MAT 371 LECTURES

JOHN QUIGG

CONTENTS

1. Sets 1
2. The real numbers 10
3. Countability 21
4. Sequences 23
5. Limits 34
6. Continuity 38
7. Differentiation 42
8. Integration 48
9. Log and exp 60
10. Series 63
11. Sequences of functions 69
12. Series of functions 72
13. Power series 73
Index 80

1. SETS

The terms set and element will be undefined.

Remark 1.1. All a set $A$ knows about is what its elements are. If $x$ is any object, the sentence “$x$ is an element of $A$”, written “$x \in A$”, is a (mathematical) proposition—that is, it is either true or false. Roughly speaking, all a set is good for is to allow the formation of such propositions. In fact, strictly speaking these are the only propositions in mathematics; more precisely, every mathematical proposition can be built up from those of the form “$x \in A$”. The rules governing these propositions comprise set theory, which is the foundation of modern mathematics. We won’t study the formal theory of sets here; instead, we’ll just record a few of the most important facts concerning sets, partly to indicate the level of rigor, and the style of writing, of our definitions, results (theorems, corollaries, propositions, and lemmas), and proofs. Virtually everything in this section, except perhaps for the general concept of families of sets occurring at the end, is prerequisite material for the course.

Date: May 6, 2002.
Many definitions in set theory are merely translations of constructions from logic; for example, the logical biconditional becomes set equality:

Definition 1.2. Let $A$ and $B$ be sets. We say $A$ equals $B$, written $A = B$, if for all $x$ we have $x \in A$ if and only if $x \in B$.

Notation and Terminology 1.3. The negation of “$x \in A$” is written “$x \not\in A$”.

Definition 1.4. The empty set is the unique set $\emptyset$ such that for all $x$ we have $x \not\in \emptyset$.

Remark 1.5. There is no set $A$ such that for all $x$ we have $x \in A$; such a set would be “too big to be”—that is, its existence would lead to (logical) contradictions, of which the most famous is Russell’s Paradox.

The set-theoretic translation of the logical conditional is the subset relation:

Definition 1.6. Let $A$ and $B$ be sets.

(i) $A$ is called a subset of $B$, written $A \subseteq B$, if for all $x$ we have $x \in A$ implies $x \in B$.

(ii) If $A \subseteq B$, we also say $B$ is a superset of $A$, written $B \supseteq A$.

(iii) A subset $A$ of $B$ is called proper if $A \neq B$.

Remark 1.7. In the above definition, we used a phrase of the form “we also say ...”. It is important to realize that this just introduces a synonym for the primary term—the definition of each synonym is the same as that of the primary term. For example, the definition of “$B \supseteq A$” is the same as that of “$A \subseteq B$”.

Example 1.8. For every set $A$, we have:

(i) $A \subseteq A$;

(ii) $\emptyset \subseteq A$.

Lemma 1.9. Let $A$ and $B$ be sets. Then $A = B$ if and only if both $A \subseteq B$ and $B \subseteq A$.

Proof. First assume $A = B$. Since $A \subseteq A$, we have both $A \subseteq B$ and $B \subseteq A$.

Conversely, assume $A \subseteq B$ and $B \subseteq A$. Let $x$ be any object. First assume $x \in A$. Then $x \in B$ since $A \subseteq B$. Conversely, assume $x \in B$. Then $x \in A$ since $B \subseteq A$. Thus $x \in A$ if and only if $x \in B$, so $A = B$.

Remark 1.10. In the forward direction of the above proof, we had to prove $A \subseteq B$ (as well as $B \subseteq A$); a more typical proof of this would start with “let $x \in A$” and would deduce “$x \in B$”, however, the hypothesis “$A = B$” was so special that it trivially implied the desired conclusion.

Actually, the above lemma has a much shorter proof, because it is just a set-theoretic version of a tautology from logic (namely, the biconditional $p \leftrightarrow q$ is equivalent to the conjunction $p \Rightarrow q$ and $q \Rightarrow p$). However, the whole point here is to indicate typical proofs with sets, not to give the most efficient proofs of the results themselves.
The set-theoretic version of the logical disjunction, conjunction, and negation are union, intersection, and complement:

**Definition 1.11.** Let $A$ and $B$ be sets.

(i) The **union** of $A$ and $B$ is defined by

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

(ii) The **intersection** of $A$ and $B$ is defined by

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

(iii) $A$ and $B$ are called **disjoint** if $A \cap B = \emptyset$.

(iv) The **difference** of $A$ and $B$ is defined by

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

(v) If the set $U$ is understood, for any subset $A \subseteq U$ the **complement** of $A$ is defined by

$$A^c := U \setminus A.$$  

**Remark 1.12.** Rather than state the elementary properties of these operations, we'll wait until we have the more general versions for families of sets (set below).

Note that the complement of $A$ is not just $\{ x : x \notin A \}$, rather it is $\{ x \in U : x \notin A \}$. Thus the complement depends upon the choice of the particular “universe” $U$; there is no single set $U$ that could serve as a universe for taking complements of every set $A$, because, as we mentioned above, there is no single set $U$ which contains all sets.

**Definition 1.13.**

(i) Let $x$ and $y$ be objects. The **ordered pair** with first coordinate $x$ and second coordinate $y$ is defined by

$$(x, y) := \{ \{ x \}, \{ x, y \} \}.$$  

(ii) The **Cartesian product** of sets $A$ and $B$ is defined by

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

**Remark 1.14.** Note that ordered pairs $(x, y)$ and $(z, w)$ are equal if and only if $x = z$ and $y = w$. However, it would be improper to try to define “ordered pair" by this property; it is important to know that we can define all our concepts completely in terms of sets.

**Notation and Terminology 1.15.** $\mathbb{R}$ denotes the set of real numbers, which we will discuss in more detail in the next section.

**Definition 1.16.** Let $A$ and $B$ be sets, and let $f \subseteq A \times B$. We say $f$ is a function from $A$ to $B$, written $f : A \to B$, if for all $x \in A$ there exists a unique $y \in B$ such that $(x, y) \in f$. We call this $y$ the value of $f$ at $x$, written $y = f(x)$.

**Remark 1.17.** It would be improper to try to define a “function” as a “rule” with certain properties, because it raises the question “what is a rule?”; again, it is important to know that we can phrase the definition in terms of sets.
Definition 1.18. Let \( f : A \to B \).

(i) The set \( A \) is called the \textit{domain} of \( f \), written \( A = \text{dom} \ f \).

(ii) The \textit{range} of \( f \) is defined by
\[
\text{ran} \ f := \{ f(x) : x \in A \}.
\]

(iii) If \( C \subseteq A \), the \textit{image} of \( C \) under \( f \) is defined by
\[
f(C) := \{ f(x) : x \in C \}.
\]

(iv) If \( D \subseteq B \), the \textit{pre-image} of \( D \) under \( f \) is defined by
\[
f^{-1}(D) := \{ x \in A : f(x) \in D \}.
\]

(v) \( f \) is called \textit{real-valued} if \( \text{ran} \ f \subseteq \mathbb{R} \).

Remark 1.19. Note that the set \( A \) is determined by the function \( f \), since \( A = \text{dom} \ f \). However, the set \( B \) is not determined by \( f \)—it could be any superset of the range of \( f \). This has an important consequence: if we have a function \( f : A \to B \), and if \( C \) is any set such that \( \text{ran} \ f \subseteq C \), then we can equally well regard \( f : A \to C \); it is still the \textit{same} function. Actually, we have made a choice here—in some areas of advanced mathematics it is important to regard the set \( B \) as an official part of the function, but we have chosen to identify the function with its graph \( \{ (x, f(x)) : x \in \text{dom} \ f \} \).

Remark 1.20. The notation \( f(C) \) for images is an abuse of the function notation; we must take care to interpret the notation correctly from the context. Even more dangerous is the notation \( f^{-1}(D) \) for pre-images, since it can be confused with the concept of an \textit{inverse function} (see below).

Notation and Terminology 1.21. Occasionally it is convenient to introduce a function but not give it a name, in which case we use the notation \( \{ x \mapsto (\text{some formula involving } x) \} \). For example, we could consider “the function \( x \mapsto x^2 \) from \( \mathbb{R} \) to \( \mathbb{R} \).”

Remark 1.22. Thus, a function \( f \) is just a set of ordered pairs with a certain property, popularly known as the “vertical line test”: for all \( x, y, z \), if \((x, y) \in f \) and \((x, z) \in f \), then \( y = z \). In particular, to prove two functions are equal is just a case of proving two sets are equal. But since functions are such special kinds of sets, there is a more efficient way to prove function equality:

Lemma 1.23. Let \( f \) and \( g \) be functions. Then \( f = g \) if and only if \( \text{dom} \ f = \text{dom} \ g \) and for all \( x \in \text{dom} \ f \) we have \( f(x) = g(x) \).

Proof. First, if \( f = g \) then trivially \( \text{dom} \ f = \text{dom} \ g \) and for all \( x \in \text{dom} \ f \) we have \( f(x) = g(x) \).

Conversely, assume \( \text{dom} \ f = \text{dom} \ g \) and for all \( x \in \text{dom} \ f \) we have \( f(x) = g(x) \). We will show \( f \subseteq g \) and \( g \subseteq f \); since the hypothesis is symmetric in \( f \) and \( g \), it suffices to show \( f \subseteq g \). Let \((x, y) \in f \). Then \( x \in \text{dom} \ f \) and \( f(x) = y \). By hypothesis we have \( x \in \text{dom} \ g \) and \( g(x) = y \). Thus \((x, y) \in g \), and we have shown \( f \subseteq g \). \(\square\)
Remark 1.24. In the converse direction of the above proof, when we were showing \( f \subseteq g \) we did not just start with something like “let \( z \in f \)” and then deduce that there exist \( x \) and \( y \) such that \( z = (x, y) \); since the elements of \( f \) are ordered pairs, it was more efficient to start with ordered-pair notation for an arbitrary element of \( f \). In general, we try to use “notation appropriate for the context”.

Definition 1.25. Let \( f : A \to B \) and \( C \subseteq A \).

(i) The function \( f|C : C \to B \) defined by \( f|C(x) = f(x) \) for each \( x \in C \) is called the restriction of \( f \) to \( C \).

(ii) If \( g \) is a function with domain \( C \), then \( f \) is called an extension of \( g \) to \( A \) if \( g = f|C \).

Proposition 1.26. Let \( f : A \to B \), and let \( C, D \subseteq A \) and \( E, F \subseteq B \). Then:

(i) \( f(C \setminus D) \supseteq f(C) \setminus f(D) \);
(ii) \( f^{-1}(E \setminus F) = f^{-1}(E) \setminus f^{-1}(F) \);
(iii) \( f^{-1}(f(C)) \supseteq C \);
(iv) \( f(f^{-1}(E)) \subseteq E \);
(v) \( C \subseteq D \Rightarrow f(C) \subseteq f(D) \);
(vi) \( E \subseteq F \Rightarrow f^{-1}(E) \subseteq f^{-1}(F) \).

Proof. (i) Let \( y \in f(C) \setminus f(D) \). Then \( y \in f(C) \) and \( y \notin f(D) \). Choose \( x \in C \) such that \( y = f(x) \). Then \( x \notin D \) since \( y \notin f(D) \). Thus \( x \in C \setminus D \). Hence \( y \in f(C \setminus D) \) since \( y = f(x) \).

(ii) First let \( x \in f^{-1}(E \setminus F) \). Then \( f(x) \in E \setminus F \). Thus \( f(x) \in E \) and \( f(x) \notin F \). Hence \( x \in f^{-1}(E) \) and \( x \notin f^{-1}(F) \). Therefore \( x \in f^{-1}(E) \setminus f^{-1}(F) \), and we have shown \( f^{-1}(E \setminus F) \subseteq f^{-1}(E) \setminus f^{-1}(F) \).

For the opposite inclusion \( f^{-1}(E) \setminus f^{-1}(F) \subseteq f^{-1}(E \setminus F) \), just reverse the above steps. Explicitly: let \( x \in f^{-1}(E) \setminus f^{-1}(F) \). Then \( x \in f^{-1}(E) \) and \( x \notin f^{-1}(F) \). Thus \( f(x) \in E \) and \( f(x) \notin F \). Hence \( f(x) \in E \setminus F \). Therefore \( x \in f^{-1}(E \setminus F) \).

