Notes on Real Analysis

Lee Larson
Section 1: Sets and Functions

1 Sets and Functions

1.1 Basic Ideas

Set theory is a large and complicated subject in its own right. There is no time in this course to touch any but the simplest parts of it. Instead, we’ll just look at a few topics from what is often called “naive set theory.”

We begin with a few definitions.

A set is a collection of objects called elements. Usually, sets are denoted by the capital letters \( A, B, \ldots, Z \). A set can consist of any type of elements. Even other sets can be elements of a set. The sets we typically deal with here have real numbers as their elements.

If \( a \) is an element of the set \( A \), we write \( a \in A \). If \( a \) is not an element of the set \( A \), we write \( a \notin A \).

If all the elements of \( A \) are also elements of \( B \), then \( A \) is a subset of \( B \). In this case, we write \( A \subset B \) or \( B \supset A \).

Two sets \( A \) and \( B \) are equal, if they have the same elements. In this case we write \( A = B \). It is easy to see that \( A = B \) iff \( A \subset B \) and \( B \subset A \). Establishing that both of these containments are true is a standard way to show that two sets are equal.

There are several ways to describe a set.

A set can be described in words such as “\( P \) is the set of all presidents of the United States.” This is cumbersome for complicated sets.

All the elements of the set could be listed in curly braces as \( S = \{2, 0, a\} \). If the set is large, this is impractical, or impossible.

More common in mathematics is set builder notation. Some examples are

\[
P = \{ p : p \text{ is a president of the United states} \} \\
= \{ \text{Washington, Adams, Jefferson, \ldots, Clinton} \}
\]

and

\[
S = \{ n : n \text{ is a prime number} \} = \{2, 3, 5, 7, 11, \ldots \}.
\]

In general, the set builder notation defines a set in the form

\[
\{ \text{formula for a typical element : object to plug into the formula} \}.
\]

A more complicated example is the set of perfect squares:

\[
S = \{ n^2 : n \text{ is an integer} \} = \{0, 1, 4, 9, \ldots \}.
\]

The existence of several sets will be assumed. The simplest of these is the empty set, which is the set with no elements. It is denoted as \( \emptyset \). The natural numbers is the set \( \mathbb{N} = \{1, 2, 3, \ldots \} \) consisting of the positive integers. The set \( \mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots \} \) is the set of all integers. Clearly, \( \emptyset \subset A \), for any set \( A \) and

\[
\emptyset \subset \mathbb{N} \subset \mathbb{Z}.
\]
Section 1: Sets and Functions

Figure 1: These are Venn diagrams showing the four standard binary operations on sets. In this figure, the set which results from the operation is shaded.

Definition 1.1. Given any set $A$, the power set of $A$, written $\mathcal{P}(A)$, is the set consisting of all subsets of $A$; i.e.,

$$\mathcal{P}(A) = \{B : B \subseteq A\}.$$ 

Problem 1. If a set $S$ has $n$ elements for $n \in \mathbb{Z}$ and $n \geq 0$, how many elements are in $\mathcal{P}(S)$?

1.2 Algebra of Sets

Let $A$ and $B$ be sets. There are four common binary operations used on sets.\footnote{In the following, some standard logical notation is used. The symbol $\lor$ is the logical nonexclusive “or.” The symbol $\land$ is the logical “and.” Their truth tables are as follows:}

The union of $A$ and $B$ is the set

$$A \cup B = \{x : x \in A \lor x \in B\}.$$ 

The intersection of $A$ and $B$ is the set

$$A \cap B = \{x : x \in A \land x \in B\}.$$ 

The difference of $A$ and $B$ is the set

$$A \setminus B = \{x : x \in A \land x \notin B\}.$$ 

The symmetric difference of $A$ and $B$ is the set

$$A \Delta B = (A \cup B) \setminus (A \cap B).$$
Another common set operation is complementation. The complement of a set $A$ is usually thought of as the set consisting of all elements which are not in $A$. But, a short reflection will convince the reader that this is not a well-stated definition because the collection of elements not in $A$ is not a precisely understood collection. To make sense of the complement of a set, there must be a well-defined universal set $U$ which contains all the sets in question. Then the complement of a set $A \subset U$ is $A^c = U \setminus A$.

**Theorem 1.1.** Let $A$, $B$ and $C$ be sets.

(a) $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$

(b) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$

**Proof.** (a) This is proved as a sequence of equivalences.

\[
x \in A \setminus (B \cup C) \iff x \in A \land x \notin (B \cup C) \\
\quad \iff x \in A \land x \notin B \land x \notin C \\
\quad \iff (x \in A \land x \notin B) \land (x \in A \land x \notin C) \\
\quad \iff x \in (A \setminus B) \cap (A \setminus C)
\]

(b) This is also proved as a sequence of equivalences.

\[
x \in A \setminus (B \cap C) \iff x \in A \land x \notin (B \cap C) \\
\quad \iff x \in A \land (x \notin B \lor x \notin C) \\
\quad \iff (x \in A \land x \notin B) \lor (x \in A \land x \notin C) \\
\quad \iff x \in (A \setminus B) \cup (A \setminus C)
\]

\[\Box\]

Theorem 1.1 is a version of a group of set equations which are often called DeMorgan’s Laws. The more usual statement of DeMorgan’s Laws are in Corollary 1.2. Corollary 1.2 is an obvious consequence of Theorem 1.1 when there is a universal set to make the complementation well-defined.

**Corollary 1.2 (DeMorgan’s Laws).** Let $A$ and $B$ be sets.

(a) $(A \cup B)^c = A^c \cup B^c$

(b) $(A \cap B)^c = A^c \cup B^c$

**Problem 2.** Prove that for any sets $A$ and $B$,

(a) $A = (A \cap B) \cup (A \setminus B)$

(b) $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$ and that the sets $A \setminus B$, $B \setminus A$ and $A \cap B$ are pairwise disjoint.

We often have occasion to work with large collections of sets. For example, we could have a sequence of sets $A_1, A_2, A_3, \ldots$, where there is a set $A_n$ associated with each $n \in \mathbb{N}$. In general, let $\Lambda$ be a set and suppose for each $\lambda \in \Lambda$ there is a set $A_\lambda$. The collection $\{A_\lambda : \lambda \in \Lambda\}$ is called a collection of sets indexed by $\Lambda$. In this case, $\Lambda$ is called the indexing set for the collection.
Section 1: Sets and Functions

Example 1.1. For each \( n \in \mathbb{N} \), let
\[
A_n = \{k \in \mathbb{Z} : k^2 \leq n\}.
\]
Then
\[
A_1 = A_2 = A_3 = \{-1, 0, 1\}, \quad A_4 = \{-2, -1, 0, 1, 2\}, \ldots, \quad A_{50} = \{-7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7\}, \ldots
\]
is a collection of sets indexed by \( \mathbb{N} \).

Several of the binary operations can be extended to work with indexed collections. In particular, using the indexed collection from the previous paragraph, we define
\[
\bigcup_{\lambda \in \Lambda} A_\lambda = \{x : x \in A_\lambda \text{ for some } \lambda \in \Lambda\}
\]
and
\[
\bigcap_{\lambda \in \Lambda} A_\lambda = \{x : x \in A_\lambda \text{ for all } \lambda \in \Lambda\}.
\]

DeMorgan’s Laws can be generalized to indexed collections.

Theorem 1.3. If \( \{B_\lambda : \lambda \in \Lambda\} \) is an indexed collection of sets and \( A \) is a set, then
\[
A \setminus \bigcup_{\lambda \in \Lambda} B_\lambda = \bigcap_{\lambda \in \Lambda} (A \setminus B_\lambda)
\]
and
\[
A \setminus \bigcap_{\lambda \in \Lambda} B_\lambda = \bigcup_{\lambda \in \Lambda} (A \setminus B_\lambda).
\]

Problem 3. Prove Theorem 1.3.

1.3 Functions and Relations

When listing the elements of a set, the order in which they are listed is unimportant: e. g., \( \{a, b\} = \{b, a\} \). If the order in which \( n \) items are listed is important, the list is called an \( n \)-tuple. (Strictly speaking, an \( n \)-tuple is not a set.) We denote an \( n \)-tuple by enclosing the ordered list in parentheses. For example, if \( x_1, x_2, x_3, x_4 \) are 4 items, the 4-tuple \( (x_1, x_2, x_3, x_4) \) is different from the \( n \)-tuple \( (x_2, x_1, x_3, x_4) \).

Because they are used so often, a 2-tuple is called an ordered pair and a 3-tuple is called an ordered triple.

Definition 1.2. The Cartesian product of \( A \) and \( B \) is the set of all ordered pairs
\[
A \times B = \{(a, b) : a \in A \land b \in B\}.
\]
Example 1.2. If $A = \{a, b, c\}$ and $B = \{1, 2\}$, then

$$A \times B = \{(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2)\}.$$ 

A useful way to visualize the Cartesian product of two sets is as a table. The Cartesian product from Example 1.2 is contained in the entries of the following table.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$(a, 1)$</td>
<td>$(a, 2)$</td>
</tr>
<tr>
<td>$b$</td>
<td>$(b, 1)$</td>
<td>$(b, 2)$</td>
</tr>
<tr>
<td>$c$</td>
<td>$(c, 1)$</td>
<td>$(c, 2)$</td>
</tr>
</tbody>
</table>

Of course, the common Cartesian plane from analytic geometry is nothing more than a variation of this idea of listing the elements of a Cartesian product as a table.

By induction, the definition of Cartesian product can be extended to the case of more than two sets. If $\{A_1, A_2, \ldots, A_n\}$ are sets, then

$$A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) : a_k \in A_k \text{ for } 1 \leq k \leq n\}$$

is a set of $n$-tuples.

Definition 1.3. If $A$ and $B$ are sets, then any $R \subset A \times B$ is a relation from $A$ to $B$. If $(a, b) \in R$, we write $aRb$.

In this case,

$$\text{dom}(R) = \{a : (a, b) \in R\}$$

is the domain of $R$ and

$$\text{ran}(R) = \{b : (a, b) \in R\}$$

is the range of $R$.

Suppose $R \subset A \times A$.

The relation $R$ is symmetric, if $aRb \iff bRa$.

The relation $R$ is reflexive, if $aRa$ whenever $a \in A$.

The relation $R$ is transitive, if $aRb \wedge bRc \implies aRc$.

The relation $R$ is an equivalence relation on $A$, if it is symmetric, reflexive and transitive.

Example 1.3. Let $R$ be the relation on $\mathbb{Z} \times \mathbb{Z}$ defined by $aRb \iff a \leq b$. Then $R$ is reflexive and transitive, but not symmetric.

Example 1.4. Let $R$ be the relation on $\mathbb{Z} \times \mathbb{Z}$ defined by $aRb \iff a^2 = b^2$. In this case, $R$ is an equivalence relation. It is evident that $aRb$ iff $b = a$ or $b = -a$.

Problem 4. Suppose $R$ is an equivalence relation on $A$. For each $x \in A$ define $C_x = \{y \in A : xRy\}$. Prove that if $x, y \in A$, then either $C_x = C_y$ or $C_x \cap C_y = \emptyset$. (The collection $\{C_x : x \in A\}$ is the set of equivalence classes induced by $R$.)
Definition 1.4. A relation \( R \subset A \times B \) is a function if \( a R b_1 \land a R b_2 \implies b_1 = b_2. \)

If \( f \subset A \times B \) is a function and \( \text{dom}(f) = A \), then we usually write \( f : A \to B \) and \( f(a) = b \) instead of \( afb \).

If \( f : A \to B \) is a function, then the usual intuitive interpretation is to regard \( f \) as a rule that assigns each element of \( A \) to a unique element of \( B \). It’s not necessarily the case that each element of \( B \) is assigned something from \( A \).

Example 1.5. Define \( f : \mathbb{N} \to \mathbb{Z} \) by \( f(n) = n^2 \) and \( g : \mathbb{Z} \to \mathbb{Z} \) by \( g(n) = n^2 \). In this case \( \text{ran}(f) = \{ n^2 : n \in \mathbb{N} \} \) and \( \text{ran}(g) = \text{ran}(f) \cup \{ 0 \} \).

Definition 1.5. If \( f : A \to B \) and \( g : B \to C \), then the composition of \( g \) with \( f \) is the function \( g \circ f : A \to C \) defined by \( g \circ f(a) = g(f(a)) \).

In Example 1.5, \( g \circ f(n) = g(f(n)) = g(n^2) = (n^2)^2 = n^4 \) makes sense for all \( n \in \mathbb{N} \), but \( g \circ f \) is undefined at \( n = 0 \).

There are several important types of functions.

Definition 1.6. A function \( f : A \to B \) is a constant function, if \( \text{ran}(f) \) has a single element; i.e., there is a \( b \in B \) such that \( f(a) = b \) for all \( a \in A \).

Definition 1.7. A function \( f : A \to B \) is surjective (or onto \( B \)), if \( \text{ran}(f) = B \).

In a sense, constant and surjective functions are the opposite extremes. A constant function has the smallest possible range and a surjective function has the largest possible range. Of course, a function \( f : A \to B \) can be both constant and surjective, if \( B \) has only one element.

Definition 1.8. A function \( f : A \to B \) is injective (or one-to-one), if \( f(a) = f(b) \) implies \( a = b \).

The terminology "one-to-one" is very descriptive in this case. An illustration of this definition is in Figure 2. In Example 1.5, \( f \) is injective while \( g \) is not.

Definition 1.9. A function \( f : A \to B \) is bijective, if it is both surjective and injective.

A bijective function can be visualized as pairing up all the elements of \( A \) and \( B \). In a sense, \( A \) and \( B \) must have the same number of elements for this to happen. This idea will be explored further in the next section.

Definition 1.10. If \( f : A \to B \), \( C \subset A \) and \( D \subset B \), then the image of \( C \) is the set \( f(C) = \{ f(a) : a \in C \} \). The inverse image of \( D \) is the set \( f^{-1}(D) = \{ a : f(a) \in D \} \).

Definitions 1.9 and 1.10 work together in the following way. Suppose \( f : A \to B \) is bijective and \( b \in B \). The fact that \( f \) is surjective guarantees that \( f^{-1}(b) \neq \emptyset \). Since \( f \) is injective, \( f^{-1}(b) \) contains exactly one element, say \( a \), where \( f(a) = b \). In this way, it is seen that \( f^{-1} \) is a rule that assigns each element of \( B \) to exactly one element of \( A \); i.e., \( f^{-1} \) is a function with domain \( B \) and range \( A \).
Figure 2: These diagrams show two functions, $f : A \to B$ and $g : A \to B$. The function $g$ is injective and $f$ is not because $f(a) = f(c)$.

**Definition 1.11.** If $f : A \to B$ is bijective, we define the *inverse* of $f$ to be a function $f^{-1} : B \to A$ with the property that $f^{-1} \circ f(a) = a$ for all $a \in A$ and $f \circ f^{-1}(b) = b$ for all $b \in B$.

**Example 1.6.** Let $A = \mathbb{N}$ and $B$ be the even natural numbers. If $f : A \to B$ is $f(n) = 2n$ and $g : B \to A$ is $g(n) = n/2$, it is clear $f$ is bijective. Since $f \circ g(n) = f(n/2) = 2n/2 = n$ and $g \circ f(n) = g(2n) = 2n/2 = n$, we see $g = f^{-1}$.

Figure 3: This is one way to visualize a general invertible function. First $f$ does something to $a$ and then $f^{-1}$ undoes it.
Example 1.7. Let \( f : \mathbb{N} \to \mathbb{Z} \) be defined by
\[
 f(n) = \begin{cases} 
 (n-1)/2, & n \text{ odd}, \\
 -n/2, & n \text{ even} 
\end{cases}
\]
It’s quite easy to see that \( f \) is bijective and
\[
 f^{-1}(n) = \begin{cases} 
 2n+1, & n \geq 0, \\
 -2n, & n < 0 
\end{cases}
\]

The following theorem will be used in Section 1.4.

**Theorem 1.4 (Schröder-Bernstein).** Let \( A \) and \( B \) be sets. If there are injective functions \( f : A \to B \) and \( g : B \to A \), then there is a bijective function \( h : A \to B \).

**Proof.** Let \( B_1 = B \setminus f(A) \). If \( B_k \subset B \) is defined for some \( k \in \mathbb{N} \), let \( A_k = g(B_k) \) and \( B_{k+1} = f(A_k) \). This inductively defines \( A_k \) and \( B_k \) for all \( k \in \mathbb{N} \). Use these sets to define \( \hat{A} = \bigcup_{k \in \mathbb{N}} A_k \) and \( h : A \to B \) as
\[
 h(x) = \begin{cases} 
 g^{-1}(x), & x \in \hat{A} \\
 f(x), & x \in A \setminus \hat{A} 
\end{cases}
\]
It must be shown that \( h \) is well-defined, injective and surjective.

To show \( h \) is well-defined, let \( x \in A \). If \( x \in A \setminus \hat{A} \), then it is clear \( h(x) = f(x) \) is defined. On the other hand, if \( x \in \hat{A} \), then \( x \in A_k \) for some \( k \). Since \( x \in A_k = g(B_k) \), we see \( h(x) = g^{-1}(x) \) is defined. Therefore, \( h \) is well-defined.

To show \( h \) is injective, let \( x, y \in A \) with \( x \neq y \). If both \( x, y \in \hat{A} \) or \( x, y \in A \setminus \hat{A} \), then the assumptions that \( g \) and \( f \) are injective, respectively, imply \( h(x) \neq h(y) \). The remaining case is when \( x \in \hat{A} \) and \( y \in A \setminus \hat{A} \). Suppose \( x \in A_k \) and \( h(x) = h(y) \). Then there is an \( x_1 \in B_1 \) such that
\[
 x = g \circ f \circ g \circ f \circ \cdots \circ f \circ g(x_1).
\]
This implies
\[
 h(x) = g^{-1}(x) = f \circ g \circ f \circ \cdots \circ f \circ g(x_1) = f(y)
\]
so that
\[
 y = g \circ f \circ g \circ f \circ \cdots \circ f \circ g(x_1) \in A_{k-1} \subset \hat{A}.
\]
This contradiction shows that \( h(x) \neq h(y) \). We conclude \( h \) is injective.

To show \( h \) is surjective, let \( y \in B \). If \( y \in B_k \) for some \( k \), then \( h(A_k) = g^{-1}(A_k) = B_k \) shows \( y \in h(A) \). If \( y \notin B_k \) for any \( k \), \( y \in f(A) \) because \( B_1 = B \setminus f(A) \), and \( g(y) \notin \hat{A} \), so \( y = h(x) = f(x) \) for some \( x \in A \). This shows \( h \) is surjective.

**Problem 5.** If \( f : A \to B \) is bijective, then \( f^{-1} \) is unique.
1.4 Cardinality

There is a way to use sets to formalize and generalize how we count. For example, suppose we want to count how many elements are in the set \{a, b, c\}. The natural way to do this would be to point at each element in succession and say “one, two, three.” What is really happening is that we’re defining a bijective function between \{a, b, c\} and the set \{1, 2, 3\}. This idea can be generalized.

**Definition 1.12.** Given \( n \in \mathbb{N} \), an initial segment of \( \mathbb{N} \) is the set \( n = \{1, 2, \ldots, n\} \).

The trivial initial segment is \( 0 = \emptyset \). A set \( S \) has cardinality \( n \), if there is a bijective function \( f : S \to n \). In this case, we write \( \text{card}(S) = n \).

The cardinalities defined in Definition 1.12 are called the finite cardinal numbers. They correspond to the everyday counting numbers we usually use.

The idea can be generalized still further.

**Definition 1.13.** Let \( A \) and \( B \) be two sets. If there is an injective function \( f : A \to B \), we say \( \text{card}(A) \leq \text{card}(B) \).

According to Theorem 1.4, the Schröder-Bernstein Theorem, if \( \text{card}(A) \leq \text{card}(B) \) and \( \text{card}(B) \leq \text{card}(A) \), then there is a bijective function \( f : A \to B \). As expected, in this case we write \( \text{card}(A) = \text{card}(B) \). When \( \text{card}(A) \leq \text{card}(B) \), but no such bijection exists, we write \( \text{card}(A) < \text{card}(B) \).

In particular, a set \( A \) is countably infinite, if \( \text{card}(A) = \text{card}(\mathbb{N}) \). In this case, it is common to write \( \text{card}(\mathbb{N}) = \aleph_0 \).

