2i-1} - \frac{1}{3} + \sum_{i=m_1+1}^{m_2} \frac{1}{2i-1} > \alpha$$

which implies

$$\sum_{i=1}^{m_1} \frac{1}{2i-1} - \frac{1}{3} + \sum_{i=m_1+1}^{m_2} \frac{1}{2i-1} - \frac{1}{9} \leq \alpha.$$

Continuing this way inductively, we obtain the sequence

$$(x_m) := \left(1, \tfrac{1}{3}, \ldots, \tfrac{1}{2m_1-1}, -\tfrac{1}{2}, \tfrac{1}{2m_1+1}, \ldots, \tfrac{1}{2m_2-1}, -\tfrac{1}{4}, \ldots\right).$$

The upshot is that we have $\sum^{\infty} x_m = \alpha$. (Check this!) Yet the sequence (x_m) is none other than the rearrangement of the sequence $\left(\frac{(-1)^{m+1}}{m}\right)$, so the series $\sum^\infty x_m$ is equal to the series $\sum^\infty \frac{(-1)^{i+1}}{i}$, except that it is summed in a different order. If such a rearrangement does not affect the value of the series, then we must conclude that $\sum^\infty \frac{(-1)^{i+1}}{i} = \alpha$. But this is absurd, for $\alpha \geq 1$ is completely arbitrary here; for instance, our conclusion yields $1 = \sum^\infty \frac{(-1)^{i+1}}{i} = 2$.[38]

This example tells us that one has to be careful in rearranging a given infinite series. Fortunately, there would be no problem in this regard if all terms of the series were nonnegative (or all were nonpositive). This simple fact is established next.

Proposition 10
For any given $(x_m) \in \mathbb{R}^\infty_+$ and bijection $\sigma : \mathbb{N} \to \mathbb{N}$, we have $\sum^\infty x_i = \sum^\infty x_{\sigma(i)}$.

Proof
Since σ is bijective, for any given $m \in \mathbb{N}$ there exist integers K_m and L_m such that $K_m \geq L_m \geq m$ and $\{1, \ldots, m\} \subseteq \{\sigma(1), \ldots, \sigma(L_m)\} \subseteq \{1, \ldots, K_m\}$. So, by nonnegativity, $\sum^m x_i \leq \sum^{L_m} x_{\sigma(i)} \leq \sum^{K_m} x_i$. Letting $m \to \infty$ yields the claim. ■

The following result gives another condition that is sufficient for any rearrangement of an infinite series to converge to the same limit as the original series. This result is often invoked when Proposition 10 does not apply because the series at hand may have terms that alternate in sign.

Dirichlet's Rearrangement Theorem
For any given $(x_m) \in \mathbb{R}^\infty$ and any bijection $\sigma : \mathbb{N} \to \mathbb{N}$, we have $\sum^\infty x_i = \sum^\infty x_{\sigma(i)}$, provided that $\sum^\infty |x_i|$ converges.

[38] This is not an idle example. According to a theorem of Bernhard Riemann that was published (posthumously) in 1867, for any convergent infinite series $\sum^\infty x_i$ such that $\sum^\infty |x_i| = \infty$ (such a series is called *conditionally convergent*), and any $\alpha \in \overline{\mathbb{R}}$, there exists a bijection $\sigma : \mathbb{N} \to \mathbb{N}$ such that $\sum^\infty x_{\sigma(i)} = \alpha$. (The proof is analogous to the one I gave above to show that the series $\sum^\infty \frac{(-1)^{i+1}}{i}$ can be rearranged to converge to *any* number.)

Bernhard Riemann (1826–1865) is a towering figure in mathematics. Argued by some to be the best mathematician who ever lived, in his short lifetime he revolutionized numerous subjects, ranging from complex and real analysis to geometry and mathematical physics. There are many books about the life and genius of this great man; I would recommend Laugwitz (1999) for an engaging account.

Proof

Note first that both $\sum^\infty x_i$ and $\sum^\infty x_{\sigma(i)}$ converge. (For instance, $\left(\sum^m x_{\sigma(i)}\right)$ is convergent, because $\sum^m x_{\sigma(i)} \leq \sum^m \left|x_{\sigma(i)}\right| \leq \sum^\infty |x_i|$ for any $m \in \mathbb{N}$.) Define $s_m := \sum^m x_i$ and $t_m := \sum^m x_{\sigma(i)}$ for each m, and let $s := \sum^\infty x_i$ and $t := \sum^\infty x_{\sigma(i)}$. We wish to show that $s = t$. For any $\varepsilon > 0$, we can clearly find an $M \in \mathbb{N}$ such that $\sum_{i=M}^\infty |x_i| < \frac{\varepsilon}{3}$ and $\sum_{i=M}^\infty \left|x_{\sigma(i)}\right| < \frac{\varepsilon}{3}$ (Exercise 45). Now choose $K \in \mathbb{N}$ large enough to guarantee that $\{1, \ldots, M\} \subseteq \{\sigma(1), \ldots, \sigma(K)\}$. Then, for any positive integer $k > K$, we have $\sigma(k) > M$, so letting $S_k := \{i \in \{1, \ldots, k\} : \sigma(i) > M\}$, we have

$$\begin{aligned}|t_k - s_M| &= \left|x_{\sigma(1)} + \cdots + x_{\sigma(k)} - x_1 - \cdots - x_M\right| \\ &\leq \sum_{i \in S_k} \left|x_{\sigma(i)}\right| \leq \sum_{i=k+1}^{\infty} |x_i| < \tfrac{\varepsilon}{3}.\end{aligned}$$

(Recall Exercise 27.) But then, for any $k > K$,

$$|t - s| \leq |t - t_k| + |t_k - s_M| + |s_M - s| < \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary here, this proves that $s = t$. ■

3.6 Infinite Products

Let (x_m) be a real sequence. We define

$$\prod_{i=1}^{m} x_i := x_1 \cdots x_m \qquad \text{for any } m = 1, 2, \ldots,$$

but write $\prod^m x_i$ for $\prod_{i=1}^m x_i$ within the text. By an **infinite product**, we mean a real sequence of the form $(\prod^m x_i)$ for some $(x_m) \in \mathbb{R}^\infty$. When the limit of this sequence exists in $\overline{\mathbb{R}}$, we denote it by $\prod_{i=1}^\infty x_i$. (But again, we often write $\prod^\infty x_i$ for $\prod_{i=1}^\infty x_i$ to simplify the notation.) That is,

$$\prod_{i=1}^{\infty} x_i := \lim_{m \to \infty} \prod_{i=1}^{m} x_i,$$

provided that $(\prod^m x_i)$ converges in $\overline{\mathbb{R}}$. We say that $(\prod^m x_i)$ (or, abusing terminology, $\prod^\infty x_i$) is **convergent** if $\lim \prod^m x_i \in \mathbb{R}$. If $(\prod^m x_i)$ diverges to ∞ or $-\infty$, that is, $\prod^\infty x_i \in \{-\infty, \infty\}$, then we say that the infinite product (or, $\prod^\infty x_i$) is **divergent**.

Exercise 50 For any $(x_m) \in \mathbb{R}^\infty$, prove the following statements.

(a) If there is an $0 < \varepsilon < 1$ such that $0 \leq |x_m| < 1 - \varepsilon$ for all but finitely many m, then $\prod^\infty x_i = 0$. (Can we take $\varepsilon = 0$ in this claim?)

(b) If $\prod^\infty x_i$ converges to a positive number, then $x_m \to 1$.

(c) If $x_m \geq 0$ for each m, then $\prod^\infty (1 + x_i)$ converges iff $\sum^\infty x_i$ converges.