(iii) If \( x \in C \), then \( f(x) \in f(C) \), so \( x \in f^{-1}(f(C)) \).

(iv) Given \( y \in f(f^{-1}(E)) \), choose \( x \in f^{-1}(E) \) such that \( f(x) = y \). Then \( f(x) \in E \), so \( y \in E \).

(v) Given \( y \in f(C) \), choose \( x \in C \) such that \( f(x) = y \). Then \( x \in D \) since \( C \subseteq D \). Hence \( f(x) \in f(D) \), so \( y \in f(D) \).

(vi) Let \( x \in f^{-1}(E) \). Then \( f(x) \in E \), so \( f(x) \in F \) since \( E \subseteq F \). Therefore \( x \in f^{-1}(F) \).

Remark 1.27. In the proof of part (ii) above, it would have been logically correct to stop after saying “reverse the above steps”. However, this must be done carefully: it is only proper to do so when the steps can really be reversed exactly—if any modifications are necessary in the proof of the reverse direction, they must be explicitly mentioned.

Actually, precisely because the steps are reversible, the proof could have been given for both directions simultaneously by using “if and only if” statements. However, again, the point is to indicate typical proofs with sets, not
just prove the result; to prove two sets are equal it is typically necessary, or at least more convenient, to prove separately that each is a subset of the other.

**Definition 1.28.** Given real-valued functions $f$ and $g$ with the same domain, define

1. $(f + g)(x) := f(x) + g(x)$;
2. $(cf)(x) := cf(x)$ if $c \in \mathbb{R}$;
3. $(fg)(x) := f(x)g(x)$;
4. $\left( \frac{f}{g} \right)(x) := \left\{ \begin{array}{ll} \frac{f(x)}{g(x)} & \text{if } 0 \not\in \text{ran } g. \end{array} \right.$

**Remark 1.29.** Slightly more generally, if $\text{dom } f \neq \text{dom } g$, it is sometimes convenient to still define the above combinations of $f$ and $g$, with domain $\text{dom } f \cap \text{dom } g$, and for $\frac{f}{g}$ we also subtract $g^{-1}(\{0\})$ from the domain.

In fact, it happens so often that we want to restrict a function to a subset of its domain that we often do it without comment; strictly speaking, we are replacing the old function with a new one, but in practice this causes no confusion.

**Definition 1.30.** Given $f : A \to B$ and $g : B \to C$, define $g \circ f : A \to C$ by

$$g \circ f(x) := g(f(x)).$$

**Remark 1.31.** Slightly more generally, if $\text{ran } f \not\subset \text{dom } g$, it is sometimes convenient to still define $g \circ f$ by the above formula, with domain $f^{-1}(\text{dom } g)$.

**Definition 1.32.**

1. The **identity function** on a set $A$ is defined by $\text{id}_A(x) := x$ for $x \in A$.
2. If $A$ and $B$ are sets and $c \in B$, the function $f : A \to B$ defined by $f(x) = c$ for $x \in A$ is called the **constant function** from $A$ to $B$ with value $c$. We also say $f$ is **identically** $c$.

**Definition 1.33.** Let $f : A \to B$.

1. $f$ is called **one-to-one**, written 1-1, if for all $x, y \in A$ we have $f(x) = f(y) \Rightarrow x = y$.
2. $f$ is called **onto** $B$ if $\text{ran } f = B$.
3. If $f$ is 1-1 and onto $B$, the unique function $f^{-1} : B \to A$ such that $f^{-1} \circ f = \text{id}_A$ and $f \circ f^{-1} = \text{id}_B$ is called the **inverse** of $f$.

**Remark 1.34.** Note that being onto is not just a property of the function $f$ itself; it depends upon the set $B$ when we regard $f : A \to B$. In fact, “onto-ness” is really a property of the set $B$ relative to $f$, namely $B$ must be the range of $f$. However, if the set $B$ is understood we can say “$f$ is onto”.
Remark 1.35. The notation \( f^{-1} \) must be used with care: for any function \( f : A \to B \) it makes sense to consider \( f^{-1}(C) \) for any \( C \subseteq B \). But for \( y \in B \) it only makes sense to consider \( f^{-1}(y) \) if we know that the function \( f^{-1} \) exists (i.e., if we know \( f \) is 1-1 onto). Fortunately, the notation is consistent: if \( f^{-1} : B \to A \) does exist and \( C \subseteq B \), then the set \( f^{-1}(C) \) is the same no matter whether we regard it as the pre-image of \( C \) under \( f \) or the image of \( C \) under the function \( f^{-1} \).

Definition 1.36. (i) A set whose elements are sets is (also) called a family of sets.

(ii) A function whose values are sets is called an indexed family of sets.

If the domain of an indexed family of sets is \( I \) and the function is \( i \mapsto A_i \), the family is denoted \( \{A_i\}_{i \in I} \). When \( I = \{1, 2, \ldots \} \) we write \( \{A_i\}_{i=1}^\infty \), and when \( I = \{1, \ldots, n\} \) we write \( \{A_i\}_{i=1}^n \).

Remark 1.37. There is no logical need for the term “family of sets”; nevertheless it is handy because it would sometimes be confusing to say “set of sets”.

Note that to every indexed family \( \{A_i\}_{i \in I} \) of sets, there is an associated family of sets, namely the range \( \{A_i : i \in I\} \) (and please note the subtle change of notation!) of the indexed family. We must take care to remember that the indexed family is very different from the associated family, since a function is different from its range. For some purposes, however, it won’t matter much whether we use a family or an indexed family. Since “family of sets” is simpler both linguistically and conceptually than “indexed family of sets”, we have a slight preference for the former.

There is an important notational aspect to consider: it is common to use “index” notation such as \( \{A_i : i \in I\} \) for a family of sets, so that the family is given as the range of an indexed family. However, when we don’t want to use index notation, the family is commonly denoted with a capital script letter such as \( \mathcal{F} \).

The set-theoretic version of the existential and universal quantifiers from logic are union and intersection for families:

Definition 1.38. Let \( \{A_i\}_{i \in I} \) be an indexed family of sets.

(i) The union of \( \{A_i\}_{i \in I} \) is defined by

\[
\bigcup_{i \in I} A_i := \{x : \text{there exists } i \in I \text{ such that } x \in A_i\}.
\]

When \( I = \{1, 2, \ldots\} \) we write \( \bigcup_{i=1}^\infty A_i = A_1 \cup A_2 \cup \cdots \), and when \( I = \{1, \ldots, n\} \) we write \( \bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \cdots \).

(ii) The intersection of \( \{A_i\}_{i \in I} \) is defined by

\[
\bigcap_{i \in I} A_i := \{x : \text{for all } i \in I, x \in A_i\}.
\]

When \( I = \{1, 2, \ldots\} \) we write \( \bigcap_{i=1}^\infty A_i = A_1 \cap A_2 \cap \cdots \), and when \( I = \{1, \ldots, n\} \) we write \( \bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \cdots \).
(iii) The Cartesian product of \( \{A_i\}_{i \in I} \), written \( \prod_{i \in I} A_i \), is defined as the set of functions \( i \mapsto x_i \) from \( I \) to \( \bigcup_{i \in I} A_i \) such that for all \( i \in I \) we have \( x_i \in A_i \). When \( I = \{1, 2, \ldots\} \) we write \( \prod_{i=1}^{\infty} A_i = A_1 \times A_2 \times \cdots \), and when \( I = \{1, \ldots, n\} \) we write \( \prod_{i=1}^{n} A_i = A_1 \times \cdots \times A_n \).

An element \( x \) of \( A_1 \times \cdots \times A_n \) is written \( (x_1, \ldots, x_n) \). If \( A_1 = \cdots = A_n = A \) we write \( A^n = A \times \cdots \times A \), and call an element \( (x_1, \ldots, x_n) \) an \( n \)-tuple of elements of \( A \).

Remark 1.39. It should be obvious how to define analogous notions of union and intersection for (unindexed) families of sets, although we must introduce the notation; for example, the union of a family \( \mathcal{F} \) is

\[
\bigcup \mathcal{F} = \bigcup_{A \in \mathcal{F}} A := \{x : \text{there exists } A \in \mathcal{F} \text{ such that } x \in A\}.
\]

If \( \{A_i\}_{i \in I} \) is an indexed family, we use the same notation \( \bigcup_{i \in I} A_i \) for the union of the indexed family as for the union of the associated family \( \{A_i : i \in I\} \); it is easy to see that the union is the same for both. Similarly for intersection.

However, it does not make sense to try to form Cartesian products of (unindexed) families of sets!

Remark 1.40. Keep in mind that although we normally use subscript notation \( x_i \) rather than ordinary function notation \( x(i) \) for values of an element \( x \) of a Cartesian product, \( x \) is still just a function.

Remark 1.41. Note that if \( A_1 \) and \( A_2 \) are sets we now have two definitions of the Cartesian product \( A_1 \times A_2 \): originally as a set of ordered pairs, and now as a certain set of functions from \( \{1, 2\} \) into \( A_1 \cup A_2 \). It is important to realize that we could not have avoided the original definition and just left everything concerning Cartesian products until now—the above general definition of Cartesian product uses the concept of function, which was defined in terms of ordered pairs. It turns out that this is a “chicken and egg” problem: we needed the concept of “ordered pair” before we could define “function”.

However, we could have omitted the original definition of union and intersection of two sets, since this is subsumed by the above definition involving arbitrary families of sets and the original definition was not used to define anything else along the way.

Definition 1.42. A family \( \mathcal{F} \) of sets is called pairwise disjoint if for all \( A, B \in \mathcal{F} \) we have either \( A = B \) or \( A \cap B = \emptyset \).

Remark 1.43. How should we define what it means for an indexed family \( \{A_i\}_{i \in I} \) of sets to be pairwise disjoint? It seems simplest to let it mean the same as for the associated family \( \{A_i : i \in I\} \). That is, we say \( \{A_i\}_{i \in I} \) is pairwise disjoint if for all \( i, j \in I \), either \( A_i = A_j \) or \( A_i \cap A_j = \emptyset \). Note that this is not the same as “either \( i = j \) or \( A_i \cap A_j = \emptyset \).
**Proposition 1.44 (Distributive Laws).** Let \( A \) be a set, and let \( \mathcal{F} \) be a family of sets. Then

\[
A \cap \bigcup_{B \in \mathcal{F}} B = \bigcup_{B \in \mathcal{F}} (A \cap B) \quad \text{and} \quad A \cup \bigcap_{B \in \mathcal{F}} B = \bigcap_{B \in \mathcal{F}} (A \cup B).
\]

**Proof.** First let \( x \in A \cap \bigcup \mathcal{F} \). Then \( x \in A \) and \( x \in \bigcup \mathcal{F} \). Thus we can choose \( B \in \mathcal{F} \) such that \( x \in B \). Since \( x \in A \), we have \( x \in A \cap B \). Hence \( x \in \bigcup_{B \in \mathcal{F}} (A \cap B) \), and we have shown \( A \cap \bigcup \mathcal{F} \subseteq \bigcup_{B \in \mathcal{F}} (A \cap B) \). These steps are reversible, so in fact \( A \cap \bigcup \mathcal{F} = \bigcup_{B \in \mathcal{F}} (A \cap B) \).

For the other part, first let \( x \in A \cup \bigcap \mathcal{F} \). Then \( x \in A \) or \( x \in \bigcap \mathcal{F} \). Case 1. \( x \in A \). For all \( B \in \mathcal{F} \), we have \( x \in A \cup B \) since \( x \in A \). Thus \( x \in \bigcap_{B \in \mathcal{F}} (A \cup B) \) in this case.

Case 2. \( x \in \bigcap \mathcal{F} \). For all \( B \in \mathcal{F} \), we have \( x \in A \cup B \) since \( x \in B \). Thus \( x \in \bigcap_{B \in \mathcal{F}} (A \cup B) \) in this case as well.