This leaves open the question whether all sets either have finite cardinality, or are countably infinite. This is answered by letting \( S = \mathbb{N} \) in the following theorem.

**Theorem 1.5.** If \( S \) is a set, \( \text{card}(S) < \text{card}(\mathcal{P}(S)) \).

**Proof.** It is easy to see \( \text{card}(S) \leq \text{card}(\mathcal{P}(S)) \), so it suffices to prove there is no surjective function \( f : S \to \mathcal{P}(S) \).

To see this, assume there is such a function \( f \) and let \( T = \{ x \in S : x \notin f(x) \} \). Since \( f \) is surjective, there is a \( t \in T \) such that \( f(t) = T \). Either \( t \in T \) or \( t \notin T \).

If \( t \in T = f(T) \), then the definition of \( T \) implies \( t \notin T \), a contradiction. On the other hand, if \( t \notin T = f(T) \), then the definition of \( T \) implies \( t \in T \). These contradictions lead to the conclusion that no such function \( f \) can exist.

A set \( S \) is said to be uncountably infinite, or just uncountable, if \( \aleph_0 < \text{card}(S) \). Theorem 1.5 implies \( \aleph_0 < \text{card}(\mathcal{P}(\mathbb{N})) \), so \( \mathcal{P}(\mathbb{N}) \) is uncountable. In fact, the same argument implies

\[ \aleph_0 = \text{card}(\mathbb{N}) < \text{card}(\mathcal{P}(\mathbb{N})) < \text{card}(\mathcal{P}(\mathcal{P}(\mathbb{N}))) < \ldots \]

So, there are an infinite number of distinct infinite cardinalities.

\[ ^2 \text{The symbol } \aleph \text{ is the Hebrew letter “aleph” and } \aleph_0 \text{ is usually pronounced “aleph nought.”} \]
Extra Credit 1. Prove that if a set $S$ is countably or uncountably infinite, then there is a proper subset $T \subsetneq S$ and a bijection $f : S \rightarrow T$. This property is often used as the definition of when a set is infinite.

Notice that Theorem 1.5 does not imply there are no sets $B$ such that $\aleph_0 < \text{card}(B) < \text{card}(\mathcal{P}(\mathbb{N}))$. In fact, for many years the question of whether such sets exist was one of the most important open questions in mathematics. The assumption that no such sets exist is called the continuum hypothesis.

The continuum hypothesis was first stated as a conjecture by Georg Cantor in 1878. Kurt Gödel proved in 1938 that the continuum hypothesis does not contradict anything in normal set theory, but he did not prove it was true. In 1963 it was proved by Paul Cohen that the continuum hypothesis is actually unprovable as a theorem in standard set theory.

So, the continuum hypothesis is a statement which is neither true nor false within the framework of ordinary set theory. This means that in an axiomatic development of set theory, the continuum hypothesis, or a suitable negation of it, can be taken as an axiom.

The proofs of these theorems are quite complicated. A well-written introduction to many of these ideas is contained in the book by Ciesielski [1].

Problem 6. Suppose that $A_k$ is a set for each positive integer $k$.

(a) Show that $x \in \bigcap_{n=1}^{\infty} \left( \bigcup_{k=n}^{\infty} A_k \right)$ iff $x \in A_k$ for infinitely many sets $A_k$.

(b) Show that $x \in \bigcup_{n=1}^{\infty} \left( \bigcap_{k=n}^{\infty} A_k \right)$ iff $x \in A_k$ for all but finitely many of the sets $A_k$.

The set $x \in \bigcap_{n=1}^{\infty} \left( \bigcup_{k=n}^{\infty} A_k \right)$ from (a) is often called the superior limit of the sets $A_k$ and $x \in \bigcup_{n=1}^{\infty} \left( \bigcap_{k=n}^{\infty} A_k \right)$ is often called the inferior limit of the sets $A_k$.

Problem 7. Given two sets $A$ and $B$, it is common to let $A^B$ be the set of all functions $f : B \rightarrow A$. Prove that for any set $A$, $\text{card} \left( A^A \right) = \text{card} \left( \mathcal{P}(A) \right)$. 

2 The real numbers as a complete ordered field

In this section are presented what can be thought of as “the rules of the game:” the axioms of the real numbers. In this work, we present these axioms as rules without justification. There are other approaches which can be used. For example, another standard technique is to begin with the Peano axioms—the axioms of the natural numbers—and build up to the real numbers through several “completions” of this system. In such a setup, our axioms are theorems.

2.1 Field Axioms

This first set of axioms are called the field axioms because any object satisfying them is called a field. They give the algebraic properties of the real numbers.

A field is a nonempty set $F$ along with two functions, multiplication $\times : F \times F \to F$ and addition $+: F \times F \to F$ satisfying the following axioms.\(^3\)

**Axiom 1 (Associative Laws).** If $a, b, c \in F$, then $(a + b) + c = a + (b + c)$ and $(a \times b) \times c = a \times (b \times c)$.

**Axiom 2 (Commutative Laws).** If $a, b \in F$, then $a + b = b + a$ and $a \times b = b \times a$.

**Axiom 3 (Distributive Law).** If $a, b, c \in F$, then $a \times (b + c) = a \times b + a \times c$.

**Axiom 4 (Existence of identities).** There are $0, 1 \in F$ such that $a + 0 = a$ and $a \times 1 = a$, $\forall a \in F$.

**Axiom 5 (Existence of an additive inverse).** For each $a \in F$ there is $-a \in F$ such that $a + (-a) = 0$.

**Axiom 6 (Existence of a multiplicative inverse).** For each $a \in F \setminus \{0\}$ there is $a^{-1} \in F$ such that $a \times a^{-1} = 1$.

Although these axioms seem to contain most of the properties of the real numbers we normally use, there are other fields besides the real numbers.

**Example 2.1.** From elementary algebra we know that the rational numbers, $\mathbb{Q}$, are a field.

**Example 2.2.** Let $F = \{0, 1, 2\}$ with with addition and multiplication calculated modulo 3. It is easy to check that the field axioms are satisfied.

**Theorem 2.1.** The additive and multiplicative identities of a field $F$ are unique.

**Proof.** Suppose $e_1$ and $e_2$ are both multiplicative identities in $F$. Then

$$e_1 = e_1 \times e_2 = e_2,$$

so the multiplicative identity is unique. The proof for the additive identity is essentially the same. \(\square\)

---

\(^3\)The functions $+$ and $\times$ are often called binary operations. The standard notation of $+(a, b) = a + b$ and $\times(a, b) = a \times b$ is used here.
Theorem 2.2. Let $\mathbb{F}$ be a field. If $a, b \in \mathbb{F}$ with $b \neq 0$, then $-a$ and $b^{-1}$ are unique.

Proof. Suppose $b_1$ and $b_2$ are both multiplicative inverses for $b \neq 0$. Then, using Axiom 1,

$$b_1 = b_1 \times 1 = b_1 \times (b \times b_2) = (b_1 \times b) \times b_2 = 1 \times b_2 = b_2.$$ 

This shows the multiplicative inverse is unique. The proof is essentially the same for the additive inverse.

From now on we will assume the standard notations for algebra; e.g., we will write $ab$ instead of $a \times b$ and $a/b$ instead of $a \times b^{-1}$. There are many other properties of fields which could be proved here, but they correspond to the usual properties of the real numbers learned in beginning algebra, so we omit them.

Problem 8. Prove that if $a, b \in \mathbb{F}$, where $\mathbb{F}$ is a field, then $(-a)b = -(ab) = a(-b)$. 
2.2 Order Axiom

The axiom of this section gives us the order properties of the real numbers.

**Axiom 7 (Order axiom.).** There is a set \( P \subset \mathbb{F} \) such that

(i) If \( a, b \in P \), then \( a + b, ab \in P \).

(ii) If \( a \in \mathbb{F} \), then exactly one of the following is true: \( a \in P \), \( -a \in P \) or \( a = 0 \).

Of course, the \( P \) is known as the set of **positive** elements of \( \mathbb{F} \). Using Axiom 7(ii), we see that \( \mathbb{F} \) is divided into three pairwise disjoint sets: \( P \), \( \{0\} \) and \( \{ -x : x \in P \} \). The latter of these is the set of **negative** elements of \( \mathbb{F} \).

**Definition 2.1.** We write \( a < b \) or \( b > a \), if \( b - a \in P \). The meanings of \( a \leq b \) and \( b \geq a \) are now as expected.

**Example 2.3.** The rational numbers \( \mathbb{Q} \) are an ordered field. This example shows there are ordered fields which are not equal to \( \mathbb{R} \).

**Extra Credit 2.** Prove there is no set \( P \subset \mathbb{Z}_3 \) which makes \( \mathbb{Z}_3 \) into an ordered field.

Following are a few standard properties of ordered fields.

**Theorem 2.3.** \( a \neq 0 \) iff \( a^2 > 0 \).

*Proof.* \((\Rightarrow)\) If \( a > 0 \), then \( a^2 > 0 \) by Axiom 7(a). If \( a < 0 \), then \( -a > 0 \) by Axiom 7(b) and above, \( a^2 = 1a^2 = (-1)(-1)a^2 = (-a)^2 > 0 \).

\((\Leftarrow)\) Since \( 0^2 = 0 \), this is obvious. \( \Box \)

**Theorem 2.4.** If \( \mathbb{F} \) is an ordered field and \( a, b, c \in \mathbb{F} \), then

(a) \( a < b \iff a + c < b + c \),

(b) \( a < b \land b < c \implies a < c \),

(c) \( a < b \land c > 0 \implies ac < bc \),

(d) \( a < b \land c < 0 \implies ac > bc \).

*Proof.*

(a) \( a < b \iff b - a \in P \iff (b + c) - (a + c) \in P \iff a + c < b + c \).

(b) By supposition, both \( b - a, c - b \in P \). Using the fact that \( P \) is closed under addition, we see \( (b - a) + (c - b) = c - a \in P \). Therefore, \( c > a \).

(c) Since \( b - a \in P \) and \( c \in P \) and \( P \) is closed under multiplication, \( c(b - a) = cb - ca \in P \) and, therefore, \( ac < bc \).

(d) By assumption, \( b - a, -c \in P \). Apply part (c) and Problem 8. \( \Box \)

**Theorem 2.5 (Two out of three rule).** *Let \( \mathbb{F} \) be an ordered field and \( a, b, c \in \mathbb{F} \). If \( ab = c \) and any two of \( a \), \( b \) or \( c \) are positive, then so is the third.*
Proof. If $a > 0$ and $b > 0$, then Axiom 7(a) implies $c > 0$. Next, suppose $a > 0$ and $c > 0$. In order to force a contradiction, suppose $b \leq 0$. In this case, Axiom 7(b) shows

$$0 \leq a(-b) = -(ab) = -c < 0,$$

which is impossible.

Corollary 2.6. Let $\mathbb{F}$ be an ordered field and $a \in \mathbb{F}$. If $a > 0$, then $a^{-1} > 0$. If $a < 0$, then $a^{-1} < 0$.

Proof. The proof is Problem 9.


Suppose $a > 0$. Since $1a = a$, Theorem 2.5 implies $1 > 0$. Applying Theorem 2.4, we see that $1 + 1 > 1 > 0$. It’s clear that by induction, we can find a copy of $\mathbb{N}$ in any ordered field. Similarly, $\mathbb{Z}$ and $\mathbb{Q}$ also have unique copies in any ordered field.

The standard notation for intervals will be used on an ordered field, $\mathbb{F}$; i.e., $(a, b) = \{x \in \mathbb{F} : a < x < b\}$, $(a, \infty) = \{x \in \mathbb{F} : a < x\}$, $[a, b] = \{x \in \mathbb{F} : a \leq x \leq b\}$, etc.

2.2.1 Metric Properties

The order axiom on a field $\mathbb{F}$ allows us to introduce the idea of a distance between points in $\mathbb{F}$. To do this, we begin with the following familiar definition.

Definition 2.2. Let $\mathbb{F}$ be an ordered field. The absolute value function on $\mathbb{F}$ is a function $|\cdot| : \mathbb{F} \to \mathbb{F}$ defined as

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

The most important properties of the absolute value function are contained in the following theorem.

Theorem 2.7. Let $\mathbb{F}$ be an ordered field. Then

(a) $|x| \geq 0$ for all $x \in \mathbb{F}$ and $|x| = 0 \iff x = 0$;
(b) $|x| = |−x|$ for all $x \in \mathbb{F}$;
(c) $−|x| \leq x \leq |x|$ for all $x \in \mathbb{F}$;
(d) $|x| \leq y \iff −y \leq x \leq y$; and,
(e) $|x + y| \leq |x| + |y|$ for all $x, y \in \mathbb{F}$.

Proof. (a) The fact that $|x| \geq 0$ for all $x \in \mathbb{F}$ follows from Axiom 7(b). Since $0 = −0$, the second part is clear.
Section 2: The Axioms for the Real Numbers

(b) If \( x \geq 0 \), then \(-x \leq 0\) so that \(|-x| = -(-x) = x = |x|\). If \( x < 0 \), then \(-x > 0\) and \(|x| = -x = |-x|\).

(c) If \( x \geq 0 \), then \(-|x| = -x \leq x = |x|\). If \( x < 0 \), then \(-|x| = -(x) = x < -x = |x|\).

(d) This is left as an exercise.

(e) Add the two sets of inequalities \(-|x| \leq x \leq |x|\) and \(-|y| \leq y \leq |y|\) to see \(-(|x| + |y|) \leq x + y \leq |x| + |y|\). Now apply (d).

\[ \square \]

Definition 2.3. Let \( S \) be a set and \( d : S \times S \rightarrow \mathbb{R} \) satisfy

(a) for all \( x, y \in S \), \( d(x, y) \geq 0 \) and \( d(x, y) = 0 \iff x = y \),

(b) for all \( x, y \in S \), \( d(x, y) = d(y, x) \), and

(c) for all \( x, y, z \in S \), \( d(x, z) \leq d(x, y) + d(y, z) \).

Then the function \( d \) is a metric on \( S \).

A metric is a function which defines the distance between any two points of a set.

Example 2.4. Let \( S \) be a set and define \( d : S \times S \rightarrow S \) by

\[ d(x, y) = \begin{cases} 1, & x \neq y \\ 0, & x = y \end{cases} \]

It is easy to prove that \( d \) is a metric on \( S \). This trivial metric is called the discrete metric.

Theorem 2.8. If \( \mathbb{F} \) is an ordered field, then \( d(x, y) = |x - y| \) is a metric on \( \mathbb{F} \).

Proof. This easily follows from various parts of Theorem 2.7 \( \square \)

Problem 10. Prove \( |x| \leq y \) iff \(-y \leq x \leq y\).
2.3 The Completeness Axiom

Definition 2.4. A subset $S$ of an ordered field $\mathbb{F}$ is bounded above, if there exists $M \in \mathbb{F}$ such that $M \geq x$ for all $x \in S$. A subset $S$ of an ordered field $\mathbb{F}$ is bounded below, if there exists $m \in \mathbb{F}$ such that $m \leq x$ for all $x \in S$. The elements $M$ and $m$ are called upper and lower bounds for $S$, respectively.

Definition 2.5. Suppose $\mathbb{F}$ is an ordered field and $S$ is bounded above in $\mathbb{F}$. A number $B \in \mathbb{F}$ is called a least upper bound of $S$ if

(a) $B$ is an upper bound for $S$, and

(b) if $\alpha$ is any upper bound for $S$, then $B \leq \alpha$.

Generally, we denote $B = \text{lub } S$.

Suppose $\mathbb{F}$ is an ordered field and $S$ is bounded below in $\mathbb{F}$. A number $b \in \mathbb{F}$ is called a greatest lower bound of $S$ if

(a) $b$ is a lower bound for $S$, and

(b) if $\alpha$ is any lower bound for $S$, then $b \geq \alpha$.

Generally, we denote $b = \text{glb } S$.

Axiom 8 (Completeness). Every set which is bounded above has a least upper bound.

This is the final axiom. Any set which satisfies all eight axioms is called a complete ordered field. We assume the existence of a complete ordered field, called the real numbers. The real numbers are denoted by $\mathbb{R}$.

It can be shown that if $\mathbb{F}_1$ and $\mathbb{F}_2$ are both complete ordered fields, then they are the same, in the following sense. There exists a unique bijective function $i : \mathbb{F}_1 \to \mathbb{F}_2$ such that $i(a + b) = i(a) + i(b)$, $i(ab) = i(a)i(b)$ and $a < b \iff i(a) < i(b)$. Such a function $i$ is called an order isomorphism. The existence of such an order isomorphism shows that the real numbers are essentially unique. Further discussion of this would take us too far afield. More reading on this topic can be done in [2].

Theorem 2.9. If $A \subset \mathbb{R}$ is bounded above, then it has a unique least upper bound. If $A \subset \mathbb{R}$ is bounded below, then it has a unique greatest lower bound.

Proof. Suppose $a_1$ and $a_2$ are both least upper bounds for $A$. By the definition of least upper bound, $a_1 \leq a_2 \leq a_1$ implies $a_1 = a_2$. The proof is similar for the greatest lower bound. $\square$

Theorem 2.10. $\alpha = \text{lub } A$ iff $(\alpha, \infty) \cap A = \emptyset$ and for all $\varepsilon > 0$, $(\alpha - \varepsilon, \alpha] \cap A \neq \emptyset$. Similarly, $\beta = \text{glb } A$ iff $(-\infty, \beta] \cap A = \emptyset$ and for all $\varepsilon > 0$, $[\beta, \beta + \varepsilon) \cap A \neq \emptyset$.

Proof. We will prove the first statement, concerning the least upper bound. The second statement, concerning the greatest lower bound, follows similarly.
Section 2: The Axioms for the Real Numbers

(⇒) If $x \in (\alpha, \infty) \cap A$, then $\alpha$ cannot be an upper bound of $A$, which is a contradiction. If there is an $\varepsilon > 0$ such that $(\alpha - \varepsilon, \alpha] \cap A = \emptyset$, then from above, we conclude $(\alpha - \varepsilon, \infty) \cap A = \emptyset$. This implies $\alpha - \varepsilon/2$ is an upper bound for $A$ which is less than $\alpha = \text{lub} A$. This contradiction shows $(\alpha - \varepsilon, \alpha]\cap A \neq \emptyset$.

(⇐) The assumption that $(\alpha, \infty) \cap A = \emptyset$ implies $\alpha \geq \text{lub} A$. On the other hand, suppose $\text{lub} A < \alpha$. By assumption, there is an $x \in (\text{lub} A, \alpha) \cap A$. This is clearly a contradiction, since $\text{lub} A < x \in A$. Therefore, $\alpha = \text{lub} A$.

Corollary 2.11. If $\alpha = \text{lub} A$ and $\alpha \notin A$, then for all $\varepsilon > 0$, $(\alpha - \varepsilon, \alpha]\cap A$ is an infinite set. Similarly, if $\beta = \text{lub} A$ and $\beta \notin A$, then for all $\varepsilon > 0$, $(\beta - \varepsilon, \beta]\cap A$ is an infinite set.

Proof. For each $n \in \mathbb{N}$, use Theorem 2.10 to choose $x_n \in (\alpha - 1/n, \alpha] \cap A$. Given $\varepsilon > 0$ let $N \in \mathbb{N}$ be large enough so that $0 < 1/N < \varepsilon$. Then, $\{x_n : n \geq N\} \subset (\alpha - \varepsilon, \alpha]\cap A$, and the corollary is proved.

Problem 11. Let $A \subset \mathbb{R}$ be bounded above and

$$B = \{x : x \text{ is an upper bound of } A\}.$$ 

Prove $\text{lub} A = \text{glb} B$.

If a set $A$ is not bounded above, then it is usual to write $\text{lub} A = \infty$. Notice that the symbol “$\infty$” is not a number. It is really a short way to say that there is no number which is an upper bound for $A$. Similarly, if $B$ has no lower bound, then $\text{glb} B = -\infty$.

An interesting observation is that $\text{lub} \emptyset = -\infty$ and $\text{glb} \emptyset = \infty$. To see the first of these, notice that every $M \in \mathbb{R}$ is an upper bound for the empty set. This is because, given $M$, there is no $x \in A$ such that $x \geq M$. Thus, the set of upper bounds for $A$ has no lower bound.