4 Real Functions

This section is a refresher on the theory of real functions on $\mathbb{R}$. Because you are familiar with the elements of this theory, we go as fast as possible. Most of the proofs are either left as exercises or given only in brief sketches.

4.1 Basic Definitions

By a **real function** (or a **real-valued function**) on a nonempty set T, we mean an element of $\mathbb{R}^T$. If $f \in \mathbb{R}^T$ equals the real number a everywhere, that is, if $f(t) = a$ for all $t \in T$, then we write $f = a$. If $f \neq a$, it follows that $f(t) \neq a$ for *some* $t \in T$. Similarly, if $f, g \in \mathbb{R}^T$ are such that $f(t) \geq g(t)$ for all $t \in T$, we write $f \geq g$. If $f \geq g$ but not $g \geq f$, we then write $f > g$. If, on the other hand, $f(t) > g(t)$ for all $t \in T$, then we write $f \gg g$. The expressions $f \leq g$, $f < g$ and $f \ll g$ are understood similarly. Note that $\geq$ is a partial order on $\mathbb{R}^T$ which is linear iff $|T| = 1$.

We define the addition and multiplication of real functions by using the binary operations $+$ and $\cdot$ *pointwise*. That is, for any $f, g \in \mathbb{R}^T$, we define $f + g$ and $fg \in \mathbb{R}^T$ as the real functions on T with

$$(f + g)(t) := f(t) + g(t) \qquad \text{and} \qquad (fg)(t) := f(t)g(t) \qquad \text{for all } t \in T.$$

Similarly, for any $a \in \mathbb{R}$, the map $af \in \mathbb{R}^T$ is defined by $(af)(t) := af(t)$. The subtraction operation is then defined on $\mathbb{R}^T$ in the straightforward way: $f - g := f + (-1)g$ for each $f, g \in \mathbb{R}^T$. Provided that $g(t) \neq 0$ for all $t \in T$, we also define $\frac{f}{g} \in \mathbb{R}^T$ by $\left(\frac{f}{g}\right)(t) := \frac{f(t)}{g(t)}$.

Remark 1. Let $n \in \mathbb{N}$. By setting $T := \{1, \ldots, n\}$, we see that the definitions above also tell us how we order, add and multiply vectors in $\mathbb{R}^n$. In particular,

$\geq$ is none other than the natural order of $\mathbb{R}^n$. Moreover, for any $\lambda \in \mathbb{R}$ and real n-vectors $x := (x_1, \ldots, x_n)$ and $y := (y_1, \ldots, y_n)$, we have

$$x + y = (x_1 + y_1, \ldots, x_n + y_n) \quad \text{and} \quad \lambda x = (\lambda x_1, \ldots, \lambda x_n).$$

Of course, these are the natural addition and scalar multiplication operations in $\mathbb{R}^n$; when we talk about $\mathbb{R}^n$ we always have these operation in mind. In particular, these operations are used to define the **line segment between** x **and** y algebraically as $\{\lambda x + (1 - \lambda)y : 0 \leq \lambda \leq 1\}$.

Similar remarks apply to real matrices and sequences. Indeed, given any positive integers m and n, by setting $T := \{1, \ldots, m\} \times \{1, \ldots, n\}$, we obtain the definitions for ordering, summing and multiplying by a real number the members of $\mathbb{R}^{m \times n}$. Similarly, by setting $T := \mathbb{N}$, we find out about the situation for $\mathbb{R}^\infty$. For instance, for any real number λ and any matrices $[a_{ij}]_{m \times n}$ and $[b_{ij}]_{m \times n}$, we have

$$[a_{ij}]_{m \times n} + [b_{ij}]_{m \times n} = [a_{ij} + b_{ij}]_{m \times n} \quad \text{and} \quad \lambda[a_{ij}]_{m \times n} = [\lambda a_{ij}]_{m \times n}.$$

Similarly, for any $\lambda \in \mathbb{R}$ and any $(x_m), (y_m) \in \mathbb{R}^\infty$, we have $(x_m) + (y_m) = (x_m + y_m)$ and $\lambda(x_m) = (\lambda x_m)$, while $(x_m) \geq (0, 0, \ldots)$ means that $x_m \geq 0$ for each m. $\square$

When $|T| \geq 2$, $(\mathbb{R}^T, +, \cdot)$ is not a field, because not every map in $\mathbb{R}^T$ has a multiplicative inverse. (What is the inverse of the map that equals 0 at a given point and 1 elsewhere, for instance?) Nevertheless, $\mathbb{R}^T$ has a pretty rich algebraic structure. In particular, it is a partially ordered linear space (see Chapter F).

When the domain of a real function is a poset, we can talk about how this map affects the ordering of things in its domain. Of particular interest in this regard is the concept of a *monotonic* function defined on a subset of $\mathbb{R}^n$, $n \in \mathbb{N}$. (Of course, we think of $\mathbb{R}^n$ as a poset with respect to its natural order (Example 2.[3]).) For any $\emptyset \neq T \subseteq \mathbb{R}^n$, the map $f \in \mathbb{R}^T$ is said to be **increasing** if, for any $x, y \in T$, $x \geq y$ implies $f(x) \geq f(y)$, and **strictly increasing** if, for any $x, y \in T$, $x > y$ implies $f(x) > f(y)$. (An obvious example of an increasing real function that is not strictly increasing is a constant function on $\mathbb{R}$.) We say that $f \in \mathbb{R}^T$ is **decreasing** or **strictly decreasing** if $-f$ is increasing or strictly increasing, respectively. By a **monotonic** function in $\mathbb{R}^T$, we understand a map in $\mathbb{R}^T$ that is either increasing or decreasing.

We say that $f \in \mathbb{R}^T$ is **bounded** if there is an $M \in \mathbb{R}$ such that $\left|f(t)\right| \leq M$ for all $t \in T$. Note that, given any $-\infty < a \leq b < \infty$, every monotonic function in $\mathbb{R}^{[a,b]}$ is bounded. Indeed, for any such function, we have $\left|f(t)\right| \leq \max\{\left|f(a)\right|, \left|f(b)\right|\}$ for all $a \leq t \leq b$. It is also easily seen that a strictly increasing function f in $\mathbb{R}^{[a,b]}$ is injective. Thus $f : [a,b] \to f([a,b])$ is then a bijection, and hence it is invertible (Proposition 2). Moreover, the inverse of f is itself a strictly increasing function on $f([a,b])$. Similar observations hold for strictly decreasing functions.

4.2 Limits, Continuity, and Differentiation

Let T be a nonempty subset of $\mathbb{R}$, and $f \in \mathbb{R}^T$. If x is an extended real number that is the limit of at least one decreasing sequence in $T\backslash\{x\}$, then we say that $y \in \overline{\mathbb{R}}$ is the **right-limit** of f at x, and write $f(x+) = y$, provided that $f(x_m) \to y$ for every sequence (x_m) in $T\backslash\{x\}$ with $x_m \searrow x$. (Notice that f does not have to be defined at x.) The **left-limit** of f at x, denoted as $f(x-)$, is defined analogously. Finally, if x is an extended real number that is the limit of at least one sequence in $T\backslash\{x\}$, we say that y is the **limit** of f at x, and write

$$\lim_{t\to x} f(t) = y,$$

provided that $f(x_m) \to y$ for every sequence (x_m) in $T\backslash\{x\}$ with $x_m \to x$.[39] Equivalently, for any such x, we have $\lim_{t\to x} f(t) = y$ iff, for each $\varepsilon > 0$, we can find a $\delta > 0$ (which may depend on x and ε) such that $\left|y - f(t)\right| < \varepsilon$ for all $t \in T\backslash\{x\}$ with $|x - t| < \delta$. (Why?) In particular, when T is an open interval and $x \in T$, we have $\lim_{t\to x} f(t) = y$ iff $f(x+) = y = f(x-)$.