We have shown that if \( x \in A \cup \bigcap \mathcal{F} \) then \( x \in \bigcap_{B \in \mathcal{F}} (A \cup B) \), and therefore \( A \cup \bigcap \mathcal{F} \subseteq \bigcap_{B \in \mathcal{F}} (A \cup B) \). For the opposite containment, let \( x \in \bigcap_{B \in \mathcal{F}} (A \cup B) \). To show \( x \in A \cup \bigcap \mathcal{F} \), we must show either \( x \in A \) or \( x \in \bigcap \mathcal{F} \). Equivalently, we must show that if \( x \notin A \) then \( x \notin \bigcap \mathcal{F} \). Assume \( x \notin A \), and let \( B \in \mathcal{F} \). Since \( x \in \bigcap_{B \in \mathcal{F}} (A \cup B) \), we have \( x \in A \cup B \). By assumption, \( x \notin A \), so we must have \( x \notin B \). We have shown that \( x \in \bigcap_{B \in \mathcal{F}} (A \cup B) \), as desired, and this concludes the proof that \( \bigcap_{B \in \mathcal{F}} (A \cup B) \subseteq A \cup \bigcap \mathcal{F} \). \( \Box \)

**Proposition 1.45 (De Morgan’s Laws).** Let \( \mathcal{F} \) be a family of sets. Then

\[
\left( \bigcup_{A \in \mathcal{F}} A \right)^c = \bigcap_{A \in \mathcal{F}} A^c \quad \text{and} \quad \left( \bigcap_{A \in \mathcal{F}} A \right)^c = \bigcup_{A \in \mathcal{F}} A^c.
\]

**Proof.** First let \( x \in \left( \bigcup_{A \in \mathcal{F}} A \right)^c \). Then \( x \notin \bigcup_{A \in \mathcal{F}} A \), so it is false that there exists \( A \in \mathcal{F} \) such that \( x \in A \). Thus for all \( A \in \mathcal{F} \) we have \( x \notin A \), hence \( x \in A^c \). We have shown \( x \in \bigcap_{A \in \mathcal{F}} A^c \), therefore \( \left( \bigcup_{A \in \mathcal{F}} A \right)^c \subseteq \bigcap_{A \in \mathcal{F}} A^c \). For the opposite containment, just reverse the above steps.

For the other part, first let \( x \notin \left( \bigcap_{A \in \mathcal{F}} A \right)^c \). Then \( x \notin \bigcap_{A \in \mathcal{F}} A \), so it is false that for all \( A \in \mathcal{F} \) we have \( x \in A \). Thus we can choose \( A \in \mathcal{F} \) such that \( x \notin A \), hence \( x \in A^c \). We have shown \( x \in \bigcup_{A \in \mathcal{F}} A^c \), therefore \( \left( \bigcap_{A \in \mathcal{F}} A \right)^c \subseteq \bigcup_{A \in \mathcal{F}} A^c \). For the opposite containment, just reverse the above steps. \( \Box \)

**Remark 1.46.** In the statement of the above proposition, it is tacitly understood that there is some “universe” \( U \) with respect to which the complements are formed.
Proposition 1.47. Let $f: A \to B$, and let $\mathcal{F}$ be a family of subsets of $A$ and $\mathcal{G}$ a family of subsets of $B$. Then

(i) \[ f \left( \bigcup_{C \in \mathcal{F}} C \right) = \bigcup_{C \in \mathcal{F}} f(C), \]

(ii) \[ f \left( \bigcap_{C \in \mathcal{F}} C \right) \subseteq \bigcap_{C \in \mathcal{F}} f(C), \]

(iii) \[ f^{-1} \left( \bigcup_{D \in \mathcal{G}} D \right) = \bigcup_{D \in \mathcal{G}} f^{-1}(D), \]

and

(iv) \[ f^{-1} \left( \bigcap_{D \in \mathcal{G}} D \right) = \bigcap_{D \in \mathcal{G}} f^{-1}(D) \]

Proof. (i). Let $y \in f \left( \bigcup_{C \in \mathcal{F}} C \right)$. Choose $x \in \bigcup_{C \in \mathcal{F}} C$ such that $y = f(x)$, then choose $C \in \mathcal{F}$ such that $x \in C$. We have $y \in f(C)$ since $y = f(x)$. Therefore $y \in \bigcup_{C \in \mathcal{F}} f(C)$. The converse direction is proven by essentially reversing the steps. More precisely, if $y \in \bigcup_{C \in \mathcal{F}} f(C)$ then we can choose $C \in \mathcal{F}$ such that $y \in f(C)$, then we can choose $x \in C$ such that $y = f(x)$, giving $y \in f \left( \bigcup_{C \in \mathcal{F}} C \right)$ since $x \in \bigcup_{C \in \mathcal{F}} C$.

(ii). Let $y \in f \left( \bigcap_{C \in \mathcal{F}} C \right)$. Choose $x \in \bigcap_{C \in \mathcal{F}} C$ such that $y = f(x)$. For each $C \in \mathcal{F}$ we have $x \in C$ and $y = f(x)$, hence $y \in f(C)$. Therefore $y \in \bigcap_{C \in \mathcal{F}} f(C)$.

(iii). Let $x \in f^{-1} \left( \bigcup_{D \in \mathcal{G}} D \right)$. Then $f(x) \in \bigcup_{D \in \mathcal{G}} D$, so we can choose $D \in \mathcal{G}$ such that $f(x) \in D$. We have $x \in f^{-1}(D)$, hence $x \in \bigcup_{D \in \mathcal{G}} f^{-1}(D)$. These steps are reversible, so $f^{-1} \left( \bigcup_{D \in \mathcal{G}} D \right) = \bigcup_{D \in \mathcal{G}} f^{-1}(D)$.

(iv). Let $x \in f^{-1} \left( \bigcap_{D \in \mathcal{G}} D \right)$. Then $f(x) \in \bigcap_{D \in \mathcal{G}} D$. For each $D \in \mathcal{G}$ we have $f(x) \in D$, hence $x \in f^{-1}(D)$. Therefore $x \in \bigcap_{D \in \mathcal{G}} f^{-1}(D)$. These steps are reversible, so $f^{-1} \left( \bigcap_{D \in \mathcal{G}} D \right) = \bigcap_{D \in \mathcal{G}} f^{-1}(D)$. \hfill \Box

Remark 1.48. In the proof of (ii) above, the steps are not reversible, because if we take $y \in \bigcap_{C \in \mathcal{F}} f(C)$, then we only know that for every $C \in \mathcal{F}$ there exists $x \in C$ such that $y = f(x)$—we cannot conclude that there is a single $x$ which works for every $C$.

2. The Real Numbers

Remark 2.1. The real numbers form the foundation of mathematical analysis. Although the real numbers can be constructed starting only from the counting numbers $1, 2, \ldots$, we shall not do this; instead we content ourselves with a careful listing of the properties of the real numbers. However, it is important to know that such a construction is possible, and that moreover the following axioms characterize the real numbers—this means that every real number system is just a relabelling of any other one.
MAT 371 LECTURES

The axioms fall into three categories: (1) the Field Axioms, listing the essential arithmetic properties, (2) the Order Axioms, listing the essential properties of inequalities, and (3) the all-important Completeness Axiom, guaranteeing that the “real number line” has no gaps.

**Axiom 2.2** (Field Axioms). There exist functions \( (x, y) \mapsto x + y \) and \( (x, y) \mapsto xy \) from \( \mathbb{R}^2 \) to \( \mathbb{R} \) such that:

(i) for all \( x, y, z \in \mathbb{R} \) we have \( (x + y) + z = x + (y + z) \) and \( (xy)z = x(yz) \);
(ii) for all \( x, y \in \mathbb{R} \) we have \( x + y = y + x \) and \( xy = yx \);
(iii) for all \( x, y, z \in \mathbb{R} \) we have \( x(y + z) = xy + xz \);
(iv) there exists \( 0 \in \mathbb{R} \) such that for all \( x \in \mathbb{R} \) we have \( 0 + x = x \);
(v) there exists \( 1 \in \mathbb{R} \) such that \( 1 \neq 0 \) and for all \( x \in \mathbb{R} \) we have \( 1x = x \);
(vi) for all \( x \in \mathbb{R} \) there exists \( -x \in \mathbb{R} \) such that \( x + (-x) = 0 \);
(vii) for all \( x \in \mathbb{R} \setminus \{0\} \) there exists \( x^{-1} \in \mathbb{R} \) such that \( xx^{-1} = 1 \).

**Remark 2.3.** The elements 0 and 1 of \( \mathbb{R} \) are unique. Also, for each \( x \in \mathbb{R} \) the elements \(-x\) and \( x^{-1} \) (if \( x \neq 0 \)) are unique.

**Remark 2.4.** It would not have been proper to try to put the universal quantifying phrase “for all \( x, y, z \in \mathbb{R} \) before the list of items (i)–(vii), because in (iv) and (v) the phrase “for all \( x \in \mathbb{R} \)” comes after an existential quantifying phrase (“there exists … such that”), and it changes the logic to reverse the order of mixed quantifiers. Also, in each item we made a point of putting the universal quantifying phrase before the propositional expression involving the quantified variable(s), and we will (usually) continue to do so in the future, again to emphasize the order of the quantifiers. Although it is popular to write things like “there exists \( x \) such that \( P(x, y) \) for all \( y \)” or symbolically “\( \exists x P(x, y) \forall y \)” this must be interpreted with care—if it is rewritten with all the quantifiers in front, where should the \( \forall y \) go? The convention is that it should be the same as “\( \exists x \forall y P(x, y) \)” but in a complicated proposition involving many quantifiers this can cause confusion, so in these notes we try to put all quantifying phrases in front.

**Definition 2.5.** For each \( x, y \in \mathbb{R} \) define

(i) \( x - y \) := \( x + (-y) \) and
(ii) \( \frac{x}{y} := xy^{-1} \) (if \( y \neq 0 \)).

**Notation and Terminology 2.6.** (i) \( \mathbb{N} \) denotes the set of natural numbers, so that

\[
\mathbb{N} = \{1, 2, 3, \ldots \}
\]

is the smallest (where here “smaller than” means “is a subset of”) subset of \( \mathbb{R} \) satisfying both

(a) \( 1 \in \mathbb{N} \), and
(b) for all \( n \in \mathbb{N} \) we have \( n + 1 \in \mathbb{N} \).

This property of \( \mathbb{N} \) is the Principle of Mathematical Induction; this means that any subset \( A \subseteq \mathbb{N} \) satisfying both \( 1 \in A \) and for all \( n \in A \) we have \( n + 1 \in A \) must coincide with \( \mathbb{N} \).
(ii) \( \mathbb{Z} \) denotes the set of integers, so that
\[
\mathbb{Z} = \mathbb{N} \cup -\mathbb{N} \cup \{0\} = \{0, \pm 1, \pm 2, \ldots \}
\]
is the smallest subset of \( \mathbb{R} \) containing \( \mathbb{N} \) and closed under subtraction.

(iii) \( \mathbb{Q} \) denotes the set of rational numbers, so that
\[
\mathbb{Q} = \left\{ \frac{n}{k} : n, k \in \mathbb{Z}, k \neq 0 \right\}
\]
is the smallest subset of \( \mathbb{R} \) containing \( \mathbb{Z} \) and closed under division.

(iv) The set of irrational numbers is \( \mathbb{R} \setminus \mathbb{Q} \).

**Definition 2.7.** For each \( x \in \mathbb{R} \) and \( n \in \mathbb{Z} \) define
\[
x^n := \begin{cases} 
\frac{n}{x \cdot x \cdots x} & \text{if } n \in \mathbb{N} \\
1 & \text{if } n = 0 \\
\frac{1}{x^{-n}} & \text{if } -n \in \mathbb{N} \text{ and } x \neq 0 \\
1 & \text{if } n = 0 \text{ and } x \neq 0.
\end{cases}
\]

**Remark 2.8.** All the usual properties of algebra follow from Axiom 2.2. For example:

(i) for all \( x \in \mathbb{R} \) we have \( 0x = 0 \);
(ii) for all \( x \in \mathbb{R} \) we have \( -x = (-1)x \);
(iii) \((-1)^2 = 1\);
(iv) for all \( x, y \in \mathbb{R} \), if \( xy = 0 \) then either \( x = 0 \) or \( y = 0 \).

**Axiom 2.9** (Order Axioms). There exists a subset \( P \subseteq \mathbb{R} \) such that:

(i) for all \( x \in \mathbb{R} \), exactly one of \( x \in P \), \( -x \in P \), or \( x = 0 \) holds;
(ii) for all \( x, y \in P \) we have \( x + y \in P \) and \( xy \in P \).