Theorem 2.12 (Archimedean Principle). If $a \in \mathbb{R}$, then there exists $n_a \in \mathbb{N}$ such that $n_a > a$.

Proof. If the theorem is false, then $a$ is an upper bound for $\mathbb{N}$. Let $\alpha = \text{lub} \mathbb{N}$. According to Theorem 2.10 there is an $m \in \mathbb{N}$ such that $m > \alpha - 1$. But, this is a contradiction because $\alpha = \text{lub} \mathbb{N} < m + 1 \in \mathbb{N}$.

Some other variations on this theme are in the following corollary.

Corollary 2.13. Let $a, b \in \mathbb{R}$ with $a > 0$.

(a) There is an $n \in \mathbb{N}$ such that $an > b$.

(b) There is an $n \in \mathbb{N}$ such that $0 < 1/n < a$.

(c) There is an $n \in \mathbb{N}$ such that $n - 1 \leq a < n$.

Proof. (a) Use Theorem 2.12 to find $n \in \mathbb{N}$ where $0 < b/a < n$.

(b) Let $b = 1$ in part (a).

(c) Theorem 2.12 guarantees that $S = \{n \in \mathbb{N} : n > a\} \neq \emptyset$. If $n$ is the least element of this set, then $n - 1 \notin S$ and $n - 1 \leq a < n$. 

□
2.4 Existence of $\sqrt{2}$

All of the above still does not establish that $\mathbb{Q}$ is different from $\mathbb{R}$. Since $\mathbb{Q} \subset \mathbb{R}$, we must find a real number which is not rational. The following two propositions show that $\sqrt{2} \in \mathbb{R} \setminus \mathbb{Q}$.

**Theorem 2.14.** There is a positive $\alpha \in \mathbb{R}$ such that $\alpha^2 = 2$.

**Proof.** Let $S = \{x > 0 : x^2 < 2\}$. Then $1 \in S$, so $S \neq \emptyset$. If $x \geq 2$, then Theorem 2.4(c) implies $x^2 \geq 4 > 2$, so $S$ is bounded above. Let $\alpha = \text{lub } S$. It will be shown that $\alpha^2 = 2$.

Suppose first that $\alpha^2 < 2$. This assumption implies $(2 - \alpha^2)/(2\alpha + 1) > 0$. According to Corollary 2.13, there is an $n \in \mathbb{N}$ large enough so that

$$0 < \frac{1}{n} < \frac{2 - \alpha^2}{2\alpha + 1} \implies 0 < \frac{2\alpha + 1}{n} < 2 - \alpha^2.$$ 

Therefore,

$$\left(\alpha + \frac{1}{n}\right)^2 = \alpha^2 + \frac{2\alpha}{n} + \frac{1}{n^2} = \alpha^2 + \frac{1}{n} \left(2\alpha + \frac{1}{n}\right)$$

$$< \alpha^2 + \frac{(2\alpha + 1)}{n} < \alpha^2 + (2 - \alpha^2) = 2$$

contradicts the fact that $\alpha = \text{lub } S$.

Next, assume $\alpha^2 > 2$. In this case, choose $n \in \mathbb{N}$ so that

$$0 < \frac{1}{n} < \frac{\alpha^2 - 2}{2\alpha} \implies 0 < \frac{2\alpha}{n} < \alpha^2 - 2.$$ 

Then

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

again contradicts that $\alpha = \text{lub } S$.

Therefore, $\alpha^2 = 2$. \qed

**Theorem 2.15.** There is no $\alpha \in \mathbb{Q}$ such that $\alpha^2 = 2$.

**Proof.** Assume to the contrary that there is $\alpha \in \mathbb{Q}$ with $\alpha^2 = 2$. Then there are $p, q \in \mathbb{N}$ such that $\alpha = p/q$ and $p$ and $q$ are relatively prime. Now,

$$\left(\frac{p}{q}\right)^2 = 2 \implies p^2 = 2q^2$$

(1)

shows $p^2$ is even. Since the square of an odd number is odd, $p$ must be even; i.e., $p = 2f$ for some $f \in \mathbb{N}$. Substituting this into (1), shows $2f^2 = q^2$. The same argument as above establishes $q$ is also even. This contradicts the assumption that $p$ and $q$ are relatively prime. Therefore, no such $\alpha$ exists. \qed
Section 3: Sequences

3 Sequences

3.1 Basic Properties

Definition 3.1. A sequence is a function \( a : \mathbb{N} \to \mathbb{R} \).

Instead of using the standard function notation of \( a(n) \) for sequences, it is usually more convenient to write the argument of the function as a subscript, \( a_n \).

Example 3.1. Let the sequence \( a_n = 1 - 1/n \). Then an easy calculation shows \( a_1 = 0 \), \( a_2 = 1/2 \), \( a_3 = 2/3 \), etc.

Example 3.2. Let the sequence \( b_n = 2^n \). It’s easy to see \( b_1 = 2 \), \( b_2 = 4 \), \( b_3 = 8 \), etc.

Definition 3.2. A sequence \( a_n \) is bounded if \( \{ a_n : n \in \mathbb{N} \} \) is a bounded set. This definition is extended in the obvious way to bounded above and bounded below.

The sequence of Example 3.1 is bounded, but the sequence of Example 3.2 is not.

Definition 3.3. A sequence \( a_n \) converges to \( L \in \mathbb{R} \) if for all \( \varepsilon > 0 \) there exists an \( N \in \mathbb{N} \) such that whenever \( n \geq N \), then \( |a_n - L| < \varepsilon \). If a sequence does not converge, then it is said to diverge.

When \( a_n \) converges to \( L \), we write \( \lim_{n \to \infty} a_n = L \), or often, more simply, \( a_n \to L \).

Example 3.3. Let \( a_n \) be as in Example 3.1. We claim \( a_n \to 1 \). To see this, let \( \varepsilon > 0 \) and choose \( N \in \mathbb{N} \) such that \( 1/N < \varepsilon \). Then, if \( n \geq N \)

\[
|a_n - 1| = |(1 - 1/n) - 1| = 1/n \leq 1/N < \varepsilon,
\]

so \( a_n \to 1 \).

The sequence \( b_n \) of Example 3.2 diverges. To see this, suppose not. Then there is an \( L \in \mathbb{R} \) such that \( b_n \to L \). If \( \varepsilon = 1 \), there must be an \( N \in \mathbb{N} \) such that \( |b_n - L| < \varepsilon \) whenever \( n \geq N \). Choose \( n \geq N \). \(|L - 2^n| < 1 \) implies \( L < 2^n + 1 \).

But, then

\[
b_{n+1} - L = 2^{n+1} - L > 2^{n+1} - (2^n + 1) = 2^n - 1 \geq 1 = \varepsilon.
\]

This violates the condition on \( N \). We conclude that for every \( L \in \mathbb{R} \) there exists an \( \varepsilon > 0 \) such that for no \( N \in \mathbb{N} \) is it true that whenever \( n \geq N \), then \( |b_n - L| < \varepsilon \). Therefore, \( b_n \) diverges.

Definition 3.4. A sequence \( a_n \) diverges to \( \infty \) if for every \( B > 0 \) there is an \( N \in \mathbb{N} \) such that \( n \geq N \) implies \( a_n > B \). The sequence \( a_n \) is said to diverge to \( -\infty \) if \(-a_n \) diverges to \( \infty \).

When \( a_n \) diverges to \( \infty \), we write \( \lim_{n \to \infty} a_n = \infty \), or often, more simply, \( a_n \to \infty \).
Example 3.4. It is easy to prove that the sequence of Example 3.2 diverges to ∞.

**Theorem 3.1.** If $a_n \to L$, then $L$ is unique.

**Proof.** Suppose $a_n \to L_1$ and $a_n \to L_2$. Let $\varepsilon > 0$. According to Definition 3.2, there exist $N_1, N_2 \in \mathbb{N}$ such that $n \geq N_1$ implies $|a_n - L_1| < \varepsilon / 2$ and $n \geq N_2$ implies $|a_n - L_2| < \varepsilon / 2$. Set $N = \max\{N_1, N_2\}$. If $n \geq N$, then

$$|L_1 - L_2| = |L_1 - a_n + a_n - L_2| \leq |L_1 - a_n| + |a_n - L_2| < \varepsilon / 2 + \varepsilon / 2 = \varepsilon.$$  

Since $\varepsilon$ is an arbitrary positive number, this implies $L_1 = L_2$. \hfill \Box

**Theorem 3.2.** $a_n \to L$ iff for all $\varepsilon > 0$, the set $\{n : a_n \notin (L - \varepsilon, L + \varepsilon)\}$ is finite.

**Proof.** ($\Rightarrow$) Let $\varepsilon > 0$. According to Definition 3.2, there is an $N \in \mathbb{N}$ such that $\{n : n \geq N\} \subset (L - \varepsilon, L + \varepsilon)$. Then $\{n : a_n \notin (L - \varepsilon, L + \varepsilon)\} \subset \{1, 2, \ldots, N - 1\}$.

($\Leftarrow$) Let $\varepsilon > 0$. By assumption $\{n : a_n \notin (L - \varepsilon, L + \varepsilon)\}$ is finite, so let $N = \max\{n : a_n \notin (L - \varepsilon, L + \varepsilon)\} + 1$. If $n \geq N$, then $a_n \in (L - \varepsilon, L + \varepsilon)$, so, by Definition 3.2, $a_n \to L$. \hfill \Box

**Corollary 3.3.** If $a_n$ converges, then $a_n$ is bounded.

**Proof.** Suppose $a_n \to L$. According to Theorem 3.2 there are a finite number of terms of the sequence lying outside $(L - 1, L + 1)$. Since any finite set is bounded, the conclusion is obvious. \hfill \Box

**Theorem 3.4.** Let $a_n$ and $b_n$ be sequences such that $a_n \to A$ and $b_n \to B$. Then

(a) $a_n + b_n \to A + B$,

(b) $ca_n \to cA$, for all $c \in \mathbb{R}$,

(c) $a_nb_n \to AB$, and

(d) $a_n/b_n \to A/B$ as long as $b_n \neq 0$ for all $n \in \mathbb{N}$ and $B \neq 0$.

**Proof.** (a) Let $\varepsilon > 0$. There are $N_1, N_2 \in \mathbb{N}$ such that $n \geq N_1$ implies $|a_n - A| < \varepsilon / 2$ and $n \geq N_2$ implies $|b_n - B| < \varepsilon / 2$. Define $N = \max\{N_1, N_2\}$. If $n \geq N$, then

$$|(a_n + b_n) - (A + B)| \leq |a_n - A| + |b_n - B| < \varepsilon / 2 + \varepsilon / 2 = \varepsilon.$$  

Therefore $a_n + b_n \to A + B$.

(b) If $c = 0$, the statement is obvious. So, assume $c \neq 0$ and let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ so that whenever $n \geq N$, then $|a_n - A| < \varepsilon / |c|$. If $n \geq N$, then

$$|ca_n - cA| = |c||a_n - A| < |c|\varepsilon / c = \varepsilon.$$  

Therefore, $ca_n \to cA$. 

Section 3: Sequences

(c) Let $\varepsilon > 0$ and $\alpha > 0$ be an upper bound for $|a_n|$. Choose $N_1, N_2 \in \mathbb{N}$ such that $n \geq N_1 \Rightarrow |a_n - A| < \varepsilon / 2(|B| + 1)$ and $n \geq N_2 \Rightarrow |b_n - B| < \varepsilon / 2\alpha$. If $n \geq N = \max\{N_1, N_2\}$, then

\[ |a_nb_n - AB| = |a_nb_n - a_nB + a_nB - AB| \]
\[ \leq |a_nb_n - a_nB| + |a_nB - AB| \]
\[ = |a_n||b_n - B| + ||B|||a_n - A| \]
\[ < \frac{\alpha \varepsilon}{2\alpha} + |B| \frac{\varepsilon}{2(|B| + 1)} \]
\[ < \varepsilon / 2 + \varepsilon / 2 = \varepsilon. \]

(d) First, notice that it suffices to show that $1/b_n \to B$, because part (c) of this theorem can be used to achieve the full result.

Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ so that $n \geq N \Rightarrow |b_n| > B/2$ and $|b_n - B| < B^2\varepsilon / 2$. Then, when $n \geq N$,

\[ \left| \frac{1}{b_n} - \frac{1}{B} \right| = \left| \frac{B - b_n}{b_nB} \right| < \frac{B^2\varepsilon / 2}{(B/2)B} = \varepsilon. \]

Therefore $1/b_n \to 1/B$. \hfill \Box

**Theorem 3.5 (Sandwich Theorem).** Suppose $a_n$, $b_n$ and $c_n$ are sequences such that $a_n \leq b_n \leq c_n$ for all $n \in \mathbb{N}$.

(a) If $a_n \to L$ and $c_n \to L$, then $b_n \to L$.

(b) If $b_n \to \infty$, then $c_n \to \infty$.

(c) If $c_n \to -\infty$, then $b_n \to -\infty$.

**Proof.** (a) Let $\varepsilon > 0$. There is an $N \in \mathbb{N}$ large enough so that when $n \geq N$, then $L - \varepsilon < a_n$ and $c_n < L + \varepsilon$. These inequalities imply $L - \varepsilon < a_n \leq b_n \leq c_n < L + \varepsilon$. Therefore, $c_n \to L$.

(b) Let $B > 0$ and choose $N \in \mathbb{N}$ so that $n \geq N \Rightarrow b_n > B$. Then $c_n \geq b_n > B$ whenever $n \geq N$. This shows $c_n \to \infty$.

(c) This is essentially the same as part (b). \hfill \Box

**Problem 12.** Show that the sequence $a_n = \frac{3n + 1}{2n + 3}$ converges.

**Extra Credit 3.** If $a_n \to L$, then what can you say about

\[ \sigma_n = \frac{a_1 + a_2 + \cdots + a_n}{n}. \]

Is there a divergent sequence $a_n$ such that $\sigma_n$ converges?

**Problem 13.** A sequence $a_n$ converges to 0 iff $|a_n|$ converges to 0.
3.2 Monotone Sequences

**Definition 3.5.** A sequence \( a_n \) is *increasing*, if \( a_{n+1} \geq a_n \) for all \( n \in \mathbb{N} \). It is *strictly increasing* if \( a_{n+1} > a_n \) for all \( n \in \mathbb{N} \).

A sequence \( a_n \) is *decreasing*, if \( a_{n+1} \leq a_n \) for all \( n \in \mathbb{N} \). It is *strictly decreasing* if \( a_{n+1} < a_n \) for all \( n \in \mathbb{N} \).

If \( a_n \) is any of the four types listed above, then it is said to be a *monotone* sequence.

**Theorem 3.6.** A bounded monotone sequence converges.

*Proof.* Suppose \( a_n \) is a bounded increasing sequence, \( L = \text{lub} \{a_n : n \in \mathbb{N}\} \) and \( \varepsilon > 0 \). Clearly, \( a_n \leq L \) for all \( n \in \mathbb{N} \). According to Theorem 2.10, there exists an \( N \in \mathbb{N} \) such that \( a_N > L - \varepsilon \). Then \( L \geq a_n \geq a_N > L - \varepsilon \) for all \( n \geq N \). This shows \( a_n \to L \).

If \( a_n \) is decreasing, let \( b_n = -a_n \) and apply the preceding argument. \( \square \)

**Theorem 3.7.** An unbounded monotone sequence diverges to \( \infty \) or \(-\infty\), depending on whether it is increasing or decreasing, respectively.

*Proof.* Suppose \( a_n \) is increasing and unbounded. If \( B > 0 \), the fact that \( a_n \) is unbounded yields an \( N \in \mathbb{N} \) such that \( a_N > B \). Since \( a_n \) is increasing, \( a_n \geq a_N > B \) for all \( n \geq N \). This shows \( a_n \to \infty \).

The proof when the sequence decreases is similar. \( \square \)

3.3 The Nested Interval Theorem

**Definition 3.6.** A collection of sets \( \{S_n : n \in \mathbb{N}\} \) is said to be *nested*, if \( S_{n+1} \subset S_n \) for all \( n \in \mathbb{N} \).

**Theorem 3.8 (Nested Interval Theorem).** If \( I_n = [a_n, b_n] \) is a nested collection of closed intervals such that \( \lim_{n \to \infty} b_n - a_n = 0 \), then there is an \( x \in \mathbb{R} \) such that \( \bigcap_{n \in \mathbb{N}} I_n = \{x\} \).

*Proof.* Since the intervals are nested, it’s clear that \( a_n \) is an increasing sequence bounded above by \( b_1 \) and \( b_n \) is a decreasing sequence bounded below by \( a_1 \). Applying Theorem 3.6 twice, we find there are \( \alpha, \beta \in \mathbb{R} \) such that \( a_n \to \alpha \) and \( b_n \to \beta \).

We claim \( \alpha = \beta \). To see this, let \( \varepsilon > 0 \) and use the “shrinking” condition on the intervals to pick \( N \in \mathbb{N} \) so that \( b_N - a_N < \varepsilon \). The nestedness of the intervals implies \( a_N \leq a_n < b_n \leq b_N \) for all \( n \geq N \). Therefore

\[
|a_N - \beta| \leq |b_n - a_N| < \varepsilon.
\]

This shows \( |\alpha - \beta| \leq |b_N - a_N| < \varepsilon \). Since \( \varepsilon > 0 \) was chosen arbitrarily, we conclude \( \alpha = \beta \).

Let \( x = \alpha = \beta \). It remains to show that \( \bigcap_{n \in \mathbb{N}} I_n = \{x\} \).

First, we show that \( x \in \bigcap_{n \in \mathbb{N}} I_n \). To do this, fix \( N \in \mathbb{N} \). Since \( a_n \) increases to \( x \), it’s clear that \( x \geq a_N \). Similarly, \( x \leq b_N \). Therefore \( x \in [a_N, b_N] \). Because \( N \) was chosen arbitrarily, it follows that \( x \in \bigcap_{n \in \mathbb{N}} I_n \).
Next, suppose there are \( x, y \in \bigcap_{n \in \mathbb{N}} I_n \) and let \( \varepsilon > 0 \). Choose \( N \in \mathbb{N} \) such that \( b_N - a_N < \varepsilon \). Then \( \{x, y\} \subseteq \bigcap_{n \in \mathbb{N}} I_n \subseteq [a_N, b_N] \) implies \(|x - y| < \varepsilon\). Since \( \varepsilon \) was chosen arbitrarily, we see \( x = y \). Therefore \( \bigcap_{n \in \mathbb{N}} I_n = \{x\} \).

**Example 3.5.** If \( I_n = (0, 1/n) \) for all \( n \in \mathbb{N} \), then the collection \( \{I_n : n \in \mathbb{N}\} \) is nested, but \( \bigcap_{n \in \mathbb{N}} I_n = \emptyset \). This shows the assumption that the intervals be closed in the Nested Interval Theorem is necessary.

**Example 3.6.** If \( I_n = [n, \infty) \) then the collection \( \{I_n : n \in \mathbb{N}\} \) is nested, but \( \bigcap_{n \in \mathbb{N}} I_n = \emptyset \). This shows that the assumption that the lengths of the intervals be bounded is necessary.

**Extra Credit 4.** If \( a_n \) is a sequence such that \( \frac{a_n - 1}{a_{n+1}} \to 0 \), then does \( \lim_{n \to \infty} a_n \) exist?

**Extra Credit 5.** Suppose a sequence is defined by \( a_1 = 0 \), \( a_1 = 1 \) and \( a_{n+1} = \frac{1}{2}(a_n + a_{n-1}) \) for \( n \geq 2 \). Prove \( a_n \) converges, and determine its limit.

**Problem 14.** Prove that the sequence \( a_n = n^3/n! \) converges.

### 3.4 Subsequences

**Definition 3.7.** Let \( a_n \) be a sequence and \( \sigma : \mathbb{N} \to \mathbb{N} \) be a strictly increasing function. Then \( b_n = a_{\sigma(n)} \) is a subsequence of \( a_n \).

The idea here is that the subsequence \( b_n \) is a new sequence formed from an old sequence \( a_n \) by possibly leaving terms out of \( a_n \). In other words, we see that all the terms of \( b_n \) must also appear in \( a_n \), and they must appear in the same order.