Let $x \in T$. If there is no sequence (x_m) in $T\backslash\{x\}$ with $x_m \to x$ (so x is an *isolated* point of T), or if there is such a sequence and $\lim_{t\to x} f(t) = f(x)$, we say that f is **continuous at** x. Intuitively, this means that f maps the points nearby x to points that are close to $f(x)$. For any nonempty subset S of T, if f is continuous at each $x \in S$, then it is said to be **continuous on** S. If $S = T$ here, then we simply say that f is **continuous**. The set of all continuous functions on T is denoted by $\mathbf{C}(T)$. (But if $T := [a,b]$ for

39 Warning. The limit of a function may fail to exist at every point on its domain. (Check the limits of the indicator function of $\mathbb{Q}$ in $\mathbb{R}$, for instance.)

some $a, b \in \mathbb{R}$ with $a \leq b$, then we write $\mathbf{C}[a, b]$ instead of $\mathbf{C}([a, b])$. It is obvious that if $f \in \mathbb{R}^T$ is continuous, then so is $f|_S$ for any $S \in 2^T\backslash\{\emptyset\}$. Put differently, continuity of a real function implies its continuity on any nonempty subset of its domain.

A function $f \in \mathbb{R}^T$ is said to be **uniformly continuous on** $S \subseteq T$ if for each $\varepsilon > 0$, there exists a $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for *all* $x, y \in S$ with $|x - y| < \delta$. If $S = T$ here, then we say that f is **uniformly continuous**. While continuity is a "local" phenomenon, uniform continuity is a "global" property that says that whenever *any* two points in the domain of the function are close to each other, so should the values of the function at these points.

It is obvious that if $f \in \mathbb{R}^T$ is uniformly continuous, then it is continuous. (Yes?) The converse is easily seen to be false. For instance, $f : (0, 1) \to \mathbb{R}$ defined by $f(t) := \frac{1}{t}$ is continuous, but not uniformly continuous. There is, however, one important case in which uniform continuity and continuity coincide.

Proposition 11

(Heine) *Let T be any subset of $\mathbb{R}$ that contains the closed interval $[a, b]$, and take any $f \in \mathbb{R}^T$. Then f is continuous on $[a, b]$ if, and only if, it is uniformly continuous on $[a, b]$.*

Proof

To derive a contradiction, assume that f is continuous on $[a, b]$, but not uniformly so. Then there exists an $\varepsilon > 0$ such that we can find two sequences (x_m) and (y_m) in $[a, b]$ with

$$|x_m - y_m| < \frac{1}{m} \quad \text{and} \quad |f(x_m) - f(y_m)| \geq \varepsilon, \qquad m = 1, 2, \ldots \tag{5}$$

(Why?) By the Bolzano-Weierstrass Theorem, there exists a convergent subsequence (x_{m_k}) of (x_m). Let $x := \lim x_{m_k}$, and note that $a \leq x \leq b$, so f is continuous at x. Then, since the first part of (5) guarantees that $\lim y_{m_k} = x$, we have $\lim f(x_{m_k}) = f(x) = \lim f(y_{m_k})$, which, of course, entails $|f(x_M) - f(y_M)| < \varepsilon$ for some $M \in \mathbb{N}$ large enough, contradicting the second part of (5). ■

Back to our review. Let x be a real number that is the limit of at least one sequence in $T\backslash\{x\}$. If the limits of $f, g \in \mathbb{R}^T$ at x exist and are finite, then we have

$$\lim_{t\to x}\big(f(t)+g(t)\big) = \lim_{t\to x} f(t) + \lim_{t\to x} g(t) \qquad \text{and}$$

$$\lim_{t\to x} f(t)g(t) = \lim_{t\to x} f(t) \lim_{t\to x} g(t). \tag{6}$$

(If $\lim_{t\to x} f(t) = \infty$, then these formulas remain valid provided that $\lim_{t\to x} g(t) \neq -\infty$ and $\lim_{t\to x} g(t) \neq 0$, respectively.) Moreover, we have

$$\lim_{t\to x} \frac{f(t)}{g(t)} = \frac{\lim_{t\to x} f(t)}{\lim_{t\to x} g(t)},$$

provided that $g(t) \neq 0$ for all $t \in T$ and $\lim_{t\to x} g(t) \neq 0$. (*Proof.* These assertions follow readily from those of Exercise 35.) It follows that $\mathbf{C}(T)$ is closed under addition and multiplication. More generally, if $f, g \in \mathbb{R}^T$ are continuous at $x \in T$, then $f + g$ and fg are continuous at x. (Of course, provided that it is well-defined on T, the same also goes for $\frac{f}{g}$.)

In passing, we note that, for any $m \in \mathbb{Z}_+$, a **polynomial of degree** m on T is a real function $f \in \mathbb{R}^T$ with

$$f(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_m t^m \qquad \text{for all } t \in T$$

for some $a_0, \ldots, a_m \in \mathbb{R}$ such that $a_m \neq 0$ if $m > 0$. The set of all polynomials (of any degree) on T is denoted as $\mathbf{P}(T)$, but again, if T is an interval of the form $[a, b]$, we write $\mathbf{P}[a, b]$ instead of $\mathbf{P}([a, b])$.

Clearly, $\mathbf{P}(T)$ is closed under addition and multiplication. Moreover, since any constant function on T, along with id_T, is continuous, and $\mathbf{C}(T)$ is closed under addition and multiplication, a straightforward application of the Principle of Mathematical Induction shows that $\mathbf{P}(T) \subseteq \mathbf{C}(T)$.

The following exercises aim to substantiate this brief review. We take up the theory of continuous functions in a much more general setting in Chapter D, where, you will be happy to know, the exposition will proceed under the speed limit.

Exercise 51 Let S and T be two nonempty subsets of $\mathbb{R}$, and take any $(f, g) \in \mathbb{R}^T \times \mathbb{R}^S$ with $f(T) \subseteq S$. Show that if f is continuous at $x \in T$ and g is continuous at $f(x)$, then $g \circ f$ is continuous at x.

Exercise 52 For any given $-\infty < a < b < \infty$, let $f \in \mathbb{R}^{(a,b)}$ be a continuous bijection. Show that f^{-1} is a continuous bijection defined on $f((a,b))$.

$\boxed{\text{Exercise 53}}^{\mathrm{H}}$ (*"Baby" Weierstrass' Theorem*) Show that for any $a, b \in \mathbb{R}$ with $a \leq b$, and any $f \in \mathbf{C}[a,b]$, there exist $x, y \in [a,b]$ such that $f(x) \geq f(t) \geq f(y)$ for all $t \in [a,b]$.

$\boxed{\text{Exercise 54}}^{\mathrm{H}}$ (*"Baby" Intermediate Value Theorem*) Let I be any interval and $a, b \in I$. Prove: If $f \in \mathbf{C}[a,b]$ and $f(a) < f(b)$, then $(f(a), f(b)) \subseteq f((a,b))$.