**Definition 2.10.** For each \( x, y \in \mathbb{R} \) define

(i) \( x < y \) if \( y - x \in P \);
(ii) \( x > y \) if \( y < x \);
(iii) \( x \leq y \) if \( x < y \) or \( x = y \);
(iv) \( x \geq y \) if \( y \leq x \).

**Remark 2.11.** All the usual properties of inequalities follow from Axiom 2.9. For example, for all \( x, y, z \in \mathbb{R} \) we have:

(i) exactly one of \( x < y \), \( x > y \), or \( x = y \) holds;
(ii) if \( x < y \) then \( x + z < y + z \);
(iii) if \( z > 0 \), then \( x < y \) implies \( xz < yz \);
(iv) if \( x < y \) and \( y < z \), then \( x < z \);
(v) \( xy > 0 \) if and only if either \( x > 0 \) and \( y > 0 \) or \( x < 0 \) and \( y < 0 \);
(vi) \( xy < 0 \) if and only if either \( x > 0 \) and \( y < 0 \) or \( x < 0 \) and \( y > 0 \).

**Remark 2.12.** We must resist the temptation to regard inequalities as mundane—in fact techniques of manipulating inequalities are crucial in the development of the theory of analysis (which includes the material in this course).
Definition 2.13. For each \( x \in \mathbb{R} \) define
\[
|x| := \begin{cases} 
  x & \text{if } x \geq 0 \\
  -x & \text{if } x < 0.
\end{cases}
\]

Lemma 2.14. For all \( a, b \in \mathbb{R} \) we have:
(i) \( \pm a \leq |a| \);  
(ii) \( |a| \leq b \text{ if and only if } -b \leq a \leq b \).

Proof. (i) More precisely, we must show two inequalities: \( a \leq |a| \) and \( -a \leq |a| \). We first show \( a \leq |a| \).
Case 1. \( a \geq 0 \). Then \( a = |a| \), so \( a \leq |a| \).
Case 2. \( a < 0 \). Then \( a = -|a| \leq 0 \leq |a| \).
For the other inequality, we have \( -a \leq |a| \).
(ii) First assume \( |a| \leq b \). Then \( a \leq b \) and \( -a \leq b \). Multiplying the latter inequality by \(-1\), we get \( a \geq -b \). Combining with \( a \leq b \), we get \( -b \leq a \leq b \).
Conversely, assume \(-b \leq a \leq b \). Then \(-a \leq b \) since \(-b \leq a \). Combining with \( a \leq b \), we get \( \pm a \leq b \). Since \( |a| \) is either \( a \) or \( -a \), we have \( |a| \leq b \). \( \Box \)

Proposition 2.15 (Triangle Inequality). For all \( x, y \in \mathbb{R} \),
\[
|x + y| \leq |x| + |y| \\
|x - y| \leq |x| - |y| \quad \text{(alternate version)}.
\]

Proof. First, \( x \leq |x| \) and \( y \leq |y| \), so
\[
x + y \leq |x| + |y|.
\]
Next, \( -x \leq |x| \) and \( -y \leq |y| \), so
\[
-(x + y) \leq |x| + |y|.
\]
Thus,
\[
-(|x| + |y|) \leq x + y \leq |x| + |y|,
\]
so
\[
|x + y| \leq |x| + |y|.
\]
For the alternate version,
\[
|x| = |x - y + y| \leq |x - y| + |y|,
\]
so
\[
|x| - |y| \leq |x - y|.
\]
Then also
\[
|y| - |x| \leq |y - x| = |x - y|.
\]
Thus
\[
||x| - |y|| \leq |x - y|.
\]
\( \Box \)

Definition 2.16. (i) Given \( a, b, x \in \mathbb{R} \), we say \( x \) is between \( a \) and \( b \) if \( a \leq x \leq b \) or \( b \leq x \leq a \). Furthermore, we say \( x \) is strictly between \( a \) and \( b \) if the inequalities are strict (this means equality is not allowed).
(ii) A subset of \( \mathbb{R} \) which contains every number between any two of its elements is called an **interval**. Every interval has one of the following forms (where \( a \) and \( b \) denote real numbers with \( a \leq b \)):

(a) \( [a,b] := \{ x : a \leq x \leq b \} \) (closed, bounded);
(b) \( [a,\infty) := \{ x : a \leq x \} \) (left-half-closed, unbounded);
(c) \( (-\infty,b] := \{ x : x \leq b \} \) (right-half-closed, unbounded);
(d) \( (-\infty,\infty) := \mathbb{R} \) (open, unbounded);
(e) \( (a,b) := \{ x : a < x < b \} \) (open, bounded);
(f) \( (a,\infty) := \{ x : a < x \} \) (open, unbounded);
(g) \( (-\infty,b) := \{ x : x < b \} \) (open, unbounded);
(h) \( [a,b) := \{ x : a \leq x < b \} \) (left-half-closed or right-half-open, bounded);
(i) \( (a,b) := \{ x : a < x \leq b \} \) (left-half-open or right-half-closed, bounded).

(iii) An interval \([a,b], (a,b), [a,b), (a,b)\) is called **degenerate** if \( a = b \) and **nondegenerate** if \( a < b \). We often assume without comment that our intervals are nondegenerate.

**Remark 2.17.** Note that if \( a = b \) then the closed interval \([a,b]\) is the **singleton** (which means a set containing exactly one element) \( \{a\} \), while the intervals \((a,b), [a,b)\), and \([a,b]\) are empty.

**Definition 2.18.** Let \( A \subseteq \mathbb{R} \) and \( x \in A \). Then \( x \) is called a **maximum** of \( A \) if for all \( y \in A \) we have \( y \leq x \). Such an \( x \) is unique if it exists, and is denoted \( \max A \). Similarly for **minimum** and \( \min A \).

**Remark 2.19.** It is easy to see that \( \min A = -\max(-A) \) (where \( -A := \{-x : x \in A\} \)). Using this and the elementary properties of inequalities, from any fact concerning maxima we can more-or-less automatically deduce a corresponding fact concerning minima.

**Remark 2.20.** If \( A \) is finite (and you can look in the next section for the definition of this!) and nonempty, then \( \max A \) exists. Similarly for \( \min A \).

**Theorem 2.21** (Well-Ordering Principle). Every nonempty subset of \( \mathbb{N} \) has a minimum.

**Proof.** Suppose \( A \subseteq \mathbb{N} \) has no minimum. Then \( 1 \notin A \), otherwise \( 1 = \min A \) since \( 1 = \min \mathbb{N} \). Let \( n \in \mathbb{N} \), and assume \( 1,2,\ldots,n \notin A \). Then \( n+1 \notin A \), otherwise \( n+1 = \min A \) since \( n+1 = \min(\mathbb{N} \setminus \{1,2,\ldots,n\}) \). By induction, for all \( n \in \mathbb{N} \) we have \( n \notin A \). Thus \( A = \emptyset \). \( \square \)

**Definition 2.22.** Let \( A \subseteq \mathbb{R} \).

(i) A real number \( x \) is called an **upper bound** of \( A \) if for all \( y \in A \) we have \( y \leq x \).

(ii) \( A \) is called **bounded above** if it has an upper bound.

(iii) Similarly for **lower bound** and **bounded below**. Also similarly elsewhere.

**Remark 2.23.** If \( A \) is bounded above, then the set of upper bounds of \( A \) is an interval which is unbounded to the right.
Trivially, every real number is simultaneously an upper and a lower bound for \( \emptyset \).

As with max and min, it is easy to see that \( x \) is a lower bound of \( A \) if and only if \(-x\) is an upper bound of \(-A\), consequently any fact concerning upper bounds gives more-or-less automatically a corresponding fact concerning lower bounds.

**Remark 2.24.** The following two statements are very easy generalizations of the Well Ordering Principle: (1) every nonempty set of integers which is bounded below has a minimum, and (2) every nonempty set of integers which is bounded above has a maximum. In fact, when these facts are invoked, we just say “by the Well Ordering Principle”.

**Definition 2.25.** For all nonnegative integers \( n, k \) with \( k \leq n \) define

\[
\binom{n}{k} := \frac{n!}{k!(n-k)!},
\]

where

\[
n! := \begin{cases} 1 \cdot 2 \cdots n & \text{if } n > 0 \\ 1 & \text{if } n = 0. \end{cases}
\]

**Proposition 2.26** (Pascal’s Triangle). With the above notation, if \( k > 0 \) then

\[
\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}.
\]

**Proof.** We have

\[
\binom{n}{k-1} + \binom{n}{k} = \frac{n!}{(k-1)!(n-(k-1))!} + \frac{n!}{k!(n-k)!}
\]

\[
= n! \left( \frac{1}{(k-1)!(n-(k-1))!} + \frac{1}{k!(n-k)!} \right)
\]

\[
= \frac{n!(k + (n - k + 1))}{k!(n-k+1)!}
\]

\[
= \frac{n!(n+1)}{k!(n+1-k)!}
\]

\[
= \frac{(n+1)!}{k!(n+1-k)!}
\]

\[
= \binom{n+1}{k}.
\]

\[\square\]

**Theorem 2.27** (Binomial Theorem). For all \( a, b \in \mathbb{R} \) and every nonnegative integer \( n \),

\[
(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.
\]

**Proof.** First, if \( n = 0 \) the left hand side is

\[
(a + b)^0 = 1,
\]

\[\square\]
while the right hand side is
\[
\sum_{k=0}^{0} \binom{0}{k} a^{0-k} b^{k} = \binom{0}{0} a^{0} b^{0} = 1,
\]
so the equation is true.

Next, let \( n \) be a nonnegative integer, and assume
\[
(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.
\]
Then
\[
(a + b)^{n+1} = (a + b)(a + b)^n = (a + b) \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k
\]
\[
= \sum_{k=0}^{n} \binom{n}{k} a^{n-k+1} b^k + \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^{k+1}
\]
\[
= a^{n+1} + \sum_{k=1}^{n} \binom{n}{k} a^{n-k+1} b^k + \sum_{k=0}^{n-1} \binom{n}{k} a^{n-k} b^{k+1} + b^{n+1}
\]
\[
= a^{n+1} + \sum_{k=1}^{n} \left( \binom{n}{k-1} + \binom{n}{k} \right) a^{n-k+1} b^k + b^{n+1}
\]
\[
= a^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} a^{n+1-k} b^k + b^{n+1}
\]
\[
= \sum_{k=0}^{n+1} \binom{n+1}{k} a^{n+1-k} b^k,
\]
so the desired equation holds for \( n + 1 \). By induction, the equation holds for every nonnegative integer \( n \).

\[\square\]

**Corollary 2.28** (Bernoulli’s Inequality). For all \( a \geq 0 \) and \( n \in \mathbb{N} \),
\[
(1 + a)^n \geq 1 + na.
\]

*Proof.* By the Binomial Theorem, the left hand side is a sum of nonnegative terms, and the right hand side comprises the first two terms.

\[\square\]

**Definition 2.29.** Let \( A \subseteq \mathbb{R} \).

(i) An upper bound \( x \) of \( A \) is called a *supremum* of \( A \) if for every upper bound \( y \) of \( A \) we have \( x \leq y \). Such an \( x \) is unique if it exists, and is denoted \( \text{sup} A \).

(ii) Similarly for *infimum* and \( \text{inf} A \). Also similarly elsewhere.
Remark 2.30. (i) Thus, if \( \sup A \) exists it is the minimum of the set of upper bounds of \( A \); indeed, the set of upper bounds of \( A \) is precisely \([\sup A, \infty)\).

(ii) \( A \) has a maximum if and only if \( A \) has a supremum and \( \sup A \in A \), in which case \( \max A = \sup A \). Similarly for min and \( \inf \).

(iii) We have \( \sup(0, 1) = 1 \) and \( 1 \notin (0, 1) \).

(iv) If \( A \) has a supremum then \( A \) is bounded above, but not conversely. The empty set gives a counterexample: we have already noticed that every real number is an upper bound for \( \emptyset \), which means the set of upper bounds of \( \emptyset \) is \( \mathbb{R} \), and of course \( \mathbb{R} \) has no minimum.

(v) Thus, if \( \sup A \) exists then \( A \) is nonempty and bounded above. The last axiom of the real numbers, certainly the deepest (indeed, its crucial role in analysis was not recognized until the 19th century), guarantees that there are no other obstructions to the existence of a supremum:

Axiom 2.31 (Completeness Axiom). Every nonempty subset of \( \mathbb{R} \) which is bounded above has a supremum.