**Example 3.7.** If \( a_n = \sin(n\pi/2) \), then some possible subsequences are

\[
b_n = a_{2n-1} \implies b_n = (-1)^{n+1},
\]

\[
c_n = a_{2n} \implies c_n = 0,
\]

and

\[
d_n = a_{n^2} \implies d_n = (1 + (-1)^{n+1})/2.
\]

**Theorem 3.9.** \( a_n \to L \) iff every subsequence of \( a_n \) converges to \( L \).

**Proof.** (\( \Rightarrow \)) Suppose \( \sigma : \mathbb{N} \to \mathbb{N} \) is strictly increasing, as in the preceding definition. Clearly, \( \sigma(1) \geq 1 \). Suppose \( \sigma(n) \geq n \) for some \( n \in \mathbb{N} \). Then \( \sigma(n + 1) > \sigma(n) \geq n \Rightarrow \sigma(n + 1) \geq n + 1 \). This simple induction argument has established \( \sigma(n) \geq n \) for all \( n \in \mathbb{N} \).

Now, suppose \( a_n \to L \) and \( b_n = a_{\sigma(n)} \) is a subsequence of \( a_n \). If \( \varepsilon > 0 \), there is an \( N \in \mathbb{N} \) such that \( n \geq N \) implies \( a_n \in (L - \varepsilon, L + \varepsilon) \). From the preceding paragraph, it follows that when \( n \geq N \), then \( b_n = a_{\sigma(n)} = a_m \) for some \( m \geq n \). So, \( b_n \in (L - \varepsilon, L + \varepsilon) \) and \( b_n \to L \).

(\( \Leftarrow \)) Since \( a_n \) is a subsequence of itself, it is obvious that \( a_n \to L \). \( \Box \)
Any sequence has an uncountable number of subsequences. Even if the original sequence diverges, it is possible there are convergent subsequences. For example, consider the divergent sequence $a_n = (-1)^n$. In this case, $a_n$ diverges, but the two subsequences $a_{2n}$ and $a_{2n+1}$ are constant sequences, so they converge.

**Problem 15.** If $a_n$ is a sequence such that every subsequence of $a_n$ has a further subsequence converging to 0, then $a_n \to 0$. 
Section 4: The Topology of $\mathbb{R}$

Definition 4.1. A set $G \subset \mathbb{R}$ is open if for every $x \in G$ there is an $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \subset G$. A set $F \subset \mathbb{R}$ is closed if $F^c$ is open.

Example 4.1. Any open interval $(a, b)$ is open. To see this, let $x \in (a, b)$ and $\varepsilon = \min\{x - a, b - x\}$. Then $(x - \varepsilon, x + \varepsilon) \subset (a, b)$.

Open half-lines are also open sets. For example, let $x \in (a, \infty)$ and $\varepsilon = x - a$. Then $(x - \varepsilon, x + \varepsilon) \subset (a, \infty)$.

A singleton set $\{a\}$ is closed. To see this, suppose $x \neq a$ and $\varepsilon = |x - a|$. Then $a \notin (x - \varepsilon, x + \varepsilon)$, and $\{a\}^c$ must be open. The definition of a closed set then implies $\{a\}$ is closed.

There are sets which are neither open nor closed. For example, consider the half-open interval $[0, 1)$. To see it isn’t open or closed, let $\varepsilon > 0$. Then $(0 - \varepsilon, 0 + \varepsilon) \not\subset [0, 1)$ shows it cannot be open. Since $(1 - \varepsilon, 1 + \varepsilon) \not\subset [0, 1)^c$, we see $[0, 1)^c$ is not open, so $[0, 1)$ cannot be closed.

Theorem 4.1. (a) If $\{G_\lambda : \lambda \in \Lambda\}$ is a collection of open sets, then $\bigcup_{\lambda \in \Lambda} G_\lambda$ is open.

(b) If $\{G_k : 1 \leq k \leq n\}$ is a finite collection of open sets, then $\bigcap_{k=1}^n G_k$ is open.

(c) Both $\emptyset$ and $\mathbb{R}$ are open.

Proof. (a) If $x \in \bigcup_{\lambda \in \Lambda} G_\lambda$, then there is a $\lambda_x \in \Lambda$ such that $x \in G_{\lambda_x}$. Since $G_{\lambda_x}$ is open, there is an $\varepsilon > 0$ such that $x \in (x - \varepsilon, x + \varepsilon) \subset G_{\lambda_x} \subset \bigcup_{\lambda \in \Lambda} G_\lambda$. This shows $\bigcup_{\lambda \in \Lambda} G_\lambda$ is open.

(b) If $x \in \bigcap_{k=1}^n G_k$, then $x \in G_k$ for $1 \leq k \leq n$. For each $G_k$ there is an $\varepsilon_k$ such that $(x - \varepsilon_k, x + \varepsilon_k) \subset G_k$. Let $\varepsilon = \min\{\varepsilon_k : 1 \leq k \leq n\}$. Then $(x - \varepsilon, x + \varepsilon) \subset G_k$ for $1 \leq k \leq n$, so $(x - \varepsilon, x + \varepsilon) \subset \bigcap_{k=1}^n G_k$. Therefore $\bigcap_{k=1}^n G_k$ is open.

(c) $\emptyset$ is open vacuously. $\mathbb{R}$ is obviously open.

Applying DeMorgan’s laws to the parts of Theorem 4.1 immediately yields the following.

Corollary 4.2. (a) If $\{F_\lambda : \lambda \in \Lambda\}$ is a collection of closed sets, then $\bigcap_{\lambda \in \Lambda} G_\lambda$ is closed.

(b) If $\{F_k : 1 \leq k \leq n\}$ is a finite collection of closed sets, then $\bigcup_{k=1}^n G_k$ is closed.

(c) Both $\emptyset$ and $\mathbb{R}$ are closed.

Notice that $\emptyset$ and $\mathbb{R}$ are both open and closed. They are the only subsets of $\mathbb{R}$ with this dual personality.
**Definition 4.2.** \( x_0 \) is a limit point of \( S \subset \mathbb{R} \) if for every \( \varepsilon > 0 \), \((x_0 - \varepsilon, x_0 + \varepsilon) \cap S \setminus \{x_0\} \neq \emptyset \). The derived set of \( S \) is

\[
S' = \{x : x \text{ is a limit point of } S\}.
\]

A point \( x_0 \in S \setminus S' \) is an isolated point of \( S \).

Notice that limit points of \( S \) need not be elements of \( S \), but isolated points of \( S \) must be elements of \( S \). In a sense, limit points and isolated points are at opposite extremes. The definitions can be restated as follows:

\[
x_0 \text{ is a limit point of } S \iff \forall \varepsilon > 0, S \cap (x_0 - \varepsilon, x_0 + \varepsilon) \setminus \{x_0\} \neq \emptyset
\]

\[
x_0 \text{ is an isolated point of } S \iff \exists \varepsilon > 0, S \cap (x_0 - \varepsilon, x_0 + \varepsilon) \setminus \{x_0\} = \emptyset
\]

**Example 4.2.** If \( S = (0, 1] \), then \( S' = [0, 1] \) and \( S \) has no isolated points.

**Example 4.3.** If \( T = \{1/n : n \in \mathbb{Z} \setminus \{0\}\} \), then \( T' = \{0\} \) and all points of \( T \) are isolated points of \( T \).

**Theorem 4.3.** \( x_0 \) is a limit point of \( S \) iff there is a sequence \( x_n \in S \setminus \{x_0\} \) such that \( x_n \to x_0 \).

**Proof.** \((\Rightarrow)\) For each \( n \in \mathbb{N} \), choose \( x_n \in S \cap (x_0 - 1/n, x_0 + 1/n) \setminus \{x_0\} \). Then \( |x_n - x_0| < 1/n \) for all \( n \in \mathbb{N} \), so \( x_n \to x_0 \).

\((\Leftarrow)\) Suppose \( x_n \) is a sequence from \( x_n \in S \setminus \{x_0\} \) converging to \( x_0 \). If \( \varepsilon > 0 \), the definition of convergence for a sequence yields an \( N \in \mathbb{N} \) such that whenever \( n \geq N \), then \( x_n \in S \cap (x_0 - \varepsilon, x_0 + \varepsilon) \setminus \{x_0\} \). This shows \( S \cap (x_0 - \varepsilon, x_0 + \varepsilon) \setminus \{x_0\} \neq \emptyset \), and \( x_0 \) must be a limit point of \( S \).

**Theorem 4.4.** A set \( S \subset \mathbb{R} \) is closed iff it contains all its limit points.

**Proof.** \((\Rightarrow)\) Suppose \( S \) is closed and \( x_0 \) is a limit point of \( S \). If \( x_0 \notin S \), then \( S^c \) open implies the existence of \( \varepsilon > 0 \) such that \((x_0 - \varepsilon, x_0 + \varepsilon) \cap S = \emptyset \). This contradicts the fact that \( x_0 \) is a limit point of \( S \). Therefore, \( x_0 \in S \) and \( S \) contains all its limit points.

\((\Leftarrow)\) Since \( S \) contains all its limit points, if \( x_0 \notin S \), there must exist an \( \varepsilon > 0 \) such that \((x_0 - \varepsilon, x_0 + \varepsilon) \cap S \neq \emptyset \). It follows from this that \( S^c \) is open. Therefore \( S \) is closed.

**Definition 4.3.** The closure of a set \( S \) is the set \( \overline{S} = S \cup S' \).

For the set \( S \) of Example 4.2, \( \overline{S} = [0, 1] \). In Example 4.3, \( \overline{T} = \{1/n : n \in \mathbb{Z} \setminus \{0\}\} \cup \{0\} \). According to Theorem 4.4, the closure of any set is a closed set.

**Problem 16.** If \( S \subset \mathbb{R} \), then \( \overline{S} \) is the smallest closed set containing \( S \). (In this case “smallest” means that if \( T \) is any closed set with \( S \subset T \), then \( \overline{S} \subset T \).)

**Theorem 4.5 (Bolzano-Weierstrass Theorem).** A set which is both bounded and infinite has a limit point.
Section 4: The Topology of $\mathbb{R}$

Proof. For the purposes of this proof, if $I = [a, b]$ is a closed interval, let $I^L = [a, (a + b)/2]$ be the closed left half of $I$ and $I^R = [(a + b)/2, b]$ be the closed right half of $I$.

Suppose $S$ is a bounded and infinite set. The assumption that $S$ is bounded implies the existence of an interval $I_1 = [-B, B]$ containing $S$. Since $S$ is infinite, at least one of the two sets $I_1^L \cap S$ or $I_1^R \cap S$ is infinite. Let $I_2$ be either $I_1^L$ or $I_1^R$ such that $I_2 \cap S$ is infinite.

If $I_n$ is such that $I_n \cap S$ is infinite, let $I_{n+1}$ be either $I_n^L$ or $I_n^R$, where $I_{n+1} \cap S$ is infinite.

In this way, a nested sequence of intervals, $I_n$ for $n \in \mathbb{N}$, is defined such that $I_n \cap S$ is infinite for all $n \in \mathbb{N}$ and the length of $I_n$ is $B/2^{n-2} \to 0$. According to the Nested Interval Theorem, there is an $x_0 \in \mathbb{R}$ such that $\bigcap_{n \in \mathbb{N}} I_n = \{x_0\}$.

To see that $x_0$ is a limit point of $S$, let $\epsilon > 0$ and choose $n \in \mathbb{N}$ so that $B/2^{n-2} < \epsilon$. Then $x_0 \in I_n \subset (x_0 - \epsilon, x_0 + \epsilon)$. Since $I_n \cap S$ is infinite, we see $S \cap (x_0 - \epsilon, x_0 + \epsilon) \setminus \{x_0\} \neq \emptyset$. Therefore, $x_0$ is a limit point of $S$. \qed

Using pretty much the same idea, the following can be proved.

Corollary 4.6. Every bounded sequence has a convergent subsequence.

Proof. For the purposes of this proof, if $I = [a, b]$ is a closed interval, let $I^L = [a, (a + b)/2]$ be the closed left half of $I$ and $I^R = [(a + b)/2, b]$ be the closed right half of $I$.

Let $a_n$ be a bounded sequence and choose $B > 0$ such that $\{a_n : n \in \mathbb{N}\} \subset I_1 = [-B, B]$. At least one of the two sets $\{n : a_n \in I^L_1\}$ or $\{n : a_n \in I^R_1\}$ must be infinite. If $\{n : a_n \in I^L_1\}$ is infinite, let $I_2 = I^L_1$. Otherwise, $I_2 = I^R_1$.

Assume that $I_m$ has been chosen for some $n \in \mathbb{N}$ such that $\{n : a_n \in I_m\}$ is infinite. At least one of the two sets $\{n : a_n \in I^L_m\}$ or $\{n : a_n \in I^R_m\}$ must be infinite. If $\{n : a_n \in I^L_m\}$ is infinite, let $I_{m+1} = I^L_m$. Otherwise, $I_{m+1} = I^R_m$.

In this way, a nested sequence of closed intervals, $I_n$, has been inductively defined, where the length of $I_n$ is $B/2^{n-2} \to 0$. An application of the Nested Interval Theorem yields $\{x\} = \bigcap_{n \in \mathbb{N}} I_n$. It suffices to find a subsequence of $a_n$ converging to $x$.

To do this, let $b_1 = a_{m_1}$, where $m_1$ is an arbitrary positive integer. Assuming $b_n = a_{m_n}$ has been chosen, pick $b_{n+1} = a_{m_{n+1}}$ from $I_{n+1}$ so that $m_{n+1} > m_n$. It is possible to do this because $\{n : a_n \in I_{n+1}\}$ is infinite. In this way, a subsequence $b_n$ of $a_n$ has been inductively defined. Since $|b_n - x| \leq B/2^{n-2} \to 0$, it’s clear $b_n \to x$. \qed

Corollary 4.7. If $\{F_n : n \in \mathbb{N}\}$ is a nested collection of nonempty closed and bounded sets, then $\bigcap_{n \in \mathbb{N}} F_n \neq \emptyset$.

Proof. Form a sequence $x_n$ by choosing $x_n \in F_n$ for each $n \in \mathbb{N}$. Since the $F_n$ are nested, $\{x_n : n \in \mathbb{N}\} \subset F_1$, and the boundedness of $F_1$ implies $x_n$ is a bounded sequence. An application of Corollary 4.6 yields a subsequence $y_n$ of $x_n$ such that $y_n \to y$. It suffices to prove $y \in F_n$ for all $n \in \mathbb{N}$.
Section 4: The Topology of $\mathbb{R}$

To do this, fix $n_0 \in \mathbb{N}$. Because $y_n$ is a subsequence of $x_n$ and $x_{n_0} \in F_{n_0}$, it is easy to see $y_n \in F_{n_0}$ for all $n \geq n_0$. Using the fact that $y_n \to y$, we see $y \in F'_{n_0}$. Since $F_{n_0}$ is closed, Theorem 4.4 shows $y \in F_{n_0}$.

\[ \square \]

**Extra Credit 6.** An uncountable subset of $\mathbb{R}$ must have a limit point.
5 Cauchy Sequences

Often the biggest problem with showing that a sequence converges using the techniques we have seen so far is that we must know ahead of time to what it converges. This is often a chicken and egg type problem because to prove a sequence converges, we must seemingly already know it converges. An escape from this dilemma is provided by Cauchy sequences.

Definition 5.1. A sequence \(a_n\) is a Cauchy sequence if for all \(\varepsilon > 0\) there is an \(N \in \mathbb{N}\) such that \(n, m \geq N\) implies \(|a_n - a_m| < \varepsilon\).

Theorem 5.1. A sequence converges iff it is a Cauchy sequence.

Proof. \((\Rightarrow)\) Suppose \(a_n \to L\) and \(\varepsilon > 0\). There is an \(N \in \mathbb{N}\) such that \(g \geq N\) implies \(|a_n - L| < \varepsilon/2\). If \(m, n \geq N\), then

\[
|a_m - a_n| = |a_m - L + L - a_n| \leq |a_m - L| + |L - a_n| < \varepsilon/2 + \varepsilon/2 = \varepsilon.
\]

This shows \(a_n\) is a Cauchy sequence.

\((\Leftarrow)\) Let \(a_n\) be a Cauchy sequence. First, we claim that \(a_n\) is bounded. To see this, let \(\varepsilon = 1\) and choose \(N \in \mathbb{N}\) such that \(n, m \geq N\) implies \(|a_n - a_m| < 1\). In this case, \(a_N - 1 < a_n < a_N + 1\) for all \(n \geq N\), so \(\{a_n : n \geq N\}\) is a bounded set. The set \(\{a_n : n < N\}\), being finite, is also bounded. Since \(\{a_n : n \in \mathbb{N}\}\) is the union of these two bounded sets, it too must be bounded.

Because \(a_n\) is a bounded sequence, Corollary 4.6 implies it has a convergent subsequence \(b_n \to L\). Let \(\varepsilon > 0\) and choose \(N \in \mathbb{N}\) so that \(n, m \geq N\) implies \(|a_n - a_m| < \varepsilon/2\). There is a \(b_k = a_{m_k}\) such that \(m_k \geq N\) and \(|b_{m_k} - L| < \varepsilon/2\). If \(n \geq N\), then

\[
|a_n - L| = |a_n - b_k + b_k - L| \leq |a_n - b_k| + |b_k - L| < |a_n - a_{m_k}| + \varepsilon/2 < \varepsilon/2 + \varepsilon/2 = \varepsilon.
\]

Therefore, \(a_n \to L\). \(\Box\)

According to this theorem, we can prove that a sequence converges without ever knowing precisely to what it converges. An example of the usefulness of this idea is contained in the following definition and theorem.

Definition 5.2. A sequence \(a_n\) is contractive if there is a \(c \in (0, 1)\) such that \(|x_{k+1} - x_k| \leq c|x_k - x_{k-1}|\) for all \(k > 1\).

Theorem 5.2. If a sequence is contractive, then it converges.

Proof. Let \(x_k\) be a contractive sequence with contraction constant \(c \in (0, 1)\).

We first claim that if \(n \in \mathbb{N}\), then

\[
|x_n - x_{n+1}| \leq c^{n-1}|x_1 - x_2|. \tag{2}
\]

This is proved by induction. When \(n = 1\), the statement is \(|x_1 - x_2| \leq c^0|x_1 - x_2| = |x_1 - x_2|\), which is trivially true. Suppose that \(|x_n - x_{n+1}| \leq c^{n-1}|x_1 - x_2|\)
Section 5: Cauchy Sequences

for some \( n \in \mathbb{N} \). Then, from the definition of a contractive sequence and the inductive hypothesis,

\[
|x_{n+1} - x_{n+2}| \leq c|x_n - x_{n+1}| \leq c(c^{n-1}|x_1 - x_2|) = c^n|x_1 - x_2|.
\]

This shows the claim is true in the case \( n + 1 \). Therefore, by induction, the claim is true for all \( n \in \mathbb{N} \).

To show \( x_n \) is a Cauchy sequence, let \( \varepsilon > 0 \). Since \( c^n \to 0 \), we can choose \( N \in \mathbb{N} \) so that

\[
\frac{c^{N+1}}{1 - c} < \frac{\varepsilon}{|x_1 - x_2|}.
\]

Let \( n > m \geq N \). Then

\[
|x_n - x_m| = |x_n - x_{n-1} + x_{n-1} - x_{n-2} + x_{n-2} - \cdots - x_{m+1} + x_{m+1} - x_m| \\
\leq |x_n - x_{n-1}| + |x_{n-1} - x_{n-2}| + \cdots + |x_{m+1} - x_m|
\]

Now, use (2) on each of these terms.

\[
\leq c^{n-2}|x_1 - x_2| + c^{n-3}|x_1 - x_2| + \cdots + c^{m-1}|x_1 - x_2|
\]

\[
= |x_1 - x_2|(c^{n-2} + c^{n-3} + \cdots + c^{m-1})
\]

Apply the formula for a geometric sum.

\[
= |x_1 - x_2|c^{m-1} \frac{1 - c^{n-m}}{1 - c}
\]

\[
< |x_1 - x_2| \frac{c^{m-1}}{1 - c}
\]

Use (3) to estimate the following.

\[
\leq |x_1 - x_2| \frac{c^{N-1}}{1 - c}
\]

\[
< |x_1 - x_2| \frac{\varepsilon}{|x_1 - x_2|}
\]

\[
= \varepsilon
\]

This shows \( x_n \) is a Cauchy sequence.