Let I be a nondegenerate interval, and take any $f \in \mathbb{R}^I$. For any given $x \in I$, we define the **difference-quotient map** $Q_{f,x} : I\backslash\{x\} \to \mathbb{R}$ by

$$Q_{f,x}(t) := \frac{f(t) - f(x)}{t - x}.$$

If the right-limit of this map at x exists as a real number, that is, $Q_{f,x}(x+) \in \mathbb{R}$, then f is said to be **right-differentiable at** x. In this case, the number $Q_{f,x}(x+)$ is called the **right-derivative** of f at x, and is denoted by $f'_+(x)$. Similarly, if $Q_{f,x}(x-) \in \mathbb{R}$, then f is said to be **left-differentiable at** x, and the **left-derivative** of f at x, denoted by $f'_-(x)$, is defined as the number $Q_{f,x}(x-)$. If x is the left end point of I and $f'_+(x)$ exists, or if x is the right end point of I and $f'_-(x)$ exists, or if x is not an end point of I and f is both right- and left-differentiable at x with $f'_+(x) = f'_-(x)$, then we say that f is **differentiable at** x. In the first case $f'_+(x)$, in the second case $f'_-(x)$, and in the third case the common value of $f'_+(x)$ and $f'_-(x)$ is denoted as either $f'(x)$ or $\frac{d}{dt}f(x)$. As you know, when it exists, the number $f'(x)$ is called the **derivative of** f **at** x. It is readily checked that f is differentiable at x iff

$$\lim_{t \to x} \frac{f(t) - f(x)}{t - x} \in \mathbb{R},$$

in which case $f'(x)$ equals precisely to this number. If J is an interval contained in I, and f is differentiable at each $x \in J$, then we say that f is **differentiable on** J. If $J = I$ here, then we simply say that f is **differentiable**. In this case the **derivative** of f is defined as the function $f' : I \to \mathbb{R}$ that maps each $x \in I$ to the derivative of f at x. (If f' is differentiable, then f is said to be **twice differentiable**, and the **second derivative** of f is defined as the function $f'' : I \to \mathbb{R}$ that maps each $x \in I$ to the derivative of f' at x.)

Similarly, if f is right-differentiable at each $x \in I$, then it is said to be **right-differentiable**, and in this case we define the **right-derivative of** f as a real function on I that maps every $x \in I$ to $f'_+(x)$. Naturally, this function is denoted as f'_+. Left-differentiability of f and the function f'_- are analogously defined.

The following exercises recall a few basic facts about the differentiation of real functions on the real line.

$\boxed{\text{Exercise 55}}$ Let I be an open interval and take any $f \in \mathbb{R}^I$.

(a) Show that if $f \in \mathbb{R}^I$ is differentiable then it is continuous.
(b) Show that if $f, g \in \mathbb{R}^I$ are differentiable and $\alpha \in \mathbb{R}$, then $\alpha f + g$ and fg are differentiable.
(c) Show that every $f \in \mathbf{P}(I)$ is differentiable.
(d) (*The Chain Rule*) Let $f \in \mathbb{R}^I$ be differentiable and $f(I)$ an open interval. Show that if $g \in \mathbb{R}^{f(I)}$ is differentiable, then so is $g \circ f$ and $(g \circ f)' = (g' \circ f)f'$.

For any $-\infty < a < b < \infty$ and $f \in \mathbf{C}[a, b]$, the definition above maintains that the derivatives of f at a and at b are $f'_+(a)$ and $f'_-(b)$, respectively. Thus, f being differentiable means that $f|_{[a,b)}$ is right-differentiable, $f|_{(a,b]}$ is left-differentiable, and $f'_+(x) = f'_-(x)$ for each $a < x < b$. If $f' \in \mathbf{C}[a, b]$, then we say that f is **continuously differentiable**—the class of all such real functions is denoted by $\mathbf{C}^1[a, b]$. If, further, $f' \in \mathbf{C}^1[a, b]$, then we say that f is **twice continuously differentiable**, and denote the class of all such maps by $\mathbf{C}^2[a, b]$. We define the classes $\mathbf{C}^3[a, b]$, $\mathbf{C}^4[a, b]$, etc., inductively. In turn, for any positive integer k, we let $\mathbf{C}^k[a, \infty)$ stand for the class of all $f \in \mathbf{C}[a, \infty)$ such that $f|_{[a,b]} \in \mathbf{C}^k[a, b]$ for every $b > a$. (The classes $\mathbf{C}^k(-\infty, b]$ and $\mathbf{C}^k(\mathbb{R})$ are defined analogously.)

Let f be differentiable on the bounded open interval (a, b). If f assumes its maximum at some $x \in (a, b)$, that is, $f(x) \geq f(t)$ for all $a < t < b$, then a fairly obvious argument shows that the derivative of f must vanish at x, that is, $f'(x) = 0$. (*Proof.* If $f'(x) > 0$ (or < 0), then we could find a small enough $\varepsilon > 0$ (< 0, respectively) such that $x + \varepsilon \in (a, b)$ and $f(x+\varepsilon) > f(x)$, contradicting that f assumes its maximum at x.) Of course, the same would be true if f assumed instead its minimum at x. (*Proof.* Just apply the previous observation to $-f$.) Combining these observations with the "baby" Weierstrass Theorem of Exercise 53 yields the following simple but very useful result.

Rolle's Theorem
Let $-\infty < a < b < \infty$ and $f \in \mathbf{C}[a, b]$. If f is differentiable on (a, b) and $f(a) = f(b)$, then $f'(c) = 0$ for some $c \in (a, b)$.

Proof
Since f is continuous, the "baby" Weierstrass Theorem (Exercise 53) implies that there exist $a \leq x, y \leq b$ such that $f(y) \leq f(t) \leq f(x)$ for all $a \leq t \leq b$. Now assume that f is differentiable on (a, b), and $f(a) = f(b)$. If $\{x, y\} \subseteq \{a, b\}$, then f must be a constant function, and hence $f'(t) = 0$ for all $a \leq t \leq b$. If this is not the case, then either $x \in (a, b)$ or $y \in (a, b)$. In the former case we have $f'(x) = 0$ (because f assumes its maximum at x), and in the latter case $f'(y) = 0$. ■

There are many important consequences of this result. The following exercise recounts some of them.

Exercise 56[H] Let $-\infty < a < b < \infty$, and take any $f \in \mathbf{C}[a, b]$ that is differentiable on (a, b).

(a) Prove the *Mean Value Theorem*: There exists a $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$.
(b) Show that if $f' = 0$, then f is a constant function.
(c) Show that if $f' \geq 0$, then f is increasing, and if $f' > 0$, then it is strictly increasing.

We shall revisit the theory of differentiation in Chapter K in a much broader context and use it to give a potent introduction to optimization theory.

4.3 Riemann Integration

Throughout this section we work mostly with two arbitrarily fixed real numbers a and b, with $a \leq b$. For any $m \in \mathbb{N}$, we denote by $[a_0, \ldots, a_m]$ the set

$$\{[a_0, a_1], [a_1, a_2], \ldots, [a_{m-1}, a_m]\} \quad \text{where} \quad a = a_0 < \cdots < a_m = b,$$

provided that $a < b$. In this case, we refer to $[a_0, \ldots, a_m]$ as a **dissection** of $[a, b]$, and we denote the class of all dissections of $[a, b]$ by $\mathcal{D}[a, b]$. By convention, we let $\mathcal{D}[a, b] := \{\{a\}\}$ when $a = b$.

For any $\mathbf{a} := [a_0, \ldots, a_m]$ and $\mathbf{b} := [b_0, \ldots, b_k]$ in $\mathcal{D}[a, b]$, we write $\mathbf{a} \Cup \mathbf{b}$ for the dissection $[c_0, \ldots, c_l] \in \mathcal{D}[a, b]$ where $\{c_0, \ldots, c_l\} = \{a_0, \ldots, a_m\} \cup \{b_0, \ldots, b_k\}$. Moreover, we say that $\mathbf{b}$ is **finer than a** if $\{a_0, \ldots, a_m\} \subseteq \{b_0, \ldots, b_k\}$. Evidently, $\mathbf{a} \Cup \mathbf{b} = \mathbf{b}$ iff $\mathbf{b}$ is finer than $\mathbf{a}$.