Remark 2.32. As with \( \max \) and \( \min \), it is easy to see that if \( A \subseteq \mathbb{R} \) is nonempty and bounded below then \( A \) has an infimum, and moreover \( \inf A = -\sup(-A) \), consequently any fact concerning suprema gives more-or-less automatically a corresponding fact concerning infima.

Remark 2.33. If \( A \subseteq \mathbb{Z} \) is nonempty and bounded above, we do not need the Completeness Axiom to know \( A \) has a supremum—the Well-Ordering Principle tells that in fact \( A \) has a maximum.

Proposition 2.34 (Approximation of Suprema). Let \( A \subseteq \mathbb{R} \) be nonempty and bounded above. Then \( \sup A \) is the unique number \( s \) satisfying both

(i) for all \( t > s \) and all \( x \in A \) we have \( x < t \);

(ii) for all \( t < s \) there exists \( x \in A \) such that \( t < x \).

Proof. We must prove two things: that \( \sup A \) satisfies properties (i)-(ii), and that there is at most one real number satisfying (i)-(ii).

First let \( t > \sup A \). Then \( t \) is an upper bound of \( A \) and is bigger than the upper bound \( \sup A \), so for all \( x \in A \) we have \( x < t \). We have shown \( \sup A \) satisfies (i). For (ii), let \( t < \sup A \). Then \( t \) is not an upper bound of \( A \), so there exists \( x \in A \) such that \( t < x \).

For the uniqueness, let \( s_1, s_2 \in \mathbb{R} \) both satisfy (i)-(ii), and suppose \( s_1 \neq s_2 \). Without loss of generality \( s_1 < s_2 \). Then letting \( s = s_2 \) in (ii) we see that we can choose \( x \in A \) such that \( s_1 < x \). But then letting \( s = s_1 \) and \( t = x \) in (i) we see that for all \( y \in A \) we have \( y < x \), which is a contradiction since \( x \in A \). \qed

Theorem 2.35 (Archimedean Principle). For all \( x \in \mathbb{R} \) there exists \( n \in \mathbb{N} \) such that \( n > x \).
Proof. Suppose not. Then there exists \( x \in \mathbb{R} \) such that for all \( n \in \mathbb{N} \) we have \( n \leq x \). In particular, \( \mathbb{N} \) is bounded above. Of course \( \mathbb{N} \neq \emptyset \), so by the Completeness Axiom \( \sup \mathbb{N} \) exists. Then by Approximation of Suprema there exists \( n \in \mathbb{N} \) such that \( n > \sup \mathbb{N} - 1 \). But then \( n + 1 \in \mathbb{N} \) and \( n + 1 > \sup \mathbb{N} \), giving a contradiction. \( \square \)

Remark 2.36. More generally (and this follows very easily from the Archimedean Principle), for every real number \( x \) there exist integers \( k \) and \( n \) such that \( k < x < n \), and when this fact is invoked we just say “by the Archimedean Principle”.

Corollary 2.37 (Floor Theorem). For all \( x \in \mathbb{R} \) there exists a unique \( n \in \mathbb{Z} \) such that
\[
    n \leq x < n + 1.
\]

Proof. Put \( A = \{ k \in \mathbb{Z} : k \leq x \} \). By the (generalized version of the) Archimedean Principle, the set \( A \) is nonempty. Moreover, by its definition, the set \( A \) has \( x \) as an upper bound. Thus \( A \) is a nonempty set of integers which is bounded above. Then by the (generalized version of the) Well-Ordering Principle, \( A \) has a maximum, which we call \( n \). So, by construction \( n \) is the largest integer less than or equal to \( x \). In particular, \( n + 1 \) must be greater than \( x \) since \( n + 1 \) is an integer larger than \( n \). Thus we have \( n \leq x < n + 1 \).

For the uniqueness, suppose \( k \in \mathbb{Z} \), \( k \leq x < k + 1 \), and \( k \neq n \). Without loss of generality \( n < k \). Then \( n + 1 \leq k \) since \( n \) and \( k \) are integers. But then we have
\[
    x < n + 1 \leq k \leq x,
\]
a contradiction. \( \square \)

Remark 2.38. We called the above result the “Floor Theorem” because the unique integer \( n \) is popularly called the “floor” of \( x \).

Theorem 2.39 (Density of Rationals). Strictly between any two distinct real numbers there exists a rational number.

Proof. Let \( a < b \). Then \( b - a > 0 \), so by the Archimedean Principle we can choose \( n \in \mathbb{N} \) such that \( n > \frac{1}{b - a} \). Then \( \frac{1}{n} < b - a \). By the Floor Theorem we can choose an integer \( k \) such that \( k - 1 \leq na < k \). Then we have
\[
    a < \frac{k}{n} \leq a + \frac{1}{n} < b,
\]
so \( k/n \) is a rational number strictly between \( a \) and \( b \). \( \square \)

Remark 2.40. The Completeness Axiom is what allows for decimal expansions of real numbers: Given \( x \geq 0 \), first define
\[
    n = \max \{ l \in \mathbb{Z} : l \leq x \},
\]
which exists by the Well-Ordering and the Archimedean Principles. Next, define $d_1, d_2, \ldots$ by

$$d_1 = \max \left\{ l \in \mathbb{Z} : n + \frac{l}{10} \leq x \right\},$$

$$d_2 = \max \left\{ l \in \mathbb{Z} : n + \frac{d_1}{10} + \frac{l}{10^2} \leq x \right\},$$

and continuing inductively. Note that each $d_i$ is an integer between 0 and 9, inclusive. Put

$$A = \left\{ n + \sum_{i=1}^{j} \frac{d_i}{10^i} : j \in \mathbb{N} \right\}.$$

Then $x = \sup A$, and we say

$$x = n.d_1d_2 \cdots$$

is a decimal expansion of $x$.

On the other hand, given a nonnegative integer $n$ and integers $d_1, d_2, \ldots$ between 0 and 9, again put

$$A = \left\{ n + \sum_{i=1}^{j} \frac{d_i}{10^i} : j \in \mathbb{N} \right\}.$$

Then $A$ is nonempty and bounded above, so we can put $x = \sup A$, and again we say $x = n.d_1d_2 \cdots$ is a decimal expansion of $x$.

If $x < 0$ we find a decimal expansion $n.d_1d_2 \cdots$ of $|x|$, and we say $x = -n.d_1d_2 \cdots$ is a decimal expansion of $x$.

We recall (without proof) some elementary properties of decimal expansions:

(i) $\mathbb{R}$ coincides with the set of all decimal expansions. That is, not only does every real number have a decimal expansion, but every decimal expansion determines a real number.

(ii) The decimal expansion of a real number $x$ is unique unless $x$ is of the form $\frac{m}{10^n}$ for some integers $n$ and $k$, in which case $x$ has two decimal expansions, one ending in 0's and the other ending in 9's.

(iii) A real number $x$ is rational if and only if its decimal expansion (more precisely, each of its decimal expansions) is eventually repeating (that is, there is a string of digits which repeats forever starting at some point in the decimal expansion). Thus, for example, the number $0.1010010001 \cdots$ is irrational since its decimal expansion does not repeat.

**Remark 2.41.** In the above preceding remark we defined $d_1, d_2, \ldots$ inductively, that is, the choice of each successive $d_n$ depended upon the choices of the preceding $d_k$'s for $k < n$. The justification for this is actually a subtle point of set theory (because we are making “infinitely many choices”), but we will allow ourselves to do this willy-nilly.
**Corollary 2.42** (Density of Irrationals). *Strictly between any two distinct real numbers there exists an irrational number.*

**Proof.** Let \( a < b \), and pick any positive irrational number \( x \). Then \( ax < bx \), so by Density of Rationals we can choose \( y \in \mathbb{Q} \) such that \( ax < y < bx \). Without loss of generality, \( y \neq 0 \). Then

\[
a < \frac{y}{x} < b,
\]

and \( \frac{y}{x} \) is irrational since \( y \) is rational and \( x \) is irrational. \( \square \)

**Remark 2.43.** The Completeness Axiom is also what guarantees the existence of roots: for all \( x \in \mathbb{R} \) and \( n \in \mathbb{N} \),

(i) if \( n \) is even and \( x \geq 0 \), then there exists a unique \( y \geq 0 \) such that \( y^n = x \), while

(ii) if \( n \) is odd then there exists a unique \( y \in \mathbb{R} \) such that \( y^n = x \).

The justification of these facts is deferred until after the Intermediate Value Theorem (and we will not run the risk of circular reasoning since we will only use roots in examples until then). In both of the above cases, \( y \) is called the \( n \)th *root* of \( x \), denoted \( y = x^{1/n} \). Then for each \( n \in \mathbb{N} \) and \( k \in \mathbb{Z} \) define

\[
x^{k/n} = (x^{1/n})^k \quad \text{(where } x \geq 0 \text{ if } n \text{ is even}).
\]

It is natural to ask “is \( x^{k/n} \) also equal to \((x^k)^{1/n}\)?”, and fortunately the answer is “yes”, at least in the cases where everything makes sense. More generally, it can be proven by a tedious algebraic argument that the usual Laws of Exponents hold for rational powers, that is, for all \( r, s \in \mathbb{Q} \) and \( x, y \in \mathbb{R} \) we have:

(i) \( x^r x^s = x^{r+s} \);

(ii) \( \frac{x^r}{x^s} = x^{r-s} \);

(iii) \( (x^r)^s = x^{rs} \);

(iv) \( (xy)^r = x^r y^r \)

(as long as we don’t try to divide by 0 or take an even root of a negative number). Since we will not need these in the formal development of the theory, we do not prove them here. Much later (after the material on integration) we will have more powerful machinery allowing us to handle even more general exponents and prove the Laws of Exponents without resorting to mundane algebraic manipulation.

Nevertheless, it is interesting to observe that rational powers tend to give irrational numbers, in the following precise sense: for all \( k, n \in \mathbb{N} \), if \( k^{1/n} \) is not an integer, then it is irrational. We only prove the case \( k = n = 2 \); the general case can be proved using the Fundamental Theorem of Arithmetic. Since we have not yet proven the existence of square roots, we must state the formal result as follows:

**Proposition 2.44.** For all \( x \in \mathbb{R} \), if \( x^2 = 2 \) then \( x \) is irrational.
MAT 371 LECTURES

Proof. We argue by contradiction. Let $x \in \mathbb{R}$, and suppose $x^2 = 2$ and $x$ is rational. Then there exist $p, q \in \mathbb{Z}$ such that $x = \frac{p}{q}$, and without loss of generality $p$ and $q$ are not both even. Then

$$2q^2 = p^2.$$ 

Thus $p^2$ is even, hence $p$ itself is even. There exists $r \in \mathbb{Z}$ with $p = 2r$. Then

$$2q^2 = 4r^2,$$

so $q^2 = 2r^2$. But then $q^2$, hence $q$, is even, giving a contradiction. \qed

3. Countability

Definition 3.1. A set $A$ is called:

(i) finite if either $A = \emptyset$ or for some $n \in \mathbb{N}$ there exists a 1-1 function from $\{1, \ldots, n\}$ onto $A$;

(ii) infinite if $A$ is not finite;

(iii) countable if $A$ is finite or there exists a 1-1 function from $\mathbb{N}$ onto $A$;

(iv) countably infinite if $A$ is countable and infinite;

(v) uncountable if $A$ is not countable.

Remark 3.2. We adopt the point of view that we understand finite sets pretty well, although strictly speaking the familiar properties of finite sets should be proven, which would involve routine induction arguments. Infinite sets, on the other hand, have some amazing properties (for example, $A$ is infinite if and only if there exists a 1-1 function from $A$ onto a proper subset of $A$), and have mystified a lot of smart people for a long time. The modern theory of infinite sets was only worked out during the late 19th and early 20th centuries. Here we will only need to know the difference between infinite sets which are “manageable” (countable) and which are “unimaginably huge” (uncountable).

One fairly obvious fact is that if there is a 1-1 function from $A$ onto $B$, then $A$ is countable if and only if $B$ is.