Example 5.1. Let \(-1 < r < 1\) and define the sequence \( s_n = \sum_{k=0}^{n} r^k \). If \( r = 0 \), the convergent os \( s_n \) is trivial. So, suppose \( r \neq 0 \). In this case

\[
\frac{|s_{n+1} - s_n|}{|s_n - s_{n-1}|} = \frac{r^{n+1}}{r^n} = |r| < 1.
\]

This shows \( s_n \) is contractive, and Theorem 5.2 implies it converges.

Problem 17. If \( x_n \) is a sequence and there is a \( c \geq 1 \) such that \( |x_{k+1} - x_k| > c|x_k - x_{k-1}| \) for all \( k > 1 \), then can \( x_n \) converge?
Section 6: Covering Properties and Compactness on $\mathbb{R}$

**Definition 6.1.** Let $S \subset \mathbb{R}$. A collection of open sets, $\mathcal{O} = \{G_\lambda : \lambda \in \Lambda\}$, is an open cover of $S$ if $S \subset \bigcup_{G \in \mathcal{O}} G$. If $\mathcal{O}' \subset \mathcal{O}$ is also an open cover of $S$, then $\mathcal{O}'$ is an open subcover of $S$ from $\mathcal{O}$.

**Example 6.1.** Let $S = (0, 1)$ and $\mathcal{O} = \{(1/n, 1) : n \in \mathbb{N}\}$. It is easy to see that $\mathcal{O}$ is an open cover of $S$. To prove this, let $x \in (0, 1)$. Choose $n_0 \in \mathbb{N}$ such that $1/n_0 < x$. Then

$$x \in (1/n_0, 1) \subset \bigcup_{n \in \mathbb{N}} (1/n, 1) = \bigcup_{G \in \mathcal{O}} G.$$ 

Since $x$ is an arbitrary element of $(0, 1)$, it follows that $(0, 1) = \bigcup_{G \in \mathcal{O}} G$.

Suppose $\mathcal{O}'$ is any infinite subset of $\mathcal{O}$ and $x \in (0, 1)$. Since $\mathcal{O}'$ is infinite, there exists an $n \in \mathbb{N}$ such that $x \in (1/n, 1) \in \mathcal{O}'$. The rest of the proof proceeds as above.

On the other hand, if $\mathcal{O}'$ is a finite subset of $\mathcal{O}$, then let $M = \max\{n : (1/n, 1) \in \mathcal{O}'\}$. If $0 < x < 1/M$, it is clear that $x \not \in \bigcup_{G \in \mathcal{O}'} G$, so $\mathcal{O}'$ is not an open cover of $(0, 1)$.

**Example 6.2.** Let $T = [0, 1)$ and $0 < \varepsilon < 1$. If $\mathcal{O} = \{(1/n, 1) : n \in \mathbb{N}\} \cup (-\varepsilon, \varepsilon)$. It is easy to see that $\mathcal{O}$ is an open cover of $T$.

It is evident that any open subcover of $T$ from $\mathcal{O}$ must contain $(-\varepsilon, \varepsilon)$, because that is the only element of $\mathcal{O}$ which contains 0. Choose $n \in \mathbb{N}$ such that $1/n < \varepsilon$. Then $\mathcal{O}' = \{(-\varepsilon, \varepsilon), (1/n, 1)\}$ is an open subcover of $T$ from $\mathcal{O}$ which contains only two elements.

**Theorem 6.1 (Lindelöf Property).** If $S \subset \mathbb{R}$ and $\mathcal{O}$ is any open cover of $S$, then $\mathcal{O}$ contains a subcover with a countable number of elements.

**Proof.** Let $\mathcal{O} = \{G_\lambda : \lambda \in \Lambda\}$ be an open cover of $S \subset \mathbb{R}$. Since $\mathcal{O}$ is an open cover of $S$, for each $x \in S$ there is a $\lambda_x \in \Lambda$ and numbers $p_x, q_x \in \mathbb{Q}$ satisfying $x \in (p_x, q_x) \subset G_{\lambda_x} \in \mathcal{O}$. The collection $\mathcal{J} = \{(p_x, q_x) : x \in S\}$ is an open cover of $S$.

Thinking of the collection $\{(p_x, q_x) : x \in S\}$ as a set of ordered pairs of rational numbers, it is seen that $\text{card} (\mathcal{J}) \leq \text{card} (\mathbb{Q} \times \mathbb{Q}) = \aleph_0$, so $\mathcal{J}$ is countable.

For each interval $I \in \mathcal{J}$, choose a $\lambda_I \in \Lambda$ such that $I \subset G_{\lambda_I}$. Then

$$S \subset \bigcup_{I \in \mathcal{J}} I \subset \bigcup_{I \in \mathcal{J}} G_{\lambda_I}$$

shows $\mathcal{O}' = \{G_{\lambda_I} : I \in \mathcal{J}\} \subset \mathcal{O}$ is an open subcover of $S$ from $\mathcal{O}$. Also, $\text{card} (\mathcal{O}') \leq \text{card} (\mathcal{J}) \leq \aleph_0$, so $\mathcal{O}'$ is a countable open subcover of $S$ from $\mathcal{O}$. \[\square\]

**Corollary 6.2.** Any open subset of $\mathbb{R}$ can be written as a countable union of pairwise disjoint open intervals.
Section 6: Covering Properties and Compactness on $\mathbb{R}$

Proof. Let $G$ be open in $\mathbb{R}$. For $x \in G$ let $\alpha_x = \text{glb} \{y : (y, x) \subset G\}$ and $\beta_x = \text{lub} \{(x, y) \subset G\}$. The fact that $G$ is open easily implies $\alpha_x < x < \beta_x$. Define $I_x = (\alpha_x, \beta_x)$.

Then $I_x \subset G$. To see this, suppose $x < w < \beta_x$. Choose $y \in (w, \beta_x)$. The definition of $\beta_x$ guarantees $w \in (x, y) \subset G$. Similarly, if $\alpha_x < w < x$, it follows that $w \in G$.

This shows $\emptyset = \{I_x : x \in G\}$ has the property that $G = \bigcup_{x \in G} I_x$.

Suppose $x, y \in G$ and $I_x \cap I_y \neq \emptyset$. There is no generality lost in assuming $x < y$. In this case, there must be a $w \in (x, y)$ such that $w \in I_x \cap I_y$. We know from above that both $[x, w] \subset G$ and $[w, y] \subset G$, so $[x, y] \subset G$. It follows easily from this that $\alpha_x = \alpha_y < x < y < \beta_x = \beta_y$ and $I_x = I_y$.

From this we conclude $\emptyset$ consists of pairwise disjoint open intervals.

To finish, apply Theorem 6.1 to extract a countable subcover from $\emptyset$.  

Corollary 6.2 can also be proved by a different strategy. Instead of using Theorem 6.1 to extract a countable subcover, we could just choose one rational number from each interval in the cover. The pairwise disjointness of the intervals in the cover guarantee that this will give a bijection between $\emptyset$ and a subset of $\mathbb{Q}$. This method has the advantage of showing that $\emptyset$ itself is countable from the start.

Definition 6.2. An open cover $\mathcal{O}$ of a set $S$ is a finite cover, if $\mathcal{O}$ has only a finite number of elements. The definition of a finite subcover is analogous.

Definition 6.3. A set $K \subset \mathbb{R}$ is compact, if every open cover of $K$ contains a finite subcover.

Theorem 6.3 (Heine-Borel). A set $K \subset \mathbb{R}$ is compact iff it is closed and bounded.

Proof. ($\Rightarrow$) Suppose $K$ is unbounded. The collection $\mathcal{O} = \{(-n, n) : n \in \mathbb{N}\}$ is an open cover of $K$. If $\mathcal{O}'$ is any finite subset of $\mathcal{O}$, then $\bigcup_{G \in \mathcal{O}'} G$ is a bounded set and cannot cover the unbounded set $K$. This shows $K$ cannot be compact, and every compact set must be bounded.

Suppose $K$ is not closed. Then there is a limit point $x$ of $K$ such that $x \notin K$. Define $\mathcal{O} = \{[x - 1/n_i, x + 1/n_i] : n_i \in \mathbb{N}\}$. Then $\mathcal{O}$ is a collection of open sets and $K \subset \bigcup_{G \in \mathcal{O}} G = \mathbb{R} \setminus \{x\}$. Let $\mathcal{O}' = \{[x - 1/n_i, x + 1/n_i] : 1 \leq i \leq N\}$ be a finite subset of $\mathcal{O}$ and $M = \max \{n_i : 1 \leq i \leq N\}$. Since $x$ is a limit point of $K$, there is a $y \in K \cap (x - 1/M, x + 1/M)$. Clearly, $y \notin \bigcup_{G \in \mathcal{O}'} G = [x - 1/M, x + 1/M]\subset$, so $\mathcal{O}'$ cannot cover $K$. This shows every compact set must be closed.

($\Leftarrow$) Let $K$ be closed and bounded and let $\mathcal{O}$ be an open cover of $K$. Applying Theorem 6.1, if necessary, we can assume $\mathcal{O}$ is countable. Thus, $\mathcal{O} = \{G_n : n \in \mathbb{N}\}$.

For each $n \in \mathbb{N}$, define

$$F_n = K \setminus \bigcup_{i=1}^{n} G_i = K \cap \bigcap_{i=1}^{n} G_i^c.$$
Then $F_n$ is a sequence of nested, bounded and closed subsets of $K$. Since $\emptyset$ covers $K$, it follows that
\[ \bigcap_{n \in \mathbb{N}} F_n \subset K \setminus \bigcup_{n \in \mathbb{N}} G_n = \emptyset. \]

According to the Cauchy criterion, the only way this can happen is if $F_n = \emptyset$ for some $n \in \mathbb{N}$. Then $K \subset \bigcup_{i=1}^{n} G_i$, and $\emptyset' = \{G_i : 1 \leq i \leq n\}$ is a finite subcover of $K$ from $\emptyset$.

Compactness shows up in several different, but equivalent ways on $\mathbb{R}$. We’ve already seen most of them, but their equivalence is not obvious. The following theorem shows a few of the most common manifestations of compactness.

**Theorem 6.4.** Let $K \subset \mathbb{R}$. The following statements are equivalent to each other.

(a) $K$ is compact.

(b) $K$ is closed and bounded.

(c) Every infinite subset of $K$ has a limit point.

(d) Every sequence $\{a_n : n \in \mathbb{N}\} \subset K$ has a convergent subsequence.

(e) If $F_n$ is a nested sequence of nonempty relatively closed subsets of $K$, then $\bigcap_{n \in \mathbb{N}} F_n \neq \emptyset$.

**Proof.** (a) $\iff$ (b) is the Heine-Borel Theorem.

That (b) $\implies$ (c) is the Bolzano-Weierstrass Theorem.

(c) $\implies$ (d) is contained in the sequence version of the Bolzano-Weierstrass theorem.

(d) $\implies$ (e) is done the same as the proof of the Cauchy criterion.

To complete the proof, it suffices to show (e) $\implies$ (b). So, suppose $K$ is such that (e) is true.

Let $F_n = K \cap ((-\infty, -n] \cup [n, \infty))$. Then $F_n$ is a sequence of sets which are relatively closed in $K$ such that $\bigcap_{n \in \mathbb{N}} F_n = \emptyset$. If $K$ is unbounded, then $F_n \neq \emptyset$, $\forall n \in \mathbb{N}$, and a contradiction of (e) is evident. Therefore, $K$ must be bounded.

If $K$ is not closed, then there must be a limit point $x$ of $K$ such that $x \notin K$. Define a sequence of relatively closed and nested subsets of $K$ by $F_n = [x - 1/n, x + 1/n] \cap K$ for $n \in \mathbb{N}$. Then $\bigcap_{n \in \mathbb{N}} F_n = \emptyset$, because $x \notin K$. This contradiction of (e) shows that $K$ must be closed.

These various ways of looking at compactness have been given different names by topologists. Property (c) is called *limit point compactness* and (d) is called *sequential compactness*.

**Problem 18.** A closed subset of a compact set is compact.
7 Connectedness

Definition 7.1. A set \( S \subset \mathbb{R} \) is disconnected if there are two open sets \( U \) and \( V \) such that \( U \cap V = \emptyset \), \( U \cap S \neq \emptyset \), \( V \cap S \neq \emptyset \) and \( S \subset U \cup V \). Otherwise, it is connected. Such sets \( U \) and \( V \) are said to be a disconnection of \( S \).

Example 7.1. Let \( S = \{x\} \) be a set containing a single point. \( S \) is connected because there cannot exist nonempty disjoint open sets \( U \) and \( V \) such that \( S \subset U \cup V \). Otherwise, it is disconnected. The sets \( U = (r, s) \subset S \) and \( V = (t, s) \subset S \) are a disconnection of \( S \).

Example 7.2. If \( S = (-1, 0) \cup (0, 1) \), then \( U = (-2, 0) \) and \( V = (0, 2) \) are open sets such that \( U \cap V = \emptyset \), \( U \cap S \neq \emptyset \), \( V \cap S \neq \emptyset \) and \( S \subset U \cup V \). This shows \( S \) is disconnected.

Example 7.3. The sets \( U = (-\infty, \sqrt{2}) \) and \( V = (\sqrt{2}, \infty) \) are open sets such that \( U \cap V = \emptyset \), \( U \cap \mathbb{Q} \neq \emptyset \), \( V \cap \mathbb{Q} \neq \emptyset \) and \( \mathbb{Q} \subset U \cup V = \mathbb{R} \setminus \{\sqrt{2}\} \). This shows \( \mathbb{Q} \) is disconnected.

Theorem 7.1. A nonempty set \( S \subset \mathbb{R} \) is connected iff it is either a single point or an interval.

Proof. (⇒) If \( S \) is not a single point or an interval, there must be numbers \( r < s < t \) such that \( r, t \in S \) and \( s \notin S \). In this case, the sets \( U = (-\infty, s) \) and \( V = (s, \infty) \) are a disconnection of \( S \).

(⇐) It was shown in Example 7.1 that a set containing a single point is connected. So, assume \( S \) is an interval.

Suppose \( S \) is not connected with \( U \) and \( V \) forming a disconnection of \( S \). Choose \( u \in U \cap S \) and \( v \in V \cap S \). There is no generality lost by assuming \( u < v \), so that \( [u, v] \subset S \).

Let \( A = \{x : [u, x] \subset U\} \).

We claim \( A \neq \emptyset \). To see this, use the fact that \( U \) is open to find \( \varepsilon \in (0, v-u) \) such that \( (u-\varepsilon, u+\varepsilon) \subset U \). Then \( u < u + \varepsilon/2 < v \), so \( u + \varepsilon/2 \in A \).

Define \( w = \text{lub} \ A \).

Since \( v \in V \) it is evident \( u < w \leq v \) and \( w \in S \).

If \( w \in U \), then \( u < w < v \) and there is \( \varepsilon \in (0, v-w) \) such that \( (w-\varepsilon, w+\varepsilon) \subset U \) and \( [u, w + \varepsilon] \subset S \) because \( w + \varepsilon < v \). This clearly contradicts the definition of \( w \), so \( w \notin U \).

If \( w \in V \), then there is an \( \varepsilon > 0 \) such that \( (w-\varepsilon, w) \subset V \). In particular, this shows \( w = \text{lub} \ A \leq w - \varepsilon < v \). This contradiction forces the conclusion that \( w \notin V \).

Now, putting all this together, we see \( w \in S \subset U \cup V \) and \( w \notin U \cup V \). This is a clear contradiction, so we're forced to conclude there is no separation of \( S \).

Problem 19. (a) Give an example of a set \( S \) such that \( S \) is disconnected, but \( S \cup \{1\} \) is connected.

(b) Prove that 1 must be a limit point of \( S \).
8 Limits of Functions

Definition 8.1. Let $D \subset \mathbb{R}$, $x_0$ be a limit point of $D$ and $f : D \to \mathbb{R}$. The limit of $f(x)$ at $x_0$ is $L$ if for each $\varepsilon > 0$ there is a $\delta > 0$ such that when $x \in D$ with $0 < |x - x_0| < \delta$, then $|f(x) - L| < \varepsilon$. When this is the case, we write $\lim_{x \to x_0} f(x) = L$.

A useful way of rewording this is to say that $\lim_{x \to x_0} f(x) = L$ iff for every $\varepsilon > 0$ there is a $\delta > 0$ such that $x \in (x_0 - \delta, x_0 + \delta) \cap D \setminus \{x_0\}$ implies $f(x) \in (L - \varepsilon, L + \varepsilon)$.

Example 8.1. If $f(x) = c$ is a constant function and $x_0 \in \mathbb{R}$, then for any positive numbers $\varepsilon$ and $\delta$,
\[
x \in (x_0 - \delta, x_0 + \delta) \cap D \setminus \{x_0\} \Rightarrow |f(x) - c| = |c - c| = 0 < \varepsilon.
\]
This shows the limit of every constant function exists at every point, and the limit is just the value of the function.

Example 8.2. Let $f(x) = x$, $x_0 \in \mathbb{R}$, and $\varepsilon = \delta > 0$. Then
\[
x \in (x_0 - \delta, x_0 + \delta) \cap D \setminus \{x_0\} \Rightarrow |f(x) - f(x_0)| = |x - x_0| < \delta = \varepsilon.
\]
This shows that the identity function has a limit at every point and its limit is just the value of the function.

Example 8.3. Let $f(x) = \frac{2x^2 - 8}{x - 2}$. In this case, the implied domain of $f$ is $D = \mathbb{R} \setminus \{2\}$. We claim that $\lim_{x \to 2} f(x) = 8$.

To see this, let $\varepsilon > 0$ and choose $\delta \in (0, \varepsilon/2)$. If $0 < |x - 2| < \delta$, then
\[
|f(x) - 8| = \left| \frac{2x^2 - 8}{x - 2} - 8 \right| = \left| 2x + 2 - 8 \right| = 2|x - 2| < \varepsilon.
\]

Example 8.4. Let $f(x) = \sqrt{x + 1}$. Then the implied domain of $f$ is $D = [-1, \infty)$. We claim that $\lim_{x \to -1} f(x) = 0$.

To see this, let $\varepsilon > 0$ and choose $\delta \in (0, \varepsilon^2)$. If $0 < x - (-1) = x + 1 < \delta$, then
\[
|f(x) - 0| = \sqrt{x + 1} < \sqrt{\delta} < \sqrt{\varepsilon^2} = \varepsilon.
\]

There is an obvious similarity between the definition of limit of a sequence and limit of a function. The following theorem makes this similarity explicit, and gives another way to prove facts about limits of functions.

Theorem 8.1. Let $f : D \to \mathbb{R}$ and $x_0$ be a limit point of $D$. $\lim_{x \to x_0} f(x) = L$ iff whenever $x_n$ is a sequence from $D \setminus \{x_0\}$ such that $x_n \to x_0$, then $f(x_n) \to L$.

Proof. ($\Rightarrow$) Suppose $\lim_{x \to x_0} f(x) = L$ and $x_n$ is a sequence from $D \setminus \{x_0\}$ such that $x_n \to x_0$. Let $\varepsilon > 0$. There exists a $\delta > 0$ such that $|f(x) - L| < \varepsilon$
whenever \( x \in (x - \delta, x + \delta) \cap D \setminus \{x_0\} \). Since \( x_n \to x_0 \), there is an \( N \in \mathbb{N} \) such that \( n \geq N \) implies \( 0 < |x_n - x_0| < \delta \). In this case, \( |f(x_n) - L| < \varepsilon \). This shows \( f(x_n) \to x_0 \).