Now let $f \in \mathbb{R}^{[a,b]}$ be any bounded function. For any $\mathbf{a} := [a_0, \ldots, a_m] \in \mathcal{D}[a, b]$, we define

$$K_{f,\mathbf{a}}(i) := \sup\{f(t) : a_{i-1} \le t \le a_i\} \quad \text{and}$$
$$k_{f,\mathbf{a}}(i) := \inf\{f(t) : a_{i-1} \le t \le a_i\}$$

for each $i = 1, \ldots, m$. (Thanks to the Completeness Axiom, everything is well-defined here.) By the **a-upper Riemann sum** of f, we mean the number

$$R_{\mathbf{a}}(f) := \sum_{i=1}^{m} K_{f,\mathbf{a}}(i) \left(a_i - a_{i-1}\right),$$

and by the **a-lower Riemann sum** of f, we mean

$$r_{\mathbf{a}}(f) := \sum_{i=1}^{m} k_{f,\mathbf{a}}(i) \left(a_i - a_{i-1}\right).$$

Clearly, $R_{\mathbf{a}}(f)$ decreases, and $r_{\mathbf{a}}(f)$ increases, as $\mathbf{a}$ becomes finer, while we always have $R_{\mathbf{a}}(f) \ge r_{\mathbf{a}}(f)$. Moreover—and this is important—

$$R(f) := \inf\{R_{\mathbf{a}}(f) : \mathbf{a} \in \mathcal{D}[a, b]\} \ge \sup\{r_{\mathbf{a}}(f) : \mathbf{a} \in \mathcal{D}[a, b]\} =: r(f).$$

($R(f)$ and $r(f)$ are called the **upper** and **lower Riemann integrals** of f, respectively.) This is not entirely obvious. Make sure you prove it before proceeding any farther.[40]

[40] *Hint.* Otherwise we would have $R_{\mathbf{a}}(f) < r_{\mathbf{b}}(f)$ for some $\mathbf{a}, \mathbf{b} \in \mathcal{D}[a, b]$. Compare $R_{\mathbf{a}\Cup\mathbf{b}}(f)$ and $r_{\mathbf{a}\Cup\mathbf{b}}(f)$.

Definition

Let $f \in \mathbb{R}^{[a,b]}$ be a bounded function. If $R(f) = r(f)$, then f is said to be **Riemann integrable**, and the number

$$\int_a^b f(t)dt := R(f)$$

is called the **Riemann integral** of f.[41] In this case, we also define

$$\int_b^a f(t)dt := -R(f).$$

Finally, if $g \in \mathbb{R}^{[a,\infty)}$ is a bounded function, then we define the **improper Riemann integral** of g as

$$\int_a^\infty g(t)dt := \lim_{b\to\infty} R(g|_{[a,b]}),$$

provided that $g|_{[a,b]}$ is Riemann integrable for each $b > a$, and the limit on the right-hand side exists (in $\overline{\mathbb{R}}$). (For any bounded $g \in \mathbb{R}^{(-\infty,a]}$, the improper Riemann integral $\int_{-\infty}^a g(t)dt$ is analogously defined.)

As you surely recall, the geometric motivation behind this formulation relates to the calculation of the area under the graph of f on the interval $[a, b]$. (When $f \geq 0$, the intuition becomes clearer.) Informally put, we approximate the area that we wish to compute from above (by an upper Riemann sum) and from below (by a lower one), and by choosing finer and finer dissections, we check if these two approximations converge to the same real number. If they do, then we call the common limit the Riemann integral of f. If they don't, then $R(f) > r(f)$, and we say that f is *not* Riemann integrable.

Almost immediate from the definitions is the following simple but very useful result.

Proposition 12

If $f \in \mathbb{R}^{[a,b]}$ is bounded and Riemann integrable, then

$$\left|\int_a^b f(t)dt\right| \leq (b-a)\sup\{|f(t)| : a \leq t \leq b\}.$$

[41] Of course, t acts as a "dummy variable" here—the expressions $\int_a^b f(t)dt$, $\int_a^b f(x)dx$ and $\int_a^b f(\omega)d\omega$ all denote the same number. (For this reason, some authors prefer to write $\int_a^b f$ for the Riemann integral of f.)

EXERCISE 57 [H] Let $\alpha \in \mathbb{R}$ and let $f, g \in \mathbb{R}^{[a,b]}$ be bounded functions. Show that if f and g are Riemann integrable, then so is $\alpha f + g$, and we have

$$\int_a^b (\alpha f + g)(t)dt = \alpha \int_a^b f(t)dt + \int_a^b g(t)dt.$$

EXERCISE 58 Take any $c \in [a,b]$ and let $f \in \mathbb{R}^{[a,b]}$ be a bounded function. Show that if f is Riemann integrable, then so is $f|_{[a,c]}$ and $f|_{[c,b]}$, and we have

$$\int_a^b f(t)dt = \int_a^c f(t)dt + \int_c^b f(t)dt.$$

(Here $\int_a^c f(t)dt$ stands for $\int_a^c f|_{[a,c]}(t)dt$, and similarly for $\int_c^b f(t)dt$.)

EXERCISE 59 Prove Proposition 12.

EXERCISE 60 Let $f \in \mathbb{R}^{[a,b]}$ be a bounded function, and define $f^+, f^- \in \mathbb{R}^{[a,b]}$ by

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

(a) Verify that $f = f^+ - f^-$ and $|f| = f^+ + f^-$.
(b) Verify that $R_{\mathbf{a}}(f) - r_{\mathbf{a}}(f) \geq R_{\mathbf{a}}(f^+) - r_{\mathbf{a}}(f^+) \geq 0$ for any $\mathbf{a} \in \mathcal{D}[a,b]$, and state and prove a similar result for f^-.
(c) Show that if f is Riemann integrable, then so are f^+ and f^-.
(d) Show that if f is Riemann integrable, then so is $|f|$, and

$$\left| \int_a^b f(t)dt \right| \leq \int_a^b |f(t)| \, dt.$$

(Here, as usual, we write $|f(t)|$ for $|f|(t)$.)

An important issue in the theory of integration concerns the identification of Riemann integrable functions. Fortunately, we don't have to spend much time on this matter. The main integrability result that we need in the sequel is quite elementary.

PROPOSITION 13
Any $f \in \mathbf{C}[a,b]$ is Riemann integrable.

PROOF
We assume $a < b$, for otherwise the claim is obvious. Take any $f \in \mathbf{C}[a,b]$, and fix an arbitrary $\varepsilon > 0$. By Proposition 11, f is uniformly continuous on

$[a,b]$. Thus, there exists a $\delta > 0$ such that $\left|f(t) - f(t')\right| < \frac{\varepsilon}{b-a}$ for *all* t, t' in $[a,b]$ with $\left|t - t'\right| < \delta$. Then, for any dissection $\mathbf{a} := [a_0, \ldots, a_m]$ of $[a,b]$ with $\left|a_i - a_{i-1}\right| < \delta$ for each $i = 1, \ldots, m$, we have

$$R_{\mathbf{a}}(f) - r_{\mathbf{a}}(f) = \sum_{i=1}^{m} (K_{f,\mathbf{a}}(i) - k_{f,\mathbf{a}}(i))(a_i - a_{i-1}) < \frac{\varepsilon}{b-a} \sum_{i=1}^{m} (a_i - a_{i-1}) = \varepsilon.$$

Since $R_{\mathbf{a}}(f) \geq R(f) \geq r(f) \geq r_{\mathbf{a}}(f)$, it follows that $\left|R(f) - r(f)\right| < \varepsilon$. Since $\varepsilon > 0$ is arbitrary here, we are done. ■

Exercise 61[H] If $f \in \mathbf{C}[a,b]$ and $f \geq 0$, then $\int_a^b f(t)dt = 0$ implies $f = 0$.