Lemma 3.3. For every nonempty set $A$, the following are equivalent:

(i) $A$ is countable;

(ii) there exist a countable set $B$ and a function from $B$ onto $A$;

(iii) there exists a function from $\mathbb{N}$ onto $A$;

(iv) there exists a 1-1 function from $A$ to $\mathbb{N}$;

(v) there exist a countable set $B$ and a 1-1 function from $A$ to $B$.

Proof. (i) $\Rightarrow$ (ii). This is trivial, since we could take the identity function on $A$.

(ii) $\Rightarrow$ (iii). Assume $B$ is countable and $f : B \to A$ is onto. Then $B \neq \emptyset$.

Case 1. $B$ is finite. For some $n \in \mathbb{N}$ we can choose a 1-1 onto function $g : \{1, \ldots, n\} \to B$. Define $h : \mathbb{N} \to \{1, \ldots, n\}$ by

$$h(k) = g(\text{min}\{k, n\}).$$

Then $h$ is onto, so $f \circ g \circ h : \mathbb{N} \to A$ is onto.
Case 2. $B$ is infinite. Choose a 1-1 onto function $g: \mathbb{N} \rightarrow B$. Then $f \circ g: \mathbb{N} \rightarrow A$ is onto.

(iii) $\Rightarrow$ (iv). Given an onto function $f: \mathbb{N} \rightarrow A$, define $g: A \rightarrow \mathbb{N}$ by

$$g(x) = \min f^{-1}(\{x\}).$$

Note that $g$ is well defined by the Well-Ordering Principle, since onto-ness of $f$ means that for every $x \in A$ the pre-image $f^{-1}(\{x\})$ is nonempty. Then $g$ is 1-1 since if $x \neq y$ then $\{x\}$ and $\{y\}$ are disjoint, hence so are $f^{-1}(\{x\})$ and $f^{-1}(\{y\})$, thus these latter two sets must have distinct minima.

(iv) $\Rightarrow$ (v). This is trivial, since $\mathbb{N}$ is countable.

(v) $\Rightarrow$ (i). By (i) $\Rightarrow$ (iv), without loss of generality $A \subseteq \mathbb{N}$.

Case 1. $A$ is finite. Then $A$ is countable.

Case 2. $A$ is infinite. Define $f: \mathbb{N} \rightarrow A$ inductively by

$$f(n) = \begin{cases} 
\min A & \text{if } n = 1 \\
\min A \setminus f(\{1, \ldots, n - 1\}) & \text{if } n > 1.
\end{cases}$$

Then $f$ is 1-1 onto. \hfill \Box

**Theorem 3.4.** The following sets are countable:

(i) $\mathbb{N}^2$;

(ii) every countable union of countable sets;

(iii) every finite product of countable sets;

(iv) $\mathbb{Z}$;

(v) $\mathbb{Z}^2$;

(vi) $\mathbb{Q}$.

**Proof.** (i) Define $f: \mathbb{N}^2 \rightarrow \mathbb{N}$ by $f(n, k) = 2^n 3^k$. Then $f$ is 1-1 by the Fundamental Theorem of Arithmetic, so $\mathbb{N}^2$ is countable.

(ii) More precisely, by a “countable union of countable sets” we mean a union $\bigcup \mathcal{F}$ where the $\mathcal{F}$ is countable and for every $A \in \mathcal{F}$ the set $A$ is countable. Without loss of generality $\mathcal{F}$ and every $A \in \mathcal{F}$ are nonempty, because $\mathcal{F} = \emptyset$ makes the whole union empty, while an empty element of $\mathcal{F}$ has no effect on the union. Let $g: \mathbb{N} \rightarrow \mathcal{F}$ be onto, and for each $A \in \mathcal{F}$ let $h_A: \mathbb{N} \rightarrow A$ be onto. Define $f: \mathbb{N}^2 \rightarrow \bigcup \mathcal{F}$ by

$$f(n, k) = h_{g(n)}(k).$$

Then $f$ is onto, so $\bigcup \mathcal{F}$ is countable since $\mathbb{N}^2$ is.

(iii) More precisely, by a “finite product of countable sets” we mean a Cartesian product $\prod_{i \in I} A_i$ where the index set $I$ is finite and each coordinate set $A_i$ is countable. Without loss of generality $I \neq \emptyset$, otherwise the product has exactly one element (because there is exactly one function whose domain is the empty set, namely the empty set of ordered pairs). Then without loss of generality $I = \{1, \ldots, n\}$ for some $n \in \mathbb{N}$. Moreover without loss of generality $n > 1$, otherwise the product is just $A_1$, which is countable by
hypothesis. Then we have
\[ \prod_{i=1}^{n} A_i = \bigcup_{x \in A_n} \left( \prod_{i=1}^{n-1} A_i \times \{x\} \right) \]
(or, more precisely, there is an obvious 1-1 function from the left hand set onto the right hand set), so the result follows by induction.

(iv) We have \( \mathbb{Z} = \mathbb{N} \cup -\mathbb{N} \cup \{0\} \), a finite union of countable sets.

(v) This is a finite product of countable sets.

(vi) Define \( f: \mathbb{Z}^2 \to \mathbb{Q} \) by
\[ f(n, k) = \begin{cases} \frac{n}{k} & \text{if } k \neq 0 \\ 0 & \text{if } k = 0. \end{cases} \]

Then \( f \) is onto, so \( \mathbb{Q} \) is countable since \( \mathbb{Z}^2 \) is. \( \square \)

**Theorem 3.5.** The following sets are uncountable:

(i) every nondegenerate interval;

(ii) \( \mathbb{R} \);

(iii) the irrational numbers.

**Proof.**

(i) It suffices to show \([0,1]\) is uncountable (because every nondegenerate interval contains an image of \([0,1]\) under a 1-1 function, and a set containing an uncountable subset must itself be uncountable), and moreover it suffices to show the subset
\[ A := \{a_1 a_2 \cdots : \text{for all } i \in \mathbb{N} \text{ we have } a_i = 1 \text{ or } 2\} \]

of \([0,1]\) (chosen to avoid the nuisance of ambiguous decimal expansions) is uncountable. Let \( f: \mathbb{N} \to A \). We show \( f \) is not onto, which suffices. For each \( n \in \mathbb{N} \) let
\[ f(n) = a_1 a_2 a_3 \cdots . \]

Define \( b = b_1 b_2 \cdots \in A \) by
\[ b_i = \begin{cases} 1 & \text{if } a_{i+1} = 2 \\ 2 & \text{if } a_{i+1} = 1. \end{cases} \]

Then \( b \) differs from each \( f(n) \) in the \( n \)th decimal place, so \( b \notin \text{ran } f \).

(ii) This follows immediately from (i), again since \( \mathbb{R} \) contains the uncountable subset \([0,1]\).

(iii) We have \( \mathbb{R} = \mathbb{Q} \cup \mathbb{Q}^c \), and \( \mathbb{Q} \) is countable, so \( \mathbb{Q}^c \) must be uncountable since \( \mathbb{R} \) is. \( \square \)

4. Sequences

**Definition 4.1.**

(i) A **sequence** is a function with domain \( \mathbb{N} \).

(ii) If \( x \) is a sequence, then for each \( n \in \mathbb{N} \) the \( n \)th **term** of \( x \) is defined as
\[ x_n := x(n). \]
Notation and Terminology 4.2. A sequence \( x \) is written \((x_n)_{n=1}^{\infty}\), or just \((x_n)\). If all the values \(x_n\) of a sequence \((x_n)\) are in a set \(S\), we say \((x_n)\) is a sequence in \(S\). Thus, the set of all sequences in \(S\) is just the Cartesian product \(\prod_{i=1}^{\infty} S\). Remember that although we use subscript notation \(x_n\), rather than ordinary function notation \(x(n)\), a sequence is just a function.

Standing Hypothesis 4.3. All our sequences will be in \(\mathbb{R}\) unless otherwise specified.

Remark 4.4. Actually, it is sometimes convenient to allow sequences to “start somewhere other than 1”, that is, to allow the domain to be of the form \(\{n \in \mathbb{Z} : n \geq k\}\) for some \(k \in \mathbb{Z}\), and in fact \(k = 0\) is frequently used. The essential properties of sequences do not depend upon the “starting point”, and for convenience we develop the general theory for sequences starting at 1.

Definition 4.5. Let \((x_n)\) be a sequence and \(x \in \mathbb{R}\). We say \((x_n)\) converges to \(x\) if for all \(\epsilon > 0\) there exists \(k \in \mathbb{N}\) such that for all \(n \in \mathbb{N}\),

\[
\text{if } n \geq k \text{ then } |x_n - x| < \epsilon.
\]

Remark 4.6. In the above \(\epsilon\)-condition for convergence, it is crucial to keep track of the dependence: the \(k\) depends upon the \(\epsilon\). More precisely, \(k\) is not uniquely determined by \(\epsilon\)—if we find one \(k\) which works, then any larger \(k\) will also work. Also, if we have a \(k\) which works for a particular \(\epsilon\), then it will also work for any larger \(\epsilon\). The important thing to remember is that if the \(\epsilon\) is decreased then we will probably have to increase the \(k\).

Notation and Terminology 4.7. When \((x_n)\) converges to \(x\), we write \(x_n \to x\) as \(n \to \infty\), or

\[
x_n \xrightarrow{n \to \infty} x,
\]

or just \(x_n \to x\) if \(n \to \infty\) is understood.

Remark 4.8. There can be at most one \(x\) for which \(x_n \to x\). This result is completely routine, but we formalize it to indicate a typical “\(\epsilon\)-argument”:

Lemma 4.9. Let \((x_n)\) be a sequence and \(x, y \in \mathbb{R}\). If \(x_n \to x\) and \(x_n \to y\), then \(x = y\).

Proof. Let \(\epsilon > 0\). Choose \(k \in \mathbb{N}\) such that if \(n \geq k\) then \(|x_n - x|, |x_n - y| < \epsilon/2\). Then

\[
|x - y| \leq |x - x_k + x_k - y| \leq |x - x_k| + |x_k - y| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\]

Since \(\epsilon > 0\) was arbitrary, we must have \(x = y\). \(\square\)

Remark 4.10. In the above proof, we used an expression of the form “\(s, t < r\)”, which is intended as an abbreviation for “\(s < r\) and \(t < r\)”. Also, we should justify how we could “choose \(k \in \mathbb{N}\) such that if \(n \geq k\) then \(|x_n - x|, |x_n - y| < \epsilon/2\)”: Since \(x_n \to x\) we can choose \(k_1 \in \mathbb{N}\) such that \(n \geq k_1\) implies \(|x_n - x| < \epsilon/2\), and similarly we can choose \(k_2 \in \mathbb{N}\) such that \(n \geq k_2\) implies \(|x_n - y| < \epsilon/2\), and then we can take \(k = \max\{k_1, k_2\}\). This sort of “choosing something to perform several (compatible) jobs simultaneously”
occurs so often that it is normally done without comment, as in the above proof.

**Remark 4.11.** By the preceding lemma, if \( x_n \to x \) then \( x \) is uniquely determined by \( (x_n) \), so it makes sense to formalize this relationship:

**Definition 4.12.** Let \( (x_n) \) be a sequence, and let \( x \in \mathbb{R} \). If \( x_n \to x \), we also say \( x \) is the limit of \( (x_n) \), written \( x = \lim_{n \to \infty} x_n \), or just \( \lim x_n \).

**Definition 4.13.** If a sequence \( (x_n) \) converges, we also say it is convergent; otherwise we say it diverges, or is divergent.

**Example 4.14.** (i) Every constant sequence converges to its constant value.

(ii) \( \frac{1}{n} \to 0 \).

**Lemma 4.15.** Let \( (x_n) \) be a sequence and \( x \in \mathbb{R} \). Then \( x_n \to x \) if and only if \( |x_n - x| \to 0 \).

**Proof.** Assume \( x_n \to x \). Given \( \epsilon > 0 \), choose \( k \in \mathbb{N} \) such that if \( n \geq k \) then \( |x_n - x| < \epsilon \). Since for every \( t \in \mathbb{R} \) we have \( |t| - 0| = |t| \), we have shown that \( |x_n - x| \to 0 \). The steps are reversible, so the desired equivalence holds. \( \square \)

**Theorem 4.16** (Squeeze Theorem). Let \( (x_n), (y_n), \) and \( (z_n) \) be sequences. Assume that for all \( n \in \mathbb{N} \) we have \( x_n \leq y_n \leq z_n \) and that \( \lim x_n = \lim z_n = x \). Then \( y_n \to x \).