\((\Leftarrow)\) Suppose that whenever \( x_n \) is a sequence from \( D \setminus \{x_0\} \) such that \( x_n \to x_0 \), then \( f(x_n) \to L \), but \( \lim_{x \to x_0} f(x) \neq L \). Then there exists an \( \varepsilon > 0 \) such that for all \( \delta > 0 \) there is an \( x \in (x_0 - \delta, x_0 + \delta) \cap D \setminus \{x_0\} \) such that \( |f(x) - L| \geq \varepsilon \). In particular, for each \( n \in \mathbb{N} \), there must exist \( x_n \in (x_0 - 1/n, x_0 + 1/n) \cap D \setminus \{x_0\} \) such that \( |f(x_n) - L| \geq \varepsilon \). Since \( x_n \to x_0 \), this is a contradiction. Therefore, \( \lim_{x \to x_0} f(x) = L \). □

**Example 8.5.** Let \( f(x) = \sin(1/x) \), \( a_n = \frac{1}{n\pi} \) and \( b_n = \frac{1}{(2n-1)\pi} \). Then \( a_n \downarrow 0 \), \( b_n \downarrow 0 \), \( f(a_n) = 0 \) and \( f(b_n) = 1 \) for all \( n \in \mathbb{N} \). An application of Theorem 8.1 shows \( \lim_{x \to 0} f(x) \) does not exist.
Section 8: Limits of Functions

Figure 6: This is the function from Example 8.6. The graph shown here is on the interval \([0, 0.5]\). There are an infinite number of oscillations from \(-x\) to \(x\) on any open interval containing the origin.

**Theorem 8.2 (Squeeze Theorem).** Suppose \(f\), \(g\) and \(h\) are all functions defined on \(D \subset \mathbb{R}\) with \(f(x) \leq g(x) \leq h(x)\) for all \(x \in D\). If \(x_0\) is a limit point of \(D\) and \(\lim_{x \to x_0} f(x) = \lim_{x \to x_0} h(x) = L\), then \(\lim_{x \to x_0} g(x) = L\).

*Proof.* Let \(x_n\) be any sequence from \(D \setminus \{x_0\}\) such that \(x_n \to x_0\). According to Theorem 8.1, both \(f(x_n) \to L\) and \(h(x_n) \to L\). Since \(f(x_n) \leq g(x_n) \leq h(x_n)\), an application of the sandwich theorem for sequences shows \(g(x_n) \to L\). Now, another use of Theorem 8.1 shows \(\lim_{x \to x_0} g(x) = L\). \(\square\)

**Example 8.6.** Let \(f(x) = x \sin(1/x)\). Since \(-1 \leq \sin(1/x) \leq 1\) when \(x \neq 0\), we see that \(-x \leq \sin(1/x) \leq x\) for \(x \neq 0\). Since \(\lim_{x \to 0} x = \lim_{x \to 0} -x = 0\), Theorem 8.2 implies \(\lim_{x \to 0} x \sin(1/x) = 0\). See Figure 6

**Theorem 8.3.** Suppose \(f: D \to \mathbb{R}\) and \(g: D \to \mathbb{R}\) and \(x_0\) is a limit point of \(D\). If \(\lim_{x \to x_0} f(x) = L\) and \(\lim_{x \to x_0} g(x) = M\), then

(a) \(\lim_{x \to x_0} (f + g)(x) = L + M\),

(b) \(\lim_{x \to x_0} (af)(x) = aL, \forall x \in \mathbb{R}\),

(c) \(\lim_{x \to x_0} (fg)(x) = LM\), and

(d) \(\lim_{x \to x_0} (1/f)(x) = 1/L\), as long as \(L \neq 0\).

*Proof.* Suppose \(a_n\) is a sequence from \(D \setminus \{x_0\}\) converging to \(x_0\). Then Theorem 8.1 implies \(f(a_n) \to L\) and \(g(a_n) \to M\). (a)-(d) follow at once from the corresponding properties for sequences. \(\square\)
Example 8.7. Let \( f(x) = 3x + 2 \). If \( g_1(x) = 3, g_2(x) = x \) and \( g_3(x) = 2 \), then 
\[
 f(x) = g_1(x)g_2(x) + g_3(x).
\]
Examples 8.1 and 8.2 along with parts (a) and (c) of Theorem 8.3 immediately show that for every \( x \in \mathbb{R} \), \( \lim_{x \to x_0} f(x) = f(x_0) \).

In the same manner as Example 8.7, it can be shown for every rational function \( f(x) \), that \( \lim_{x \to x_0} f(x) = f(x_0) \) whenever \( f(x_0) \) exists.

**Extra Credit 7.** If \( \mathbb{Q} = \{q_n : n \in \mathbb{N}\} \) is an enumeration of the rational numbers and

\[
 f(x) = \begin{cases} 
 1/n, & x = q_n \\
 0, & x \in \mathbb{Q}^c 
\end{cases}
\]

then \( \lim_{x \to a} f(x) = 0 \), for all \( a \in \mathbb{Q}^c \).
9 Unilateral Limits

Definition 9.1. Let \( f : D \rightarrow \mathbb{R} \) and \( x_0 \) be a limit point of \( D \cap (-\infty, x_0) \). \( f \) has \( L \) as its left-hand limit at \( x_0 \) if for all \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that \( f((x_0 - \delta, x_0) \cap D) \subset (L - \varepsilon, L + \varepsilon) \). In this case, we write \( \lim_{x \to x_0^-} f(x) = L \).

Let \( f : D \rightarrow \mathbb{R} \) and \( x_0 \) be a limit point of \( D \cap (x_0, \infty) \). \( f \) has \( L \) as its right-hand limit at \( x_0 \) if for all \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that \( f((x_0, x_0 + \delta) \cap D) \subset (L - \varepsilon, L + \varepsilon) \). In this case, we write \( \lim_{x \to x_0^+} f(x) = L \).

Another standard notation for the unilateral limits is \( \lim_{x \to x_0^-} f(x) = \lim_{x \to x_0^-} f(x) \) and \( \lim_{x \to x_0^+} f(x) = \lim_{x \to x_0^+} f(x) \).

Example 9.1. Let \( f(x) = |x|/x \). Then \( \lim_{x \to 0^-} f(x) = 1 \) and \( \lim_{x \to 0^+} f(x) = -1 \). (See Figure 7.)

Theorem 9.1. Let \( f : D \rightarrow \mathbb{R} \) and \( x_0 \) be a limit point of \( D \).

\[
\lim_{x \to x_0^-} f(x) = L \quad \iff \quad \lim_{x \to x_0^+} f(x) = L = \lim_{x \to x_0} f(x)
\]

Proof. This proof is left as an exercise.

Theorem 9.2. If \( f : (a, b) \rightarrow \mathbb{R} \) is monotone, then both unilateral limits of \( f \) exist at every point of \((a, b)\).

Proof. To be specific, suppose \( f \) is increasing and \( x_0 \in (a, b) \). Let \( \varepsilon > 0 \) and \( L = \text{lub}\{f(x) : a < x < x_0\} \). According to Corollary 11, there must exist an \( x \in (a, x_0) \) such that \( L - \varepsilon < f(x) \leq L \). Define \( \delta = x_0 - x \). If \( y \in (x_0 - \delta, x_0) \), then \( L - \varepsilon = f(x) < f(y) \leq L \). This shows \( \lim_{x \to x_0^-} f(x) = L \).

The proof that \( \lim_{x \to x_0^+} f(x) \) exists is similar.

To handle the case when \( f \) is decreasing, consider \(-f\) instead of \( f \).
10 Continuity

Definition 10.1. Let \( f : D \rightarrow \mathbb{R} \) and \( x_0 \in D \). \( f \) is continuous at \( x_0 \) if for every \( \varepsilon > 0 \) there exists a \( \delta > 0 \) such that when \( x \in D \) with \( |x - x_0| < \delta \), then \( |f(x) - f(x_0)| < \varepsilon \). The set of all points at which \( f \) is continuous is denoted \( C(f) \).

Several useful ways of rephrasing this are contained in the following theorem, the proof of which is left to the reader.

Theorem 10.1. Let \( f : D \rightarrow \mathbb{R} \) and \( x_0 \in D \). The following statements are equivalent.

(a) \( x_0 \in C(f) \),

(b) For all \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that \( x \in (x_0 - \delta, x_0 + \delta) \cap D \Rightarrow f(x) \in (f(x_0) - \varepsilon, f(x_0) + \varepsilon) \), and

(c) For all \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that \( f((x_0 - \delta, x_0 + \delta)) \subset (f(x_0) - \varepsilon, f(x_0) + \varepsilon) \).

Example 10.1. Define

\[
f(x) = \begin{cases} 
\frac{2x^2 - 8}{x - 2}, & x \neq 2 \\
8, & x = 2
\end{cases}
\]

It follows easily from Example 8.3 that \( 2 \in C(f) \).

There is a subtle difference between the treatment of the domain of the function between the definitions of limit and continuity. In the definition of limit, the “target point,” \( x_0 \) is required to be a limit point of the domain. There is no such stipulation in the definition of continuity. To see a consequence of this difference, consider the following example.

Example 10.2. If \( f : \mathbb{Z} \rightarrow \mathbb{R} \) is an arbitrary function, then \( C(f) = \mathbb{Z} \). To see this, let \( n_0 \in \mathbb{Z} \), \( \varepsilon > 0 \) and \( \delta = 1 \). If \( x \in \mathbb{Z} \) with \( |x - n_0| < \delta \), then \( x = n_0 \). It’s now obvious that \( |f(x) - f(n_0)| = 0 < \varepsilon \), so \( f \) is continuous at \( n_0 \).

This leads to the following theorem.

Theorem 10.2. Let \( f : D \rightarrow \mathbb{R} \) and \( x_0 \in D \). If \( x_0 \) is an isolated point of \( D \), then \( x_0 \in C(f) \). If \( x_0 \) is a limit point of \( D \), then \( x_0 \in C(f) \) iff \( \lim_{x \to x_0} f(x) = f(x_0) \).

Proof. If \( x_0 \) is isolated in \( D \), then there is an \( \delta > 0 \) such that \( (x_0 - \delta, x_0 + \delta) \cap D = \{x_0\} \). For any \( \varepsilon > 0 \), the definition of continuity is satisfied with this \( \delta \).

Next, suppose \( x_0 \) is a limit point of \( D \).

The definition of continuity says that \( f \) is continuous at \( x_0 \) iff for all \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that when \( x \in (x_0 - \delta, x_0 + \delta) \cap D \), then \( f(x) \in (f(x_0) - \varepsilon, f(x_0) + \varepsilon) \).
The definition of limit says that \( \lim_{x \to x_0} f(x) = f(x_0) \) iff for all \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that when \( x \in (x_0 - \delta, x_0 + \delta) \cap D \setminus \{x_0\} \), then \( f(x) \in (f(x_0) - \varepsilon, f(x_0) + \varepsilon) \).

Comparing these two definitions, it is clear that \( x_0 \in C(f) \) implies

\[
\lim_{x \to x_0} f(x) = f(x_0).
\]

On the other hand, suppose \( \lim_{x \to x_0} f(x) = f(x_0) \) and \( \varepsilon > 0 \). Choose \( \delta \) according to the definition of limit. When \( x \in (x_0 - \delta, x_0 + \delta) \cap D \setminus \{x_0\} \), then \( f(x) \in (f(x_0) - \varepsilon, f(x_0) + \varepsilon) \). It is easy to see from this that when \( x = x_0 \), then \( f(x) - f(x_0) = f(x_0) - f(x_0) = 0 < \varepsilon \). Therefore, when \( x \in (x_0 - \delta, x_0 + \delta) \cap D \), then \( f(x) \in (f(x_0) - \varepsilon, f(x_0) + \varepsilon) \), and \( x_0 \in C(f) \), as desired.

**Example 10.3.** If \( f(x) = c \), for some \( c \in \mathbb{R} \), then Example 8.1 and Theorem 10.2 show that \( f \) is continuous at every point.

**Example 10.4.** If \( f(x) = x \), then Example 8.2 and Theorem 10.2 show that \( f \) is continuous at every point.

**Corollary 10.3.** Let \( f : D \to \mathbb{R} \) and \( x_0 \in D \). \( x_0 \in C(f) \) iff whenever \( x_n \) is a sequence from \( D \) with \( x_n \to x_0 \), then \( f(x_n) \to f(x_0) \).

**Proof.** Combining Theorem 10.2 with Theorem 8.1 shows this to be true.

**Example 10.5.** Suppose

\[
f(x) = \begin{cases} 
1, & x \in \mathbb{Q} \\
0, & x \notin \mathbb{Q}. 
\end{cases}
\]

For each \( x \in \mathbb{Q} \), there is a sequence of irrational numbers converging to \( x \), and for each \( y \in \mathbb{Q}^c \) there is a sequence of rational numbers converging to \( y \). Corollary 10.3 shows \( C(f) = \emptyset \).

**Example 10.6 (Salt and Pepper Function).** Since \( \mathbb{Q} \) is a countable set, it can be written as a sequence, \( \mathbb{Q} = \{q_n : n \in \mathbb{N}\} \). Define

\[
f(x) = \begin{cases} 
1/n, & x = q_n, \\
0, & x \in \mathbb{Q}^c. 
\end{cases}
\]

If \( x \in \mathbb{Q} \), then \( x = q_n \), for some \( n \) and \( f(x) = 1/n > 0 \). There is a sequence \( x_n \) from \( \mathbb{Q}^c \) such that \( x_n \to x \) and \( f(x_n) = 0 \not\to f(x) = 1/n \). Therefore \( C(f) \cap \mathbb{Q} = \emptyset \).

On the other hand, let \( x \in \mathbb{Q}^c \) and \( \varepsilon > 0 \). Choose \( N \in \mathbb{N} \) large enough so that \( 1/N < \varepsilon \) and let \( \delta = \min\{|x - q_n| : 1 \leq n \leq N\} \). If \( |x - y| < \delta \), there are two cases to consider. If \( y \notin \mathbb{Q} \), then \( |f(y) - f(x)| = |0 - 0| = 0 < \varepsilon \). If \( y \in \mathbb{Q} \), then the choice of \( \delta \) guarantees \( y = q_n \) for some \( n > N \). In this case, \( |f(y) - f(x)| = f(y) = f(q_n) = 1/n < 1/N < \varepsilon \). Therefore, \( x \in C(f) \).

This shows that \( C(f) = \mathbb{Q}^c \).
It is a consequence of an advanced result known as the Baire category theorem that there is no function $f$ such that $\mathcal{C}(f) = \mathbb{Q}$.

The following theorem is an almost immediate consequence of Theorem 8.3.

**Theorem 10.4.** Let $f : D_f \to \mathbb{R}$ and $g : D_g \to \mathbb{R}$. If $x_0 \in \mathcal{C}(f) \cap \mathcal{C}(g)$, then

(a) $x_0 \in \mathcal{C}(f + g)$,

(b) $x_0 \in \mathcal{C}(\alpha f)$, $\forall \alpha \in \mathbb{R}$,

(c) $x_0 \in \mathcal{C}(fg)$, and

(d) $x_0 \in \mathcal{C}(f/g)$ when $g(x_0) \neq 0$.

**Corollary 10.5.** If $f$ is a rational function, then $f$ is continuous at each point of its domain.

**Proof.** This is a consequence of Examples 10.3 and 10.4 used with Theorem 10.4.

**Theorem 10.6.** Suppose $f : D_f \to \mathbb{R}$ and $g : D_g \to \mathbb{R}$ such that $f(D_f) \subset D_g$. If there is an $x_0 \in \mathcal{C}(f)$ such that $f(x_0) \in \mathcal{C}(g)$, then $x_0 \in \mathcal{C}(g \circ f)$.

**Proof.** Let $\varepsilon > 0$ and choose $\delta_1 > 0$ such that $g((f(x_0) - \delta_1, f(x_0) + \delta_1) \cap D_g) \subset (g \circ f(x_0) - \varepsilon, g \circ f(x_0) + \varepsilon)$. Choose $\delta_2 > 0$ such that $f((x_0 - \delta_2, x_0 + \delta_2) \cap D_f) \subset (f(x_0) - \delta_1, f(x_0) + \delta_1)$. Then

$$g \circ f((x_0 - \delta_2, x_0 + \delta_2) \cap D_f) \subset g((f(x_0) - \delta_1, f(x_0) + \delta_1) \cap D_g) \subset (g \circ f(x_0) - \delta_2, g \circ f(x_0) + \delta_2) \cap D_f).$$

Since this shows Theorem 10.1(c) is satisfied at $x_0$ with the function $g \circ f$, it follows that $x_0 \in \mathcal{C}(g \circ f)$.

**Problem 20.** Prove that $f(x) = \sqrt{x}$ is continuous on $[0, \infty)$.

**Example 10.7.** If $f$ is as in Problem 20, then Theorem 10.6 shows $f \circ f = \sqrt[4]{x}$ is continuous on $[0, \infty)$.

In the same way, it can be shown by induction that $f(x) = x^{m/2^n}$ is continuous on $[0, \infty)$ for all $m, n \in \mathbb{Z}$. 
11 Unilateral Continuity

Definition 11.1. Let \( f : D \to \mathbb{R} \) and \( x_0 \in D \). \( f \) is left-continuous at \( x_0 \) if for every \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that \( f((x_0 - \delta, x_0] \cap D) \subset (f(x_0) - \varepsilon, f(x_0) + \varepsilon) \).

Let \( f : D \to \mathbb{R} \) and \( x_0 \in D \). \( f \) is right-continuous at \( x_0 \) if for every \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that \( f([x_0, x_0 + \delta) \cap D) \subset (f(x_0) - \varepsilon, f(x_0) + \varepsilon) \).

Theorem 11.1. Let \( f : D \to \mathbb{R} \) and \( x_0 \in D \). \( x_0 \in C(f) \) iff \( f \) is both right and left-continuous at \( x_0 \).

Proof. The proof of this theorem is left as an exercise. \( \square \)

According to Theorem 9.1, when \( f \) is monotone on an interval \((a, b)\), the unilateral limits of \( f \) exist at every point. In order for such a function to be continuous at \( x_0 \in (a, b) \), it must be the case that

\[
\lim_{x \searrow x_0} f(x) = f(x_0) = \lim_{x \nearrow x_0} f(x).
\]

If either of the two inequalities is violated, the function is not continuous at \( x_0 \).

In the case when \( \lim_{x \searrow x_0} f(x) \neq \lim_{x \nearrow x_0} f(x) \), it is said that a jump discontinuity occurs at \( x_0 \).

Example 11.1. The function

\[
f(x) = \begin{cases} 
|x|/x, & x \neq 0 \\
0, & x = 0
\end{cases}
\]

has a jump discontinuity at \( x = 0 \).

In the case when \( \lim_{x \searrow x_0} f(x) = \lim_{x \nearrow x_0} f(x) \neq f(x_0) \), it is said that \( f \) has a removable discontinuity at \( x_0 \). The discontinuity is called “removable” because in this case, the function can be made continuous at \( x_0 \) merely by redefining its value at the single point, \( x_0 \), to be the value of the two one-sided limits.

Example 11.2. The function \( f(x) = \frac{x^2 - 4}{x - 2} \) is not continuous at \( x = 2 \) because 2 is not in the domain of \( f \). Since \( \lim_{x \to 2} f(x) = 4 \), if the domain of \( f \) is extended by setting \( f(2) = 4 \), then this extended \( f \) is continuous everywhere. (See Figure 8.)

Theorem 11.2. If \( f : (a, b) \to \mathbb{R} \) is monotone, then \((a, b) \setminus C(f)\) is countable.

Proof. In light of the discussion above and Theorem 9.1, it is apparent that the only types of discontinuities \( f \) can have are jump discontinuities.

To be specific, suppose \( f \) is increasing and \( x_0, y_0 \in (a, b) \setminus C(f) \) with \( x_0 < y_0 \). In this case, the fact that \( f \) is increasing implies

\[
\lim_{x \searrow x_0} f(x) < \lim_{x \searrow y_0} f(x) \leq \lim_{x \to y_0} f(x) < \lim_{x \nearrow y_0} f(x).
\]

This implies that for any two \( x_0, y_0 \in (a, b) \setminus C(f) \), there are disjoint open intervals, \( I_{x_0} = (\lim_{x \searrow x_0} f(x), \lim_{x \nearrow x_0} f(x)) \) and \( I_{y_0} = (\lim_{x \searrow y_0} f(x), \lim_{x \nearrow y_0} f(x)) \).
Figure 8: The function from Example 11.2. Note that the graph is a line with one “hole” in it. Plugging up the hole removes the discontinuity.

For each $x \in (a, b) \setminus C(f)$, choose $q_x \in I_x \cap \mathbb{Q}$. Because of the pairwise disjointness of the intervals $\{I_x : x \in (a, b) \setminus C(f)\}$, this defines an bijection between $(a, b) \setminus C(f)$ and a subset of $\mathbb{Q}$. Therefore, $(a, b) \setminus C(f)$ must be countable.