Exercise 62 Let f be a bounded real map on $[a,b]$ which is continuous at all but finitely many points of $[a,b]$. Prove that f is Riemann integrable.

We conclude with a (slightly simplified) statement of the Fundamental Theorem of Calculus, which you should carry with yourself at all times. As you might recall, this result makes precise in what way one can think of the "differentiation" and "integration" as inverse operations. Its importance cannot be overemphasized.

The Fundamental Theorem of Calculus

For any $f \in \mathbf{C}[a,b]$ and $F \in \mathbb{R}^{[a,b]}$, we have

$$F(x) = F(a) + \int_a^x f(t)dt \qquad \text{for all } a \leq x \leq b, \tag{7}$$

if, and only if, $F \in \mathbf{C}^1[a,b]$ and $F' = f$.

Proof

Take any $f \in \mathbf{C}[a,b]$ and $F \in \mathbb{R}^{[a,b]}$ such that (7) holds. Consider any $a \leq x < b$, and let ε be a fixed but arbitrary positive number. Since f is continuous at x, there exists a $\delta > 0$ such that $\left|f(t) - f(x)\right| < \varepsilon$ for any $a < t < b$ with $|t - x| < \delta$. Thus, for any $x < y < b$ with $y - x < \delta$, we have

$$\left|\frac{F(y) - F(x)}{y - x} - f(x)\right| \leq \frac{1}{y - x} \int_x^y \left|f(t) - f(x)\right| dt \leq \varepsilon$$

by Exercise 58 and Proposition 12. It follows that $F|_{[a,b)}$ is right-differentiable and $F'_+(x) = f(x)$ for each $a \leq x < b$. Moreover, an analogous argument would show that $F|_{(a,b]}$ is left-differentiable and $F'_-(x) = f(x)$ for each $a < x \leq b$. Conclusion: F is differentiable and $F' = f$.

Conversely, take any $f \in \mathbf{C}[a,b]$ and $F \in \mathbf{C}^1[a,b]$ such that $F' = f$. We wish to show that (7) holds. Fix any $a \leq x \leq b$, and let $\varepsilon > 0$. It is easy to see that, since f is Riemann integrable (Proposition 13), there exists a dissection $\mathbf{a} := [a_0, \ldots, a_m]$ in $\mathcal{D}[a,x]$ such that $R_{\mathbf{a}}(f) - r_{\mathbf{a}}(f) < \varepsilon$. (Yes?) By the Mean Value Theorem (Exercise 56), for each $i = 1, \ldots, m$ there exists an $x_i \in (a_{i-1}, a_i)$ with $F(a_i) - F(a_{i-1}) = f(x_i)(a_i - a_{i-1})$. It follows that

$$F(x) - F(a) = \sum_{i=1}^{m}(F(a_i) - F(a_{i-1})) = \sum_{i=1}^{m} f(x_i)(a_i - a_{i-1}),$$

and hence $R_{\mathbf{a}}(f) \geq F(x) - F(a) \geq r_{\mathbf{a}}(f)$. Since $R_{\mathbf{a}}(f) - r_{\mathbf{a}}(f) < \varepsilon$ and $R_{\mathbf{a}}(f) \geq \int_a^x f(t)dt \geq r_{\mathbf{a}}(f)$, therefore,

$$\left| \int_a^x f(t)dt - (F(x) - F(a)) \right| < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary here, the theorem is proved. ■

REMARK 2. In the statement of the Fundamental Theorem of Calculus, we may replace (7) by

$$F(x) = F(b) - \int_x^b f(t)dt \qquad \text{for all } a \leq x \leq b.$$

The proof goes through (almost) verbatim. □

EXERCISE 63 (*Integration by Parts Formula*) Prove: If $f, g \in \mathbf{C}^1[a,b]$, then

$$\int_a^b f(t)g'(t)dt = f(b)g(b) - f(a)g(a) - \int_a^b f'(t)g(t)dt.$$

4.4 Exponential, Logarithmic, and Trigonometric Functions

Other than the polynomials, we use only four types of special real functions in this book: the exponential, the logarithmic, and the two most basic trigonometric functions. The rigorous development of these functions from scratch is a tedious task that we do not wish to get into here. Instead, by using integral calculus, we introduce these functions here at a far quicker pace.

Let us begin with the **logarithmic function**: We define the map $x \mapsto \ln x$ on $\mathbb{R}_{++}$ by

$$\ln x := \int_1^x \tfrac{1}{t}dt.$$

This map is easily checked to be strictly increasing and continuous, and of course, $\ln 1 = 0$. (Verify!) By the Fundamental Theorem of Calculus (and Remark 2), the logarithmic function is differentiable, and we have $\frac{d}{dx} \ln x = \frac{1}{x}$ for any $x > 0$. Two other important properties of this function are:

$$\ln xy = \ln x + \ln y \qquad \text{and} \qquad \ln \tfrac{x}{y} = \ln x - \ln y \tag{8}$$

for any $x, y > 0$. To prove the first assertion, fix any $y > 0$, and define $f : \mathbb{R}_{++} \to \mathbb{R}$ by $f(x) := \ln xy - \ln x - \ln y$. Observe that f is differentiable, and $f'(x) = 0$ for all $x > 0$ by the Chain Rule. Since $f(1) = 0$, it follows that $f(x) = 0$ for all $x > 0$. (Verify this by using Exercise 56.) To prove the second claim in (8), on the other hand, set $x = \frac{1}{y}$ in the first equation of (8) to find $\ln \frac{1}{y} = -\ln y$ for any $y > 0$. Using this fact and the first equation of (8) again, we obtain $\ln \frac{x}{y} = \ln x + \ln \frac{1}{y} = \ln x - \ln y$ for any $x, y > 0$.

Finally, we note that

$$\lim_{x \to 0} \ln x = -\infty \qquad \text{and} \qquad \lim_{x \to \infty} \ln x = \infty. \tag{9}$$

(See Figure 3.) Let us prove the second assertion here, the proof of the first claim being analogous. Take any $(x_m) \in \mathbb{R}^{\infty}_{++}$ with $x_m \to \infty$. Clearly, there exists a strictly increasing sequence (m_k) of natural numbers such that $x_{m_k} \geq 2^k$ for all $k = 1, 2, \ldots$. Since $x \mapsto \ln x$ is increasing, we thus have $\ln x_{m_k} \geq \ln 2^k = k \ln 2$. (The final equality follows from the Principle of Mathematical Induction and the first equation in (8).) Since $\ln 2 > 0$, it is obvious that $k \ln 2 \to \infty$ as $k \to \infty$. It follows that the strictly increasing sequence $(\ln x_m)$ has a subsequence that diverges to ∞, which is possible only if $\ln x_m \to \infty$. (Why?)