**Proof.** Given \( \epsilon > 0 \), choose \( k \in \mathbb{N} \) such that if \( n \geq k \) then \( |x_n - x|, |z_n - x| < \epsilon \). Then \( n \geq k \) implies

\[
x_n - \epsilon < x_n \leq y_n \leq z_n < x + \epsilon,
\]

so \( |y_n - x| < \epsilon \). \( \square \)

**Definition 4.17.**

(i) A subset \( A \) of \( \mathbb{R} \) is called bounded if it is bounded above and below.

(ii) A real-valued function is called bounded if its range is bounded in \( \mathbb{R} \). More generally, if \( B \subseteq \text{dom } f \), we say \( f \) is bounded on \( B \) if the restriction \( f|B \) is bounded. In particular, a sequence \( (x_n) \) is called bounded if its range \( \{x_n : n \in \mathbb{N}\} \) is bounded in \( \mathbb{R} \).

**Remark 4.18.** For all \( A \subseteq \mathbb{R} \), the following are equivalent:

(i) \( A \) is bounded;

(ii) there exists \( M \in \mathbb{R} \) such that for all \( x \in A \) we have \( |x| \leq M \);

(iii) there exist \( t, M \in \mathbb{R} \) such that for all \( x \in A \) we have \( |x - t| \leq M \);

(iv) for all \( t \in \mathbb{R} \) there exists \( M \in \mathbb{R} \) such that for all \( x \in A \) we have \( |x - t| \leq M \).

**Proposition 4.19.** Every convergent sequence is bounded.

**Proof.** Let \( x_n \to x \). Choose \( k \in \mathbb{N} \) such that if \( n \geq k \) then \( |x_n - x| \leq 1 \). Put \( M = \max\{1, |x_1 - x|, \ldots, |x_{k-1} - x|\} \) (where without loss of generality \( k > 1 \)). Then for all \( n \in \mathbb{N} \) we have \( |x_n - x| \leq M \). \( \square \)
Remark 4.20. In the above proof we applied the definition of convergence with $\epsilon = 1$ to choose $k$ such that $n \geq k$ implies $|x_n - x| \leq 1$. However, the inequality given to us by the definition is the strict one: “$|x_n - x| < 1$”. In this particular case we only needed the weak inequality, which can certainly be satisfied since a strict inequality implies the corresponding weak inequality. This sort of “taking a weaker conclusion than the definition (or theorem, hypothesis, etc.) allows” is done often, since it is clearer to use only the property specifically required in the context.

Definition 4.21. Let $A \subseteq \mathbb{R}$ and $t \in \mathbb{R}$. We say $t$ is a cluster point of $A$ if for all $\epsilon > 0$ there exists $x \in A \setminus \{t\}$ such that $|x - t| < \epsilon$.

Notation and Terminology 4.22. $A'$ denotes the set of all cluster points of $A$.

Remark 4.23. In the definition of cluster point, once we specify $x \in A$ we can satisfy the requirement that $x \neq t$ by requiring $0 < |x - t|$. Thus we can rephrase the condition as “for all $\epsilon > 0$ there exists $x \in A$ such that $0 < |x - t| < \epsilon$”.

Thus, $t$ is a cluster point of $A$ if and only if we can find points of $A$ “arbitrarily close to but different from” $t$. It turns out that we can find lots of them:

Lemma 4.24. If $A \subseteq \mathbb{R}$ and $t \in A'$ then for all $\epsilon > 0$ the set $(t - \epsilon, t + \epsilon) \cap A$ is infinite.

Proof. Given $\epsilon > 0$, choose $x_1 \in A$ such that $0 < |x_1 - t| < \epsilon$, and for all $n = 2, 3, \ldots$ inductively choose $x_n \in A$ such that

$$0 < |x_n - t| < |x_{n-1} - t|.$$  

Then for each $n$ we have $|x_n - t| < |x_{1} - t| < \epsilon$, so $x_n \in (t - \epsilon, t + \epsilon)$. Also, if $n > k$ then $|x_n - t| < |x_k - t|$, so we must have $x_n \neq x_k$. Thus the sequence $(x_n)$ is 1-1, hence has infinite range. Therefore $\{x_n : n \in \mathbb{N}\}$ is an infinite subset of $A \cap (t - \epsilon, t + \epsilon)$. \qed

Proposition 4.25. Let $A \subseteq \mathbb{R}$ and $t \in \mathbb{R}$. Then $t \in A'$ if and only if there exists a sequence in $A \setminus \{t\}$ converging to $t$.

Proof. First assume $t \in A'$. For each $n \in \mathbb{N}$ choose $x_n \in A$ such that $0 < |x_n - t| < 1/n$. Then $(x_n)$ is a sequence in $A \setminus \{t\}$, and $|x_n - t| \to 0$ by the Squeeze Theorem, so $x_n \to t$.

Conversely, let $(x_n)$ be a sequence in $A \setminus \{t\}$ converging to $t$, and let $\epsilon > 0$. Choose $n \in \mathbb{N}$ such that $|x_n - t| < \epsilon$. Since $x_n \neq t$, we have $0 < |x_n - t|$, and we have shown $t \in A'$. \qed

Definition 4.26. Given sequences $(x_n)$ and $(y_k)$, we say $(y_k)$ is a subsequence of $(x_n)$ if there exist $n_1 < n_2 < \cdots \in \mathbb{N}$ such that for all $k \in \mathbb{N}$ we have $y_k = x_{n_k}$.
Remark 4.27. Thus, a subsequence of \((x_n)\) is a composition of the function \(n \mapsto x_n : \mathbb{N} \to \mathbb{R}\) with a certain function \(k \mapsto n_k : \mathbb{N} \to \mathbb{N}\). Note that the requirement \(n_1 < n_2 < \cdots\) implies that for all \(k \in \mathbb{N}\) we have \(n_k \geq k\).

Of course, \((x_n)\) is a subsequence of itself, taking \(n_k = k\) for all \(k\). A moment’s reflection reveals that a given sequence has a bewildering profusion of subsequences; certainly there is no simple way to write them all down. Nevertheless, there is some commonality among them:

**Lemma 4.28.** Let \((x_n)\) be a sequence and \(x \in \mathbb{R}\). If \((x_n)\) converges to \(x\) then every subsequence of \((x_n)\) also converges to \(x\).

**Proof.** Let \(x_n \to x\) and \(n_1 < n_2 < \cdots\) in \(\mathbb{N}\). Given \(\varepsilon > 0\), choose \(j \in \mathbb{N}\) such that if \(n \geq j\) then \(|x_n - x| < \varepsilon\). Then \(k \geq j\) implies \(n_k \geq k \geq j\), hence \(|x_{n_k} - x| < \varepsilon\). Thus \(x_{n_k} \to x\). \(\square\)

Remark 4.29. A given sequence \((x_n)\) may not converge, but it may have many convergent subsequences. It is useful to have a test for which real numbers are limits of subsequences of \((x_n)\):

**Lemma 4.30.** Let \((x_n)\) be a sequence and \(x \in \mathbb{R}\). Then the following are equivalent:

(i) \((x_n)\) has a subsequence converging to \(x\);  
(ii) either \(x\) is a cluster point of the range \(\{x_n : n \in \mathbb{N}\}\) or the set \(\{n \in \mathbb{N} : x_n = x\}\) is infinite;  
(iii) for all \(\varepsilon > 0\) and \(k \in \mathbb{N}\) there exists \(n \geq k\) such that \(|x_n - x| < \varepsilon\).

**Proof.** (i) \(\Rightarrow\) (ii). Let \((y_k)\) be a subsequence of \((x_n)\) converging to \(x\), and further assume \(x_n = x\) for only finitely many \(n\). We’ll show \(x \in (\text{ran}(x_n))'\). Let \(\varepsilon > 0\), and choose \(j \in \mathbb{N}\) such that \(k \geq j\) implies \(|y_k - x| < \varepsilon\). Then \(|x_n - x| < \varepsilon\) for infinitely many \(n\), and we can have \(|x_n - x| = 0\) for only finitely many \(n\), so there exists \(n\) such that \(0 < |x_n - x| < \varepsilon\), hence \(x\) is a cluster point of \(\text{ran}(x_n)\).

(ii) \(\Rightarrow\) (iii). Let \(\varepsilon > 0\) and \(k \in \mathbb{N}\). First assume \(x \in (\text{ran}(x_n))'\). Then the set \(\{n \in \mathbb{N} : |x_n - x| < \varepsilon\}\) is infinite, hence contains an element \(n\) such that \(n \geq k\). On the other hand, assume the set \(\{n \in \mathbb{N} : x_n = x\}\) is infinite. Then it contains an element \(n\) such that \(n \geq k\), and we certainly have \(|x_n - x| < \varepsilon\).

(iii) \(\Rightarrow\) (i). Choose \(n_1 \in \mathbb{N}\) such that \(|x_{n_1} - x| < 1\), and for each \(k = 2, 3, \ldots\) inductively choose \(n_k > n_{k-1}\) such that \(|x_{n_k} - x| < 1/k\). Then we have \(n_1 < n_2 < \cdots\), so that \((x_{n_k})\) is a subsequence. By the Squeeze Theorem, this subsequence converges to \(x\). \(\square\)

Remark 4.31. With the notation of the above lemma, if \(x\) is a cluster point of the set \(A := \{x_n : n \in \mathbb{N}\}\), it might seem that Proposition 4.25 tells us there is a subsequence of \((x_n)\) converging to \(x\); not only does it not tell us this, but in fact Proposition 4.25 gives us no information: although it tells us there is a sequence in the set \(A\) (in fact in the slightly smaller set \(A \setminus \{x\}\)) converging to \(x\), there is nothing in the statement of that lemma which gives us a subsequence of \((x_n)\).
Lemma 4.32. For all sequences \((x_n)\) and \((y_n)\), if \(x_n \to 0\) and \((y_n)\) is bounded, then \(x_n y_n \to 0\).

Proof. Choose \(M \in \mathbb{R}\) such that for all \(n \in \mathbb{N}\) we have \(|y_n| \leq M\). Without loss of generality \(M > 0\). Given \(\epsilon > 0\), choose \(k \in \mathbb{N}\) such that if \(n \geq k\) then \(|x_n| < \epsilon / M\). Then \(n \geq k\) implies
\[
|x_n y_n| \leq M |x_n| < \epsilon.
\]

\(\square\)

Remark 4.33. In the above proof we imposed a further condition on \(M\), namely positivity. First of all, there really was no loss of generality: of course we have \(M \geq 0\) since it is an upper bound for \(\{|y_n| : n \in \mathbb{N}\}\); if we happened to initially choose \(M = 0\) (which would only be possible if the sequence \((y_n)\) were identically \(0\)), we could swap it for a larger \(M\) and still have an upper bound. Anyway, the reason we imposed the further condition \(M > 0\) was so that \(\epsilon / M\) would be defined. An alternative trick would be to use \(\epsilon / (M + 1)\), which makes sense no matter what our initial choice of \(M\) was. Then we would arrive at the inequalities
\[
|x_n y_n| \leq M |x_n| \leq \frac{M \epsilon}{M + 1} < \epsilon.
\]

Proposition 4.34. For all convergent sequences \((x_n)\) and \((y_n)\),

(i) \(\lim (x_n + y_n) = \lim x_n + \lim y_n\);
(ii) \(\lim (x_n y_n) = (\lim x_n)(\lim y_n)\);
(iii) \(\lim (cx_n) = c \lim x_n\) if \(c \in \mathbb{R}\);
(iv) \(\lim \frac{x_n}{y_n} = \frac{\lim x_n}{\lim y_n}\) if \(y_n \neq 0\) for all \(n\) and \(y \neq 0\).

Proof. Let \(x = \lim x_n\) and \(y = \lim y_n\).

(i) Given \(\epsilon > 0\), choose \(k \in \mathbb{N}\) such that if \(n \geq k\) then \(|x_n - x|, |y_n - y| < \frac{\epsilon}{2}\).