A similar argument holds for a decreasing function.

Theorem 11.2 implies that a monotone function is continuous at “nearly every” point in its domain. Characterizing the points of discontinuity as countable is the best that can be hoped for. To see this, let $D = \{d_n : n \in \mathbb{N}\}$ be a countable set and define $J_x = \{n : d_n < x\}$. Using this, we define

$$f(x) = \sum_{n \in J_x} \frac{1}{2^n}. \quad (4)$$

**Extra Credit 7.** If $f$ is defined as in (4), then $D = C(f)^c$.

**Problem 21.** If $f : \mathbb{R} \to \mathbb{R}$ is monotone, then there is a countable set $D$ such that the values of $f$ can be altered on $D$ in such a way that the altered function is left-continuous at every point of $\mathbb{R}$.
12 Continuous Functions

Definition 12.1. Let \( f : D \to \mathbb{R} \) and \( A \subset D \). We say \( f \) is continuous on \( A \) if \( A \subset C(f) \). If \( D = C(f) \), then \( f \) is continuous.

Theorem 12.1. \( f : D \to \mathbb{R} \) is continuous iff whenever \( G \) is open in \( \mathbb{R} \), then \( f^{-1}(G) \) is relatively open in \( D \).

Proof. (\( \Rightarrow \)) Assume \( f \) is continuous on \( D \) and let \( G \) be open in \( \mathbb{R} \). Let \( x \in f^{-1}(G) \) and choose \( \varepsilon > 0 \) such that \( (f(x) - \varepsilon, f(x) + \varepsilon) \subset G \). Using the continuity of \( f \) at \( x \), we can find a \( \delta > 0 \) such that \( f((x - \delta, x + \delta) \cap D) \subset G \). This implies at once that \( (x - \delta, x + \delta) \cap D \subset f^{-1}(G) \). Because \( x \) was an arbitrary element of \( f^{-1}(G) \), it follows that \( f^{-1}(G) \) is open.

(\( \Leftarrow \)) Choose \( x \in D \) and let \( \varepsilon > 0 \). By assumption, the set \( f^{-1}((f(x) - \varepsilon, f(x) + \varepsilon)) \) is relatively open in \( D \). This implies the existence of a \( \delta > 0 \) such that \( (x - \delta, x + \delta) \cap D \subset f^{-1}((f(x) - \varepsilon, f(x) + \varepsilon)) \). It follows at once from this that \( f((x - \delta, x + \delta) \cap D) \subset (f(x) - \varepsilon, f(x) + \varepsilon) \), and \( x \in C(f) \).

Theorem 12.2. If \( f \) is continuous on a compact set \( K \), then \( f(K) \) is compact.

Proof. Let \( \emptyset \) be an open cover of \( f(K) \) and \( \mathcal{J} = \{ f^{-1}(G) : G \in \emptyset \} \). By Theorem 12.1, \( \mathcal{J} \) is a collection of sets which are relatively open in \( K \). Since \( \mathcal{J} \) covers \( f(K) \), its easy to see, \( \mathcal{J} \) is an open cover of \( K \). Using the fact that \( K \) is compact, we can choose a finite subcover of \( K \) from \( \mathcal{J} \), say \( \{ G_1, G_2, \ldots, G_n \} \). There are \( \{ H_1, H_2, \ldots, H_n \} \subset \emptyset \) such that \( f^{-1}(H_k) = G_k \) for \( 1 \leq k \leq n \). Then

\[
f(K) \subset f \left( \bigcup_{1 \leq k \leq n} G_k \right) = \bigcup_{1 \leq k \leq n} H_k.
\]

Thus, \( \{ H_1, H_2, \ldots, H_3 \} \) is a subcover of \( f(K) \) from \( \emptyset \).

Corollary 12.3. If \( f : K \to \mathbb{R} \) is continuous and \( K \) is compact, then \( f \) is bounded.

Proof. By Theorem 12.2, \( f(K) \) is compact. Now, use the Bolzano-Weierstrass theorem to conclude \( f \) is bounded.

Corollary 12.4. If \( f : K \to \mathbb{R} \) is continuous and \( K \) is compact, then there are \( m, M \in K \) such that \( f(m) \leq f(x) \leq f(M) \) for all \( x \in K \).

Proof. According to Theorem 12.2 and the Bolzano-Weierstrass theorem, \( f(K) \) is closed and bounded. Because of this, \( \text{glb} f(K) \in f(K) \) and \( \text{lub} f(K) \in f(K) \). It suffices to choose \( m \in f^{-1}(\text{glb} f(K)) \) and \( M \in f^{-1}(\text{lub} f(K)) \).

Theorem 12.5. If \( f : K \to \mathbb{R} \) is continuous and invertible and \( K \) is compact, then \( f^{-1} : f(K) \to K \) is continuous.
Proof. Let $G$ be open in $K$. According to Theorem 12.1, it suffices to show $f(G)$ is open in $f(K)$.

To do this, note that $K \setminus G$ is compact, so by Theorem 12.2, $f(K \setminus G)$ is compact, and therefore closed. Because $f$ is injective, $f(G) = f(K) \setminus f(K \setminus G)$. This shows $f(G)$ is open in $f(K)$. □

**Theorem 12.6.** If $f$ is continuous on a connected set $K$, then $f(K)$ is connected.

**Proof.** If $f(K)$ is not connected, there must exist two disjoint open sets, $U$ and $V$, such that $f(K) \subset U \cup V$ and $f(K) \cap U \neq \emptyset \neq f(K) \cap V$. In this case, Theorem 12.1 implies $f^{-1}(U)$ and $f^{-1}(V)$ are both open. They are clearly disjoint and $f^{-1}(U) \cap K \neq \emptyset \neq f^{-1}(V) \cap K$. But, this implies $f^{-1}(U)$ and $f^{-1}(V)$ disconnect $K$, which is a contradiction. Therefore, $f(K)$ is connected. □

**Corollary 12.7.** If $f : [a, b] \to \mathbb{R}$ is continuous and $\alpha$ is between $f(a)$ and $f(b)$, then there is a $c \in [a, b]$ such that $f(c) = \alpha$.

**Proof.** This is an easy consequence of Theorem 12.6 and Theorem 7.1. □

**Definition 12.2.** A function $f : D \to \mathbb{R}$ has the Darboux property if whenever $a, b \in D$ and $\gamma$ is between $f(a)$ and $f(b)$, then there is a $c$ between $a$ and $b$ such that $f(c) = \gamma$.

The Darboux property is also often called the intermediate value property. Corollary 12.7 shows that a function continuous on an interval has the Darboux property. The next example shows continuity is not necessary for the Darboux property to hold.

**Example 12.1.** The function

$$f(x) = \begin{cases} \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

is not continuous, but does have the Darboux property. (See Figure 5.) It can be seen from Example 8.5 that $0 \notin C(f)$.

To see $f$ has the Darboux property, choose two numbers $a < b$.

If $a > 0$ or $b < 0$, then $f$ is continuous on $[a, b]$ and Corollary 12.7 suffices to finish the proof.

On the other hand, if $0 \in [a, b]$, then there must exist an $n \in \mathbb{Z}$ such that both $\frac{4}{(4n+1)\pi}, \frac{4}{(4n+3)\pi} \in [a, b]$. Since $f\left(\frac{4}{(4n+1)\pi}\right) = 1$, $f\left(\frac{4}{(4n+3)\pi}\right) = -1$ and $f$ is continuous on the interval between them, we see $f([a, b]) = [-1, 1]$, which is the entire range of $f$. The claim now follows easily.

**Problem 22.** Let $f$ and $g$ be two functions which are continuous on a set $D \subseteq \mathbb{R}$. Prove or give a counter example: $\{x \in D : f(x) > g(x)\}$ is open.
Problem 23. If $f : [a, b] \to \mathbb{R}$ is continuous, not constant,

$$m = \text{glb} \{ f(x) : a \leq x \leq b \} \quad \text{and} \quad M = \text{lub} \{ f(x) : a \leq x \leq b \},$$

then $f([a, b]) = [m, M]$.

Extra Credit 8. If $F \subset \mathbb{R}$ is closed, then there is an $f : \mathbb{R} \to \mathbb{R}$ such that $F = C(f)^c$. 
13 Uniform Continuity

Most of the ideas contained in this section will not be needed until we begin developing the theory of integration.

**Definition 13.1.** A function \( f : D \to \mathbb{R} \) is uniformly continuous if for all \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that when \( x, y \in D \) with \( |x - y| < \delta \), then \( |f(x) - f(y)| < \varepsilon \).

The idea here is that in the ordinary definition of continuity, the \( \delta \) in the definition depends on both the \( \varepsilon \) and the \( x \) at which continuity is being tested. With uniform continuity, \( \delta \) only depends on \( \varepsilon \); i.e., the same \( \delta \) works uniformly across the whole domain.

**Theorem 13.1.** If \( f : D \to \mathbb{R} \) is uniformly continuous, then it is continuous.

**Proof.** This proof is left as an exercise. \( \square \)

**Example 13.1.** Let \( f(x) = 1/x \) on \( D = (0,1) \) and \( \varepsilon > 0 \). It’s clear that \( f \) is continuous on \( D \). Let \( \delta > 0 \) and choose \( m, n \in \mathbb{N} \) such that \( m > 1/\delta \) and \( n - m > \varepsilon \). If \( x = 1/m \) and \( y = 1/n \), then \( 0 < y < x < \delta \) and \( f(y) - f(x) = n - m > \varepsilon \). Therefore, \( f \) is not uniformly continuous.

**Theorem 13.2.** If \( f : D \to \mathbb{R} \) is continuous and \( D \) is compact, then \( f \) is uniformly continuous.

**Proof.** Suppose \( f \) is not uniformly continuous. Then there is an \( \varepsilon > 0 \) such that for every \( n \in \mathbb{N} \) there are \( x_n, y_n \in D \) with \( |x_n - y_n| < 1/n \) and \( |f(x_n) - f(y_n)| \geq \varepsilon \). An application of the Bolzano-Weierstrass theorem yields a subsequence \( x_{n_k} \) of \( x_n \) such that \( x_{n_k} \to x_0 \in D \).

Since \( f \) is continuous at \( x_0 \), there is a \( \delta > 0 \) such that whenever \( x \in (x_0 - \delta, x_0 + \delta) \cap D \), then \( |f(x) - f(x_0)| < \varepsilon/2 \). Choose \( n_k \in \mathbb{N} \) such that \( 1/n_k < \delta/2 \) and \( x_{n_k} \in (x_0 - \delta/2, x_0 + \delta/2) \). Then both \( x_{n_k}, y_{n_k} \in (x_0 - \delta, x_0 + \delta) \) and

\[
\varepsilon \leq |f(x_{n_k}) - f(y_{n_k})| = |f(x_{n_k}) - f(x_0) + f(x_0) - f(y_{n_k})| \\
\leq |f(x_{n_k}) - f(x_0)| + |f(x_0) - f(y_{n_k})| < \varepsilon/2 + \varepsilon/2 = \varepsilon,
\]

which is a contradiction.

Therefore, \( f \) must be uniformly continuous. \( \square \)

The following corollary is an immediate consequence of Theorem 13.2.

**Corollary 13.3.** If \( f : [a, b] \to \mathbb{R} \) is continuous, then \( f \) is uniformly continuous.

**Problem 24.** Prove that an unbounded function on a bounded open interval cannot be uniformly continuous.

**Problem 25.** Prove Theorem 13.1.
14 Differentiation

Definition 14.1. Let \( f \) be a function on a neighborhood of \( x_0 \). \( f \) is differentiable at \( x_0 \) with value \( f'(x) \) if

\[
f'(x) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.
\]

Define \( D(f) = \{ x : f'(x) \text{ exists} \} \).

The standard notations for the derivative will be used; e.g., \( f'(x) \), \( \frac{df(x)}{dx} \), \( Df(x) \), etc.

Another way of stating this definition is to note that if \( x_0 \in D(f) \), then

\[
f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.
\]

This can be interpreted in the standard way as the limiting slope of the secant line as the points of intersection approach each other.

Example 14.1. If \( f(x) = c \) for some \( c \in \mathbb{R} \), then

\[
\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{h \to 0} \frac{c - c}{h} = 0.
\]

So, \( f'(x) = 0 \) everywhere.

Example 14.2. If \( f(x) = x \), then

\[
\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{h \to 0} \frac{x_0 + h - x_0}{h} = \lim_{h \to 0} \frac{h}{h} = 1.
\]

So, \( f'(x) = 1 \) everywhere.

Theorem 14.1. For any function \( f \), \( D(f) \subset C(f) \).

Proof. Suppose \( x_0 \in D(f) \). Then

\[
\lim_{x \to x_0} |f(x) - f(x_0)| = \lim_{x \to x_0} \left| \frac{f(x) - f(x_0)}{x - x_0} (x - x_0) \right| = f'(x_0) |0| = 0.
\]

This shows \( \lim_{x \to x_0} f(x) = f(x_0) \), and \( x_0 \in C(f) \).

Example 14.3. The function \( f(x) = |x| \) is continuous on \( \mathbb{R} \), but

\[
\lim_{h \to 0} \frac{f(0 + h) - f(0)}{h} = 1 = -\lim_{h \to 0} \frac{f(0 + h) - f(0)}{h},
\]

so \( f'(0) \) fails to exist.
Theorem 14.2. Suppose \( f \) and \( g \) are functions such that \( x_0 \in D(f) \cap D(g) \).

(a) \( x_0 \in D(f + g) \) and \( (f + g)'(x_0) = f'(x_0) + g'(x_0) \).

(b) If \( a \in \mathbb{R} \), then \( x_0 \in D(af) \) and \( (af)'(x_0) = af'(x_0) \).

(c) \( x_0 \in D(fg) \) and \( (fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0) \).

(d) If \( g(x_0) \neq 0 \), then \( x_0 \in D(f/g) \) and

\[
\left( \frac{f}{g} \right)(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{(g(x_0))^2}.
\]

Proof. (a)

\[
\lim_{h \to 0} \frac{(f + g)(x_0 + h) - (f + g)(x_0)}{h} = \lim_{h \to 0} \frac{f(x_0 + h) + g(x_0 + h) - f(x_0) - g(x_0)}{h} = \lim_{h \to 0} \left( \frac{f(x_0 + h) - f(x_0)}{h} + \frac{g(x_0 + h) - g(x_0)}{h} \right) = f'(x_0) + g'(x_0)
\]

(b)

\[
\lim_{h \to 0} \frac{(af)(x_0 + h) - (af)(x_0)}{h} = a \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = af'(x_0)
\]

(c)

\[
\lim_{h \to 0} \frac{(fg)(x_0 + h) - (fg)(x_0)}{h} = \lim_{h \to 0} \frac{f(x_0 + h)g(x_0 + h) - f(x_0)g(x_0)}{h}
\]

Now, “slip a 0” into the numerator and factor the fraction.

\[
= \lim_{h \to 0} \frac{f(x_0 + h)g(x_0 + h) - f(x_0)g(x_0) + f(x_0)g(x_0 + h) - f(x_0)g(x_0)}{h}
= \lim_{h \to 0} \left( \frac{f(x_0 + h) - f(x_0)}{h}g(x_0 + h) + f(x_0)\frac{g(x_0 + h) - g(x_0)}{h} \right)
\]

Finally, use the definition of the derivative and the continuity of \( f \) and \( g \) at \( x_0 \).

\[
= f'(x_0)g(x_0) + f(x_0)g'(x_0)
\]
Section 14: Differentiation

(d) It will be proved that if \( g(x_0) \neq 0 \), then \( (1/g)'(x_0) = -g'(x_0)/(g(x_0))^2 \). This statement, combined with (c), yields (d).

\[
\lim_{h \to 0} \frac{(1/g)(x_0 + h) - (1/g)(x_0)}{h} = \lim_{h \to 0} \frac{1}{g(x_0 + h)} - \frac{1}{g(x_0)} \frac{g(x_0) - g(x_0 + h)}{h} \frac{1}{g(x_0 + h)g(x_0)} \frac{1}{g(x_0) - y_0} \frac{y - y_0}{y - y_0} = \frac{-g'(x_0)}{(g(x_0))^2}
\]

Plug this into (c) to see

\[
\left( \frac{f}{g} \right)'(x_0) = \left( \frac{1}{g} \right)'(x_0) = f'(x_0) \frac{1}{g(x_0)} + f(x_0) \frac{-g'(x_0)}{(g(x_0))^2} = f'(x_0)g(x_0) - f(x_0)g'(x_0) \frac{1}{(g(x_0))^2}.
\]

\[
\Box
\]

Combining Examples 14.1 and 14.2 with Theorem 14.2, the following theorem is immediate.

**Theorem 14.3.** A rational function is differentiable at every point of its domain.

**Theorem 14.4 (Chain Rule).** If \( f \) and \( g \) are functions such that \( x_0 \in D(f) \) and \( f(x_0) \in D(g) \), then \( x_0 \in D(g \circ f) \) and \( (g \circ f)'(x_0) = g' \circ f(x_0) f'(x_0) \).

**Proof.** Let \( y_0 = f(x_0) \). By assumption, there is an open interval \( J \) containing \( f(x_0) \) such that \( g \) is defined on \( J \). Since \( J \) is open and \( x_0 \in C(f) \), there is an open interval \( I \) containing \( x_0 \) such that \( f(I) \subset J \).

Define \( h : J \to \mathbb{R} \) by

\[
h(y) = \begin{cases} 
\frac{g(y) - g(y_0)}{y - y_0} - g'(y_0), & y \neq y_0 \\
0, & y = y_0
\end{cases}
\]

Since \( y_0 \in D(f) \), we see

\[
\lim_{y \to y_0} h(y) = \lim_{y \to y_0} \frac{g(y) - g(y_0)}{y - y_0} - g'(y_0) = g'(y_0) - g'(y_0) = 0 = h(0),
\]

so \( y_0 \in C(h) \). Now, \( x_0 \in C(f) \) and \( f(x_0) = y_0 \in C(h) \), so Theorem 10.6 implies \( x_0 \in C(h \circ f) \). In particular

\[
\lim_{x \to x_0} h \circ f(x) = 0.
\]

(5)
Section 14: Differentiation

From the definition of \( h \circ f \) for \( x \in I \) with \( f(x) \neq f(x_0) \), we can solve for

\[
g \circ f(x) - g \circ f(x_0) = (h \circ f(x) + g' \circ f(x_0))(f(x) - f(x_0)). \tag{6}
\]

Notice that (6) is also true when \( f(x) = f(x_0) \). Divide both sides of (6) by \( x - x_0 \), and use (5) to obtain

\[
\lim_{x \to x_0} \frac{g \circ f(x) - g \circ f(x_0)}{x - x_0} = \lim_{x \to x_0} \frac{(h \circ f(x) + g' \circ f(x_0))(f(x) - f(x_0))}{x - x_0}
\]

\[
= (0 + g' \circ f(x_0))f'(x_0)
\]

\[
= g' \circ f(x_0)f'(x_0).
\]

\[\square\]

**Theorem 14.5.** Suppose \( f : [a, b] \to \mathbb{R} \) is continuous and invertible. If \( x_0 \in D(f) \) and \( f'(x_0) \neq 0 \) for some \( x_0 \in (a, b) \), then \( f(x_0) \in D(f^{-1}) \) and \((f^{-1})'(f(x_0)) = 1/f'(x_0)\).

**Proof.** Let \( y_0 = f(x_0) \) and suppose \( y_n \) is any sequence in \( f([a, b]) \setminus \{y_0\} \) converging to \( y_0 \) and \( x_n = f^{-1}(y_n) \). By Theorem 12.5, \( f^{-1} \) is continuous, so

\[
x_0 = f^{-1}(y_0) = \lim_{n \to \infty} f^{-1}(y_n) = \lim_{n \to \infty} x_n.
\]

Therefore,

\[
\lim_{n \to \infty} \frac{f^{-1}(y_n) - f^{-1}(y_0)}{y_n - y} = \lim_{n \to \infty} \frac{x_n - x_0}{f(x_n) - f(x_0)} = \frac{1}{f'(x_0)}.
\]

\[\square\]

**Example 14.4.** It follows easily from Theorem 14.2 that \( f(x) = x^3 \) is differentiable everywhere with \( f'(x) = 3x^2 \). Define \( g(x) = \sqrt[3]{x} \). Then \( g(x) = f^{-1}(x) \). Suppose \( g(y_0) = x_0 \) for some \( y_0 \in \mathbb{R} \). According to Theorem 14.5,

\[
g'(y_0) = \frac{1}{f'(x_0)} = \frac{1}{3x_0^2} = \frac{1}{3g(y_0)^2} = \frac{1}{3(\sqrt[3]{y_0})^2} = \frac{1}{3y_0^{2/3}}.
\]

In the same manner as Example 14.4, the following corollary can be proved.