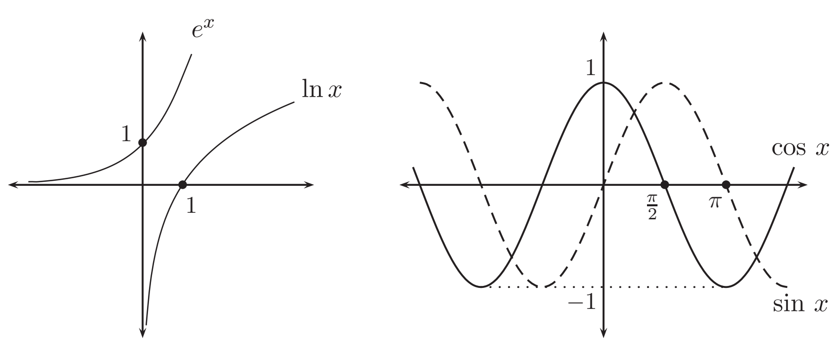


Figure 3

Since the logarithmic function is continuous, (9) and an appeal to the "baby" Intermediate Value Theorem (Exercise 54) entail that the range of this map is the entire $\mathbb{R}$. Since it is also strictly increasing, the logarithmic function is invertible, with its inverse function being a strictly increasing map from $\mathbb{R}$ onto $\mathbb{R}_{++}$. The latter map, denoted as $x \mapsto e^x$, is called the **exponential function**. (See Figure 3.) By definition, we have

$$\ln e^x = x \quad \text{and} \quad e^{\ln x} = x \quad \text{for all } x > 0.$$

Of course, the real number e^1 is denoted as e.[42]

The following property of the exponential function is basic:

$$e^{x+y} = e^x e^y \quad \text{for all } x, y > 0. \tag{10}$$

Indeed, by (8),

$$\ln e^x e^y = \ln e^x + \ln e^y = x + y = \ln e^{x+y}$$

for all $x, y > 0$, so, since the logarithmic function is injective, we get (10).[43]

Finally, let us show that the exponential function is differentiable, and compute its derivative. Since the derivative of the logarithmic function at 1 equals 1, we have $\lim_{y \to 1} \frac{\ln y}{y-1} = 1$, which implies that $\lim_{y \to 1} \frac{y-1}{\ln y} = 1$. (Why?) Then, since $e^{x_m} \to 1$ for any real sequence (x_m) with $x_m \to 0$, we have $\frac{e^{x_m}-1}{x_m} \to 1$. (Why?) Since (x_m) is arbitrary here, we thus have $\lim_{\varepsilon \to 0} \frac{e^\varepsilon - 1}{\varepsilon} = 1$. It follows that $x \mapsto e^x$ is differentiable at 0, and its derivative equals 1 there. Therefore, by (10),

$$\frac{d}{dx} e^x = \lim_{\varepsilon \to 0} \frac{e^{x+\varepsilon} - e^x}{\varepsilon} = e^x \lim_{\varepsilon \to 0} \frac{e^\varepsilon - 1}{\varepsilon} = e^x, \qquad -\infty < x < \infty.$$

We conclude that the exponential map is differentiable, and the derivative of this function is equal to the exponential function itself.

Among the trigonometric functions, we only need to introduce the *sine* and the *cosine* functions, and we will do this again by using integral calculus. Let us define first the real number π by the equation

$$\pi := 2 \int_0^1 \frac{1}{\sqrt{1-t^2}} dt,$$

[42] Therefore, e is the (unique) real number with the property $\int_1^e \frac{1}{t} dt = 1$, but of course, there are various other ways of defining the number e (Exercise 47).

[43] By the way, do you think there is another increasing map f on $\mathbb{R}$ with $f(1) = e$ and $f(x+y) = f(x)f(y)$ for any $x, y \in \mathbb{R}$? (This question will be answered in Chapter D.)

that is, we define π as the area of the circle with radius 1. Now define the function $f \in \mathbb{R}^{(-1,1)}$ by

$$f(x) := \int_0^x \frac{1}{\sqrt{1-t^2}}dt \qquad \text{for any } x \geq 0$$

and

$$f(x) := -\int_x^0 \frac{1}{\sqrt{1-t^2}}dt \qquad \text{for any } x < 0.$$

This map is a bijection from $(-1,1)$ onto $(-\frac{\pi}{2}, \frac{\pi}{2})$. (Why?) We define the map $x \mapsto \sin x$ on $(-\frac{\pi}{2}, \frac{\pi}{2})$ as the inverse of f, and then extend it to $[-\frac{\pi}{2}, \frac{\pi}{2}]$ by setting $\sin \frac{-\pi}{2} := -1$ and $\sin \frac{\pi}{2} = 1$. (How does this definition relate to the geometry behind the sine function?) Finally, the **sine function** is defined on the entire $\mathbb{R}$ by requiring the following periodicity: $\sin(x+\pi) = -\sin x$ for all $x \in \mathbb{R}$. It is easy to see that this function is an *odd* function, that is, $\sin(-x) = -\sin x$ for any $x \in \mathbb{R}$ (Figure 3).

Now define the map $x \mapsto \cos x$ on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ by $\cos x := \sqrt{1-(\sin x)^2}$, and then extend it to $\mathbb{R}$ by requiring the same periodicity with the sine function: $\cos(x+\pi) = -\cos x$ for any $x \in \mathbb{R}$. The resulting map is called the **cosine function.** This is an *even* function, that is, $\cos(-x) = \cos x$ for any $x \in \mathbb{R}$, and we have $\cos 0 = 1$ and $\cos \frac{\pi}{2} = 0 = \cos \frac{-\pi}{2}$ (Figure 3).

Exercise 64 Show that the sine and cosine functions are differentiable, and $\frac{d}{dx} \sin x = \cos x$ and $\frac{d}{dx} \cos x = -\sin x$ for all $x \in \mathbb{R}$.

Exercise 65 Prove: $\lim_{x \to 0} \frac{\sin x}{x} = 1$.

4.5 Concave and Convex Functions

Let $n \in \mathbb{N}$, and recall that a subset T of $\mathbb{R}^n$ is said to be **convex** if the line segment connecting any two elements of T lies entirely within T, that is, $\lambda x + (1-\lambda)y \in T$ for all $x, y \in T$ and $0 \leq \lambda \leq 1$. Given any such nonempty set T, a function $\varphi \in \mathbb{R}^T$ is called **concave** if

$$\varphi(\lambda x + (1-\lambda)y) \geq \lambda\varphi(x) + (1-\lambda)\varphi(y) \qquad \text{for any } x, y \in T \text{ and } 0 \leq \lambda \leq 1,$$

and **strictly concave** if this inequality holds strictly for any distinct $x, y \in T$ and $0 < \lambda < 1$. The definitions of convex and strictly convex functions are obtained by reversing these inequalities. Equivalently, φ is called (**strictly**) **convex** if $-\varphi$ is (strictly) concave. (This observation allows us to convert any property that a concave function may possess into a property for convex functions in a straightforward manner.) Finally, φ is said to be **affine** if it is both concave and convex.

If φ and ψ are concave functions in $\mathbb{R}^T$, and $\alpha \geq 0$, then $\alpha\varphi + \psi$ is a concave function in $\mathbb{R}^T$. Similarly, if S is an interval with $\varphi(T) \subseteq S$, and $\varphi \in \mathbb{R}^T$ and $\psi \in \mathbb{R}^S$ are concave, then so is $\psi \circ \varphi$. The following exercises provide two further examples of functional operations that preserve the concavity of real functions.

Exercise 66 For any given $n \in \mathbb{N}$, let T be a nonempty convex subset of $\mathbb{R}^n$ and $\mathcal{F}$ a (nonempty) class of concave functions in $\mathbb{R}^T$. Show that if $\inf\{\varphi(x) : \varphi \in \mathcal{F}\} > -\infty$ for all $x \in T$, then the map $x \mapsto \inf\{\varphi(x) : \varphi \in \mathcal{F}\}$ is a concave function in $\mathbb{R}^T$.

Exercise 67 For any given $n \in \mathbb{N}$, let T be a nonempty convex subset of $\mathbb{R}^n$ and (φ_m) a sequence of concave functions in $\mathbb{R}^T$. Show that if $\lim \varphi_m(x) \in \mathbb{R}$ for each $x \in T$, then the map $x \mapsto \lim \varphi_m(x)$ is a concave function in $\mathbb{R}^T$.

We now specialize to concave functions defined on an open interval $I \subseteq \mathbb{R}$. The first thing to note about such a function is that it is continuous. In fact, we can prove a stronger result with the aid of the following useful observation about the boundedness of concave functions defined on a bounded interval.

Lemma 2

For any given $-\infty < a \leq b < \infty$, if $f \in \mathbb{R}^{[a,b]}$ is concave (or convex), then

$$\inf f([a,b]) > -\infty \quad \textit{and} \quad \sup f([a,b]) < \infty.$$

Proof

Let f be a concave real map on $[a,b]$. Obviously, for any $a \leq t \leq b$, we have $t = \lambda_t a + (1-\lambda_t)b$ for some $0 \leq \lambda_t \leq 1$, whereas $f(\lambda_t a + (1-\lambda_t)b) \geq \min\{f(a), f(b)\}$ by concavity. It follows that $\inf f([a,b]) > -\infty$.

The proof of the second claim is trickier. Let us denote the midpoint $\frac{a+b}{2}$ of the interval $[a,b]$ by M, and fix an arbitrary $a \leq t \leq b$. Note that there is a real number c_t such that $|c_t| \le \frac{b-a}{2}$ and $t = M + c_t$. (Simply define c_t by the latter equation.) Then, $M - c_t$ belongs to $[a, b]$, so, by concavity,

$$\begin{aligned} f(M) &= f\left(\tfrac{1}{2}(M + c_t) + \tfrac{1}{2}(M - c_t)\right) \\ &\ge \tfrac{1}{2} f(M + c_t) + \tfrac{1}{2} f(M - c_t) \\ &= \tfrac{1}{2} f(t) + \tfrac{1}{2} f(M - c_t), \end{aligned}$$

so $f(t) \le 2f(M) - \inf f([a, b]) < \infty$. Since t was chosen arbitrarily in $[a, b]$, this proves that $\sup f([a, b]) < \infty$. ■

Here is the main conclusion we wish to derive from this observation.

Proposition 14
Let I be an open interval and $f \in \mathbb{R}^I$. If f is concave (or convex), then for every $a, b \in \mathbb{R}$ with $a \le b$ and $[a, b] \subset I$, there exists a $K > 0$ such that

$$\left|f(x) - f(y)\right| \le K\left|x - y\right| \qquad \textit{for all } a \le x, y \le b.$$

Proof
Since I is open, there exists an $\varepsilon > 0$ such that $[a-\varepsilon, b+\varepsilon] \subseteq I$. Let $a' := a-\varepsilon$ and $b' := b + \varepsilon$. Assume that f is concave, and let $\alpha := \inf f([a', b'])$ and $\beta := \sup f([a', b'])$. By Lemma 2, α and β are real numbers. Moreover, if $\alpha = \beta$, then f is constant, so all becomes trivial. We thus assume that $\beta > \alpha$.

For any distinct $x, y \in [a, b]$, let

$$z := y + \varepsilon\frac{y - x}{\left|y - x\right|} \qquad \text{and} \qquad \lambda := \frac{\left|y - x\right|}{\varepsilon + \left|y - x\right|}.$$

Then $a' \le z \le b'$ and $y = \lambda z + (1 - \lambda)x$—we defined z the way we did in order to satisfy these two properties. Hence, by concavity of f, $f(y) \ge \lambda(f(z) - f(x)) + f(x)$, that is,

$$\begin{aligned} f(x) - f(y) &\le \lambda(f(x) - f(z)) \\ &\le \lambda(\beta - \alpha) \\ &= \frac{\beta - \alpha}{\varepsilon + \left|y - x\right|}\left|y - x\right| \\ &< \frac{\beta - \alpha}{\varepsilon}\left|y - x\right|. \end{aligned}$$

Interchanging the roles of x and y in this argument and letting $K := \frac{\beta-\alpha}{\varepsilon}$ complete the proof. ■

> Corollary 2
> *Let I be an open interval and $f \in \mathbb{R}^I$. If f is concave (or convex), then it is continuous.*

Exercise 68[H] Show that every concave function on an open interval is both right-differentiable and left-differentiable. (Of course, such a map need not be differentiable.)

In passing, we recall that, provided that f is a differentiable real map on an open interval I, then it is (strictly) concave iff f' is (strictly) decreasing. (Thus $x \mapsto \ln x$ is a concave map on $\mathbb{R}_{++}$, and $x \mapsto e^x$ is a convex map on $\mathbb{R}$.) Provided that f is twice differentiable, it is concave iff $f'' \leq 0$, while $f'' < 0$ implies the strict concavity of f. (The converse of the latter statement is false; for instance, the derivative of the strictly concave function $x \mapsto x^2$ vanishes at 0.) These are elementary properties, and they can easily be proved by using the Mean Value Theorem (Exercise 56). We will not, however, lose more time on this matter here.

This is all we need in terms of concave and convex real functions on the real line. In later chapters we will revisit the notion of concavity in much broader contexts. For now, we conclude by noting that a great reference that specializes on the theory of concave and convex functions is Roberts and Varberg (1973). That book certainly deserves a nice spot on the bookshelves of any economic theorist.

4.6 Quasiconcave and Quasiconvex Functions

With T being a nonempty convex subset of $\mathbb{R}^n$, $n \in \mathbb{N}$, we say that a function $\varphi \in \mathbb{R}^T$ is **quasiconcave** if

$$\varphi(\lambda x + (1-\lambda)y) \geq \min\{\varphi(x), \varphi(y)\} \qquad \text{for any } x, y \in T \text{ and } 0 \leq \lambda \leq 1,$$

and **strictly quasiconcave** if this inequality holds strictly for any distinct $x, y \in T$ and $0 < \lambda < 1$. (φ is called (**strictly**) **quasiconvex** if $-\varphi$ is (strictly) quasiconcave.) It is easy to show that φ is quasiconcave iff $\varphi^{-1}([a, \infty))$ is a convex set for any $a \in \mathbb{R}$. It is also plain that every concave function in $\mathbb{R}^T$ is quasiconcave, but not conversely.[44]

Quasiconcavity plays an important role in optimization theory, and it is often invoked to establish the uniqueness of a solution for a maximization problem. Indeed, if $f \in \mathbb{R}^T$ is strictly quasiconcave, and there exists an $x \in T$ with $f(x) = \max f(T)$, then x must be the only element of T with this property. For, if $x \neq y = \max f(T)$, then $f(x) = f(y)$, so $f(\frac{x}{2} + \frac{y}{2}) > f(x) = \max f(T)$ by strict quasiconcavity. Since $x, y \in T$ and T is convex, this is impossible.

Exercise 69 Give an example of two quasiconcave functions on the real line the sum of which is not quasiconcave.

Exercise 70 Let I be an interval, $f \in \mathbb{R}^I$, and let $g \in \mathbb{R}^{f(I)}$ be a strictly increasing function. Show that if f is quasiconcave, then so is $g \circ f$. Would $g \circ f$ be necessarily concave if f was concave?

[44] If $\emptyset \neq T \subseteq \mathbb{R}$, then every monotonic function in $\mathbb{R}^T$ is quasiconcave, but of course, not every monotonic function in $\mathbb{R}^T$ is concave.