**Corollary 14.6.** Suppose \( q \in \mathbb{Q} \), \( f(x) = x^q \) and \( D \) is the domain of \( f \). Then \( f'(x) = qx^{q-1} \) on the set

\[
\begin{cases}
D, & \text{when } q \geq 1 \\
D \setminus \{0\}, & \text{when } q < 1
\end{cases}
\]

As is learned in calculus, the derivative is a powerful tool for determining the behavior of functions. The following theorems form the basis for much of differential calculus. First, we state a few familiar definitions.
Definition 14.2. Suppose \( f : D \to \mathbb{R} \) and \( x_0 \in D \). \( f \) is said to have a relative maximum at \( x_0 \) if there is a \( \delta > 0 \) such that \( f(x) \leq f(x_0) \) for all \( x \in (x_0 - \delta, x_0 + \delta) \cap D \). \( f \) has a relative minimum at \( x_0 \) if \( -f \) has a relative maximum at \( x_0 \). If \( f \) has either a relative maximum or a relative minimum at \( x_0 \), then it is said that \( f \) has a relative extreme value at \( x_0 \).

The absolute maximum of \( f \) occurs at \( x_0 \) if \( f(x_0) \geq f(x) \) for all \( x \in D \). The definitions of absolute minimum and absolute extreme are analogous.

Examples like \( f(x) = x \) on \((0,1)\) show that even the nicest functions need not have relative extrema. Corollary 12.4 shows that if \( D \) is compact, then any continuous function defined on \( D \) assumes both an absolute maximum and an absolute minimum on \( D \).

Theorem 14.7. Suppose \( f : (a, b) \to \mathbb{R} \) is differentiable. If \( f \) has a relative extreme value at \( x_0 \), then \( f'(x_0) = 0 \).

Proof. Suppose \( f(x_0) \) is a relative maximum value of \( f \). Then there must be a \( \delta > 0 \) such that \( f(x) \leq f(x_0) \) whenever \( x \in (x_0 - \delta, x_0 + \delta) \). Since \( f'(x_0) \) exists,

\[
\lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0} \leq 0
\]

and

\[
\lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0} \geq 0.
\]

Combining (7) and (8) shows \( f'(x_0) = 0 \).

If \( f(x_0) \) is a relative minimum value of \( f \), apply the previous argument to \( -f \).

Theorem 14.7 is, of course, the basis for much of a beginning calculus course. If \( f : [a, b] \to \mathbb{R} \), then the extreme values of \( f \) occur at points of the set

\[
C = \{ x \in (a, b) : f'(x) = 0 \} \cup \{ x \in [a, b] : f'(x) \text{ does not exist} \}.
\]

The elements of \( C \) are often called the critical points of \( f \) on \([a, b]\). To find the maximum and minimum values of \( f \) on \([a, b]\), it suffices to find its maximum and minimum on the smaller set \( C \).

Problem 26. If \( f \) is defined on an open set containing \( x_0 \), the symmetric derivative of \( f \) at \( x_0 \) is defined as

\[
f^s(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0 - h)}{2h}.
\]

Prove that if \( f'(x) \) exists, then so does \( f^s(x) \). Is the converse true?
15 Differentiable Functions

Definition 15.1. The function \( f \) is differentiable on an open interval \( I \) if \( I \subset D(f) \). If \( f \) is differentiable on its domain, then it is said to be differentiable. In this case, the function \( f' \) is called the derivative of \( f \).

Lemma 15.1 (Rolle’s Theorem). If \( f : [a, b] \to \mathbb{R} \) is continuous on \([a, b]\), differentiable on \((a, b)\) and \( f(a) = 0 = f(b) \), then there is a \( c \in (a, b) \) such that \( f'(c) = 0 \).

Proof. Since \([a, b]\) is compact, Corollary 12.4 implies the existence of \(x_m, x_M \in [a, b]\) such that \( f(x_m) \leq f(x) \leq f(x_M) \) for all \( x \in [a, b] \). If \( f(x_m) = f(x_M) \), then \( f \) is constant on \([a, b]\) and any \( c \in (a, b) \) satisfies the lemma. Otherwise, either \( f(x_m) < 0 \) or \( f(x_M) > 0 \). If \( f(x_m) < 0 \), then \( x_m \in (a, b) \) and Theorem 14.7 implies \( f'(x_m) = 0 \). If \( f(x_M) > 0 \), then \( x_M \in (a, b) \) and Theorem 14.7 implies \( f'(x_M) = 0 \).

Theorem 15.2 (Cauchy Mean Value Theorem). If \( f : [a, b] \to \mathbb{R} \) and \( g : [a, b] \to \mathbb{R} \) are such that \( f \) and \( g \) are continuous on \([a, b]\) and differentiable on \((a, b)\), then there is a \( c \in (a, b) \) such that

\[
g'(c)(f(b) - f(a)) = f'(c)(g(b) - g(a)).
\]

Proof. Let

\[
h(x) = (g(b) - g(a))(f(a) - f(x)) + (g(x) - g(a))(f(b) - f(a)).
\]

Then \( h \) is continuous on \([a, b]\) and differentiable on \((a, b)\) with \( h(a) = h(b) = 0 \). Theorem 15.1 yields a \( c \in (a, b) \) such that \( h'(c) = 0 \). Then

\[
0 = h'(c) = -(g(b) - g(a))f'(c) + g'(c)(f(b) - f(a))
\]

\[
\implies g'(c)(f(b) - f(a)) = f'(c)(g(b) - g(a)).
\]

Corollary 15.3 (Mean Value Theorem). If \( f : [a, b] \to \mathbb{R} \) is continuous on \([a, b]\) and differentiable on \((a, b)\), then there is a \( c \in (a, b) \) such that \( f(b) - f(a) = f'(c)(b - a) \).

Proof. Let \( g(x) = x \) in Theorem 15.2.

Theorem 15.4. Suppose \( f : (a, b) \to \mathbb{R} \) is a differentiable function. \( f \) is increasing on \((a, b)\) iff \( f'(x) \geq 0 \) for all \( x \in (a, b) \).

Proof. Choose \( \alpha, \beta \in (a, b) \) with \( \alpha < \beta \). According to Corollary 15.3, there is a \( c \in (\alpha, \beta) \) such that

\[
f(\beta) - f(\alpha) = f'(c)(\beta - \alpha) \geq 0.
\]

\( \Box \)

\( ^4 \)Theorem 15.2 is also often called the “Generalized Mean Value Theorem.”
Corollary 15.5. Let \( f : (a, b) \to \mathbb{R} \) be a differentiable function. \( f \) is constant iff \( f'(x) = 0 \) for all \( x \in (a, b) \).

**Theorem 15.6 (Darboux’s Theorem).** If \( f \) is differentiable on an open set containing \([a, b]\) and \( \gamma \) is between \( f'(a) \) and \( f'(b) \), then there is a \( c \in [a, b] \) such that \( f'(c) = \gamma \).

**Proof.** If \( f'(a) = f'(b) \), then \( c = a \) satisfies the theorem. So, we may as well assume \( f'(a) \neq f'(b) \). There is no generality lost in assuming \( f'(a) < f'(b) \), for, otherwise, we just replace \( f \) with \( g = -f \).

Let \( h(x) = f(x) - \gamma(x - \alpha) \) so that \( D(f) = D(h) \) and \( h'(x) = f'(x) - \gamma \). In particular, this implies \( h'(a) < 0 < h'(b) \). Because of this, there must be an \( h > 0 \) small enough so that

\[
\frac{f(a + h) - f(a)}{h} < 0 \implies f(a + h) < f(a)
\]

and

\[
\frac{f(b) - f(b - h)}{h} > 0 \implies f(b - h) < f(b).
\]

(See Figure 10.) In light of these two inequalities and Theorem 12.4, there must be a \( c \in (a, b) \) such that \( f(c) = \text{glb} \{ f(x) : x \in [a, b] \} \). Now Theorem 14.7 gives \( 0 = h'(c) = f'(c) - \gamma \), and the theorem follows. \qed
16 Applications of the Mean Value Theorem

For the following sections, we require the standard idea of higher order derivatives. If \( n \in \mathbb{N} \), then the \( n \)th order derivative of \( f \) at \( x_0 \) is written \( f^{(n)}(x_0) \). We also use the convention that \( f^{(0)} = f \).

16.1 Taylor’s Theorem

The motivation behind Taylor’s theorem is the attempt to approximate a function \( f \) near a number \( a \) by a polynomial. The polynomial of degree 0 which does the best job is clearly \( p_0(x) = f(a) \). The best polynomial of degree 1 is the tangent line to the graph of the function \( p_1(x) = f(a) + f'(a)(x-a) \). Continuing in this way, we approximate \( f \) near \( a \) by the polynomial \( p_n \) of degree \( n \) such that

\[
f^{(k)}(a) = p_n^{(k)}(a) \quad \text{for} \quad k = 0, 1, \ldots, n.
\]

A simple induction argument shows that

\[
p_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (x-a)^k.
\]

This is the well-known Taylor polynomial of \( f \) at \( a \).

The fact which makes the Taylor polynomial important is that in many cases it is possible to determine how large \( n \) must be to achieve a desired accuracy in the approximation of \( f \) by \( p_n \). This is accomplished by using Taylor’s Theorem, which is also known as the Extended Mean Value Theorem.

**Theorem 16.1 (Taylor’s Theorem).** If \( f \) is a function such that \( f, f', \ldots, f^{(n)} \) are continuous on \([a, b]\) and \( f^{(n+1)} \) exists on \((a, b)\), then there is a \( c \in (a, b) \) such that

\[
f(b) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (b-a)^k + \frac{f^{(n+1)}(c)}{(n+1)!} (b-a)^{n+1}.
\]

**Proof.** Let the constant \( \alpha \) be defined by

\[
f(b) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (b-a)^k + \frac{\alpha}{(n+1)!} (b-a)^{n+1}
\]

and define

\[
F(x) = f(b) - \left( \sum_{k=0}^{n} \frac{f^{(k)}(x)}{k!} (b-x)^k + \frac{\alpha}{(n+1)!} (b-x)^{n+1} \right).
\]

From (10) we see that \( F(a) = 0 \). Direct substitution in the definition of \( F \) shows that \( F(b) = 0 \). From the assumptions in the statement of the theorem, it is easy to see that \( F \) is continuous on \([a, b]\) and differentiable on \((a, b)\). An application of Rolle’s Theorem yields a \( c \in (a, b) \) such that

\[
0 = F'(c) = - \left( \frac{f^{(n+1)}(c)}{n!} (b-c)^n - \frac{\alpha}{n!} (b-c)^n \right) \implies \alpha = f^{(n+1)}(c),
\]
Section 15: Differentiable Functions

Figure 11: Here are several of the Taylor polynomials for the function \( f(x) = \cos(x) \) graphed along with \( f \).

as desired.

Now, suppose \( f \) is defined on an open interval \( I \) with \( a, x \in I \). If \( f \) is \( n+1 \) times differentiable on \( I \), then Theorem 16.1 implies there is a \( c \) between \( a \) and \( x \) such that

\[
f(x) = p_n(x) + R_f(n, x, a),
\]

where \( R_n(c, x, a) = \frac{f^{(n+1)}(c)}{(n+1)!}(x - a)^{n+1} \) is the error in the approximation.

**Example 16.1.** Let \( f(x) = \cos(x) \). Suppose we want to approximate \( f(2) \) to 5 decimal places of accuracy. Since it’s an easy point to work with, we’ll choose \( a = 0 \). Then, for some \( c \in (0, 2) \),

\[
|R_f(n, 2, 0)| = \left| \frac{f^{(n+1)}(c)}{(n + 1)!} 2^{n+1} \right| \le \frac{2^{n+1}}{(n + 1)!}.
\]

A bit of experimentation with a calculator shows that \( n = 12 \) is the smallest \( n \) such that the right-hand side of (11) is less than \( 5 \times 10^{-6} \). After doing some arithmetic, it follows that

\[
p_{12}(2) = 1 - \frac{2^2}{2!} + \frac{2^4}{4!} - \frac{2^6}{6!} + \frac{2^8}{8!} - \frac{2^{10}}{10!} + \frac{2^{12}}{12!} = -27809 \approx -0.41614.
\]

is a 5 decimal place approximation to \( \cos(2) \).
16.2 L’Hôpital’s Rules and Indeterminate Forms

According to Theorem 8.3,
\[
\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}
\]
whenever \(\lim_{x \to a} f(x)\) and \(\lim_{x \to a} g(x)\) both exist and \(\lim_{x \to a} g(x) \neq 0\). But, it is easy to find examples where both \(\lim_{x \to a} f(x) = 0\) and \(\lim_{x \to a} g(x) = 0\) and \(\lim_{x \to a} f(x)/g(x)\) exists, as well as similar examples where \(\lim_{x \to a} f(x)/g(x)\) fails to exist. Because of this, such a limit problem is said to be in the indeterminate form 0/0. The following theorem allows us to determine many such limits.

**Theorem 16.2 (Easy L’Hôpital’s Rule).** Suppose \(f\) and \(g\) are each continuous on \([a, b]\), differentiable on \((a, b)\) and \(f(b) = g(b) = 0\). If \(f'(x) \neq 0\) on \((a, b)\) and \(\lim_{x \to b} f'(x)/g'(x) = L\), where \(L\) could be infinite, then \(\lim_{x \to b} f(x)/g(x) = L\).

**Proof.** Let \(x \in [a, b]\), so \(f\) and \(g\) are continuous on \([x, b]\) and differentiable on \((x, b)\). Cauchy’s Mean Value Theorem, Theorem 15.2, implies there is a \(c(x) \in (x, b)\) such
\[
f'(c(x))g(x) = g'(c(x))f(x) \implies \frac{f(x)}{g(x)} = \frac{f'(c(x))}{g'(c(x))}.
\]
Since \(x < c(x) < b\), it follows that \(\lim_{x \to b} c(x) = c\). This shows that
\[
L = \lim_{x \to b} \frac{f'(x)}{g'(x)} = \lim_{x \to b} \frac{f'(c(x))}{g'(c(x))} = \lim_{x \to b} \frac{f(x)}{g(x)}.
\]

Several things should be noted about this proof. First, there is nothing special about the left-hand limit used in the statement of the theorem. It could just as easily be written in terms of the right-hand limit. Second, if \(\lim_{x \to a} f(x)/g(x)\) is not of the indeterminate form 0/0, then applying L’Hôpital’s rule will give a wrong answer. To see this, consider
\[
\lim_{x \to 0} \frac{x}{x + 1} = 0 \neq 1 = \lim_{x \to 0} \frac{1}{x}.
\]

**Corollary 16.3.** Suppose \(f\) and \(g\) are differentiable on \((a, \infty)\) and \(\lim_{x \to \infty} f(x) = \lim_{x \to \infty} g(x) = 0\). If \(f'(x) \neq 0\) on \((a, \infty)\) and \(\lim_{x \to \infty} f'(x)/g'(x) = L\), where \(L\) could be infinite, then \(\lim_{x \to \infty} f(x)/g(x) = L\).

**Proof.** There is no generality lost by assuming \(a > 0\). Let
\[
F(x) = \begin{cases} f(1/x), & x \in [a, \infty) \\ 0, & x = 0 \end{cases} \quad \text{and} \quad G(x) = \begin{cases} g(1/x), & x \in [a, \infty) \\ 0, & x = 0 \end{cases}.
\]
Then
\[
\lim_{x \to 0} F(x) = \lim_{x \to \infty} f(x) = 0 = \lim_{x \to \infty} g(x) \lim_{x \to 0} G(x),
\]
so both \( F \) and \( G \) are continuous at 0. It follows that both \( F \) and \( G \) are continuous on \([0, 1/a]\) and differentiable on \((0, 1/a)\) with \( G'(x) = -g'(x)/x^2 \neq 0 \) on \((0, 1/a)\) and \( \lim_{x \to 0} F'(x)/G'(x) = \lim_{x \to \infty} f'(x)/g'(x) = L \). The rest follows from Theorem 16.2.

The other standard indeterminate form is when \( \lim_{x \to \infty} f(x) = \infty = \lim_{x \to \infty} g(x) \). This is called an \( \infty/\infty \) indeterminate form. This is handled by the following theorem.

**Theorem 16.4 (Hard L'Hôpital's Rule).** Suppose that \( f \) and \( g \) are differentiable on \((a, \infty)\) and \( g'(x) \neq 0 \) on \((a, \infty)\). If
\[
\lim_{x \to \infty} f(x) = \lim_{x \to \infty} g(x) = \infty \quad \text{and} \quad \lim_{x \to \infty} \frac{f(x)}{g(x)} = L \in \mathbb{R} \cup \{-\infty, \infty\},
\]
then
\[
\lim_{x \to \infty} \frac{f(x)}{g(x)} = L.
\]

**Proof.** First, suppose \( L \in \mathbb{R} \) and let \( \varepsilon > 0 \). Choose \( a_1 > a \) large enough so that
\[
\left| \frac{f'(x)}{g'(x)} - L \right| < \varepsilon, \quad \forall x > a_1.
\]

(12)
Since \( \lim_{x \to \infty} f(x) = \infty = \lim_{x \to \infty} g(x) \), we can assume there is an \( a_2 > a_1 \) such that both \( f(x) > 0 \) and \( g(x) > 0 \) when \( x > a_2 \). Finally, choose \( a_3 > a_2 \) such that whenever \( x > a_3 \), then \( f(x) > f(a_2) \) and \( g(x) > g(a_2) \).

Let \( x > a_3 \) and apply Cauchy's Mean Value Theorem, Theorem 15.2, to \( f \) and \( g \) on \([a_2, x]\) to find a \( c(x) \in (a_2, x) \) such that
\[
\frac{f'(c(x))}{g'(c(x))} \left( \frac{1 - f(a_2)}{f(x)} \right) = \frac{f(x) - f(a_2)}{g(x) - g(a_2)} = \frac{f(x) \left( 1 - f(a_2)/f(x) \right)}{g(x) \left( 1 - g(a_2)/g(x) \right)}.
\]

(13)
If
\[
h(x) = \frac{1 - g(a_2)/f(x)}{1 - f(a_2)/f(x)},
\]
then (13) implies
\[
\frac{f(x)}{g(x)} = \frac{f'(c(x))}{g'(c(x))} h(x).
\]
Section 15: Differentiable Functions

Since \( \lim_{x \to \infty} h(x) = 1 \), there is an \( a_4 > a_3 \) such that whenever \( x > a_4 \), then \( |h(x) - 1| < \varepsilon \). If \( x > a_4 \), then

\[
\left| \frac{f(x)}{g(x)} - L \right| = \left| \frac{f'(c(x))}{g'(c(x))} \right| h(x) - L \\
\leq \left| \frac{f'(c(x))}{g'(c(x))} h(x) - L h(x) + L h(x) - L \right| \\
< \varepsilon (1 + \varepsilon) + |L| \varepsilon = (1 + |L| + \varepsilon) \varepsilon.
\]

Therefore \( \lim_{x \to \infty} \frac{f(x)}{g(x)} = L \).

The case when \( L = \infty \) is done similarly by first choosing a \( B > 0 \) and adjusting (13) so that \( f'(x)/g'(x) > B \) when \( x > a_1 \). A similar adjustment is necessary when \( L = -\infty \). \( \square \)

There is a companion corollary to Theorem 16.4 which is proved in the same way as Corollary 16.3.

**Corollary 16.5.** Suppose that \( f \) and \( g \) are continuous on \([a, b]\) and differentiable on \((a, b)\) with \( g'(x) \neq 0 \) on \((a, b)\). If

\[
\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = \infty \quad \text{and} \quad \lim_{x \to a} \frac{f(x)}{g(x)} = L \in \mathbb{R} \cup \{-\infty, \infty\},
\]

then

\[
\lim_{x \to a} \frac{f(x)}{g(x)} = L.
\]
References