Then \(n \geq k\) implies
\[
|(x_n + y_n) - (x + y)| \leq |x_n - x| + |y_n - y| < \epsilon.
\]

(ii) We have
\[
x_n y_n - xy = x_n y_n - x y_n + x y_n - xy = (x_n - x) y_n + x (y_n - y) \to 0,
\]

since \(x_n - x\) and \(y_n - y\) both go to 0, and \((y_n)\) and the constant sequence \((x)\) are both bounded.

(iii) This is immediate from (ii) and convergence of a constant sequence to its constant value.

(iv) By (ii), it suffices to show \(\frac{1}{y_n} \to \frac{1}{y}\). Given \(\epsilon > 0\), choose \(l \in \mathbb{N}\) such that if \(n \geq l\) then \(|y_n - y| < |y|/2\). Then \(n \geq l\) implies \(|y_n| \geq |y|/2\). Now choose \(k \geq l\) such that if \(n \geq k\) then \(|y_n - y| < |y|^2/2\). Then \(n \geq k\) implies
\[
\left| \frac{1}{y_n} - \frac{1}{y} \right| \leq \frac{|y_n - y|}{|y_n y|} \leq \frac{2|y_n - y|}{|y|^2} < \epsilon.
\]

\(\square\)
**Remark 4.35.** Note that a couple of times in the above proof we used the shortcut of choosing something to perform several consistent jobs at once.

**Proposition 4.36.** For all convergent sequences \((x_n)\) and \((y_n)\), if \(x_n \leq y_n\) for all \(n \in \mathbb{N}\), then \(\lim x_n \leq \lim y_n\).

**Proof.** Put \(z_n = y_n - x_n\) and \(z = \lim z_n\). Then \(z_n \geq 0\) for all \(n\), and by the above proposition it suffices to show \(z \geq 0\). Suppose not, that is, assume \(z < 0\). Choose \(n \in \mathbb{N}\) such that
\[
|z_n - z| < |z|.
\]
Then
\[
z_n < z + |z| = 0,
\]
a contradiction. \(\Box\)

**Definition 4.37.**
(i) The set of extended real numbers is defined as
\[
\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm \infty\},
\]
where \(+\infty\) and \(-\infty\) are distinct objects which are not real numbers.
(ii) We write \(\infty\) for \(+\infty\).
(iii) The ordering of \(\mathbb{R}\) is extended to \(\overline{\mathbb{R}}\) by
\[
-\infty < x < \infty \quad \text{for } x \in \mathbb{R}.
\]
(iv) Terms such as maximum or minimum, upper or lower bound, and supremum or infimum are applied to sets of extended real numbers in the obvious way.

**Proposition 4.38.** Let \(A \subseteq \overline{\mathbb{R}}\). Then:
(i) \(A\) is bounded above, and if \(A \neq \emptyset\) then \(A\) has a supremum.
(ii) If \(A \subseteq \mathbb{R}\) and \(A \neq \emptyset\), then \(A\) is bounded above in \(\mathbb{R}\) if and only if \(\sup A < \infty\).
(iii) Similarly for bounded below and infimum.

**Proof.** (i) Since \(\overline{\mathbb{R}}\) has a maximum (namely, \(\infty\)), it is bounded above, hence so is every subset, in particular \(A\). For the other part, assume \(A \neq \emptyset\).
   Case 1. \(\infty \in A\). Then \(\infty = \max A\), so \(\infty = \sup A\).
   Case 2. \(A = \{-\infty\}\). Then \(-\infty = \sup A\).
   Case 3. \(\infty \notin A\) and \(B := A \cap \mathbb{R} \neq \emptyset\). If \(B\) is bounded above in \(\mathbb{R}\), then \(s := \sup B \in \mathbb{R}\) by the completeness axiom, and we have \(s = \sup A\) since \(-\infty \leq s\). On the other hand, if \(B\) is unbounded above in \(\mathbb{R}\), then for all \(M \in \mathbb{R}\) there exists \(x \in B\) such that \(x > M\), so \(\infty\) is the only upper bound of \(B\) in \(\mathbb{R}\), hence the only upper bound of \(A\), thus \(\infty = \sup A\).

(ii) First, if \(A\) is bounded above in \(\mathbb{R}\), then \(\sup A \in \mathbb{R}\), so \(\sup A < \infty\).
   Conversely, the proof of (i) shows that if \(A\) is unbounded above in \(\mathbb{R}\) then \(\sup A = \infty\).
   (iii) Either use an argument similar to the above, or multiply by \(-1\) and use the above results. \(\Box\)

**Definition 4.39** (Infinite Limits). Let \((x_n)\) be a sequence.
(i) We say \((x_n)\) diverges to \(\infty\) if for all \(M \in \mathbb{R}\) there exists \(k \in \mathbb{N}\) such that for all \(n \in \mathbb{N}\),

\[
\text{if } n \geq k \text{ then } x_n > M.
\]

(ii) Our notation for infinite limits is similar to that for finite limits; for example, we write \(x_n \to \infty\) or \(\lim x_n = \infty\).

(iii) Similarly for \(-\infty\) (reverse the inequalities).

(iv) We write \(x_n \to x\) in \(\mathbb{R}\) if either \(x \in \mathbb{R}\) and \((x_n)\) converges to \(x\) or \(x = \pm \infty\) and \((x_n)\) diverges to \(x\).

Example 4.40. \(n \to \infty\).

**Proposition 4.41.** For any sequence \((x_n)\), we have \(x_n \to \infty\) if and only if there exists \(k \in \mathbb{N}\) such that both

(i) for all \(n \in \mathbb{N}\), if \(n \geq k\) then \(x_n > 0\), and

(ii) the sequence \(\left(\frac{1}{x_n}\right)_{n \geq k}\) converges to 0.

Similarly for \(-\infty\).

**Proof.** We only give the argument for the case \(\infty\); the case \(-\infty\) is similar. Since the important properties of the sequence \((x_n)\) do not change if we delete finitely many terms, without loss of generality \(x_n > 0\) for all \(n \in \mathbb{N}\). First assume \(x_n \to \infty\). Given \(\epsilon > 0\), choose \(l \in \mathbb{N}\) such that \(n \geq l\) implies \(x_n > \frac{1}{\epsilon}\). Then \(n \geq l\) implies \(\frac{1}{x_n} < \epsilon\).

Conversely, assume \(\frac{1}{x_n} \to 0\). Given \(M \in \mathbb{R}\), without loss of generality \(M > 0\), so we can choose \(l \in \mathbb{N}\) such that \(n \geq l\) implies \(\frac{1}{x_n} < \frac{1}{M}\). Then \(n \geq l\) implies \(x_n > M\). \(\square\)

**Proposition 4.42.** Let \((x_n)\) and \((y_n)\) be sequences, and assume \(x_n \to \infty\). Then:

(i) \((y_n)\) bounded below implies \(x_n + y_n \to \infty\);

(ii) \(\inf y_n > 0\) implies \(x_ny_n \to \infty\);

(iii) \((y_n)\) bounded and \(x_n \neq 0\) for all \(n\) implies \(\frac{y_n}{x_n} \to 0\).

Similarly for \(x_n \to -\infty\), etc.

**Proof.** (i) Choose \(t \in \mathbb{R}\) such that \(y_n \geq t\) for all \(n \in \mathbb{N}\). Given \(M \in \mathbb{R}\), choose \(k \in \mathbb{N}\) such that \(n \geq k\) implies \(x_n > M - t\). Then \(n \geq k\) implies \(x_n + y_n \geq x_n + t > M\).

(ii) Choose \(t > 0\) such that \(y_n \geq t\) for all \(n \in \mathbb{N}\). Given \(M \in \mathbb{R}\), choose \(k \in \mathbb{N}\) such that \(n \geq k\) implies \(x_n > M/t\). Then \(n \geq k\) implies \(x_ny_n \geq x_nt > M\).

(iii) By the preceding proposition we have \(1/x_n \to 0\). Then \(y_n/x_n = y_n(1/x_n) \to 0\) since \((y_n)\) is bounded. \(\square\)

**Remark 4.43.** Many results about convergent sequences, for example the Squeeze Theorem, have obvious analogues for infinite limits.

**Corollary 4.44.** For all \(x \in \mathbb{R}\),

(i) if \(x > 1\) then \(x^n \to \infty\), and
(ii) if $|x| < 1$ then $x^n \to 0$.

Proof. (i) We have $x = 1 + a$ with $a > 0$. Then by Bernoulli’s Inequality,

$$x^n = (1 + a)^n \geq 1 + na,$$

Since $1 + na \to \infty$ by the preceding proposition, we have $x^n \to \infty$ by the Squeeze Theorem for infinite limits.

(ii) Case 1. $x = 0$. Then $x^n \to 0$ trivially.

Case 2. $x \neq 0$. Then $\frac{1}{|x|} > 1$, so

$$\left(\frac{1}{|x|}\right)^n \to \infty$$

by the first part of the corollary. Hence $|x|^n \to 0$ by the rules of infinite limits. Thus $x^n \to 0$, since $|x^n| = |x|^n$. \hfill \qed

Lemma 4.45. If $A$ is a nonempty subset of $\mathbb{R}$, then there exists a sequence $(x_n)$ in $A$ such that $x_n \to \sup A$. Similarly for inf.

Proof. Put $s := \sup A$. Case 1. $s = \infty$. For each $n \in \mathbb{N}$ choose $x_n \in A$ such that $x_n > n$. Then $x_n \to \infty$ by the Squeeze Theorem.

Case 2. $s \in \mathbb{R}$. For each $n \in \mathbb{N}$ choose $x_n \in A$ such that $x_n > s - 1/n$. Then

$$s - \frac{1}{n} < x_n \leq s < s + \frac{1}{n},$$

so $|x_n - s| < 1/n$. Hence $x_n \to s$ by the Squeeze Theorem. \hfill \qed

Definition 4.46. Let $(x_n)$ be a sequence.

(i) The limit supremum of $(x_n)$ is defined as

$$\limsup_{n \to \infty} x_n := \inf \sup_{k \in \mathbb{N}, n \geq k} x_n.$$

(ii) Similarly for limit infimum.

Remark 4.47. In more detail: for each $k \in \mathbb{N}$ let $s_k$ be the supremum of the tail end $\{x_n : n \geq k\}$. Then $\{s_k : k \in \mathbb{N}\}$ is a nonempty subset of $\mathbb{R}$. As such, this set has an infimum, which is defined to be the lim sup of $(x_n)$.

Lemma 4.48. For any sequence $(x_n)$, we have $\liminf x_n = -\limsup(-x_n)$.

Lemma 4.49. For any sequence $(x_n)$, we have $\limsup x_n < \infty$ if and only if $(x_n)$ is bounded above. Similarly for lim inf and bounded below.

Proof. $\limsup x_n < \infty$ if and only if there exists $k \in \mathbb{N}$ such that $\sup_{n \geq k} x_n < \infty$, if and only if $\sup x_n < \infty$. \hfill \qed

Lemma 4.50. Let $(x_n)$ be a sequence, and let $x = \limsup x_n$. Then:

(i) for all $a < x$ and for all $k \in \mathbb{N}$ there exists $n \geq k$ such that $x_n > a$, and

(ii) for all $a > x$ there exists $k \in \mathbb{N}$ such that for all $n \geq k$ we have $x_n < a$.

Similarly for lim inf.
Proof. (i) Let \( k \in \mathbb{N} \). Since \( a < x \leq \sup_{n \geq k} x_n \), by Approximation of Suprema there exists \( n \geq k \) such that \( x_n > a \).

(ii) Since \( a > \inf_{k \in \mathbb{N}} \sup_{n \geq k} x_n \), by Approximation of Infima there exists \( k \in \mathbb{N} \) such that \( \sup_{n \geq k} x_n < a \), so that for all \( n \geq k \) we have \( x_n < a \). \( \square \)

**Theorem 4.51.** For any sequence \( (x_n) \), \( \lim \sup x_n \) is the maximum extended real number \( x \) such that \( (x_n) \) has a subsequence \((y_k)\) with \( y_k \to x \). Similarly for \( \lim \inf \) and minimum.

*Proof.* Put \( x = \lim \sup x_n \).

Case 1. \( x = \infty \). Then for all \( M \in \mathbb{R} \) and for all \( k \in \mathbb{N} \) there exists \( n \geq k \) such that \( x_n > M \). Letting \( M = k = 1 \), we can choose \( n_1 \in \mathbb{N} \) such that \( x_{n_1} > 1 \). Then for all \( k > 1 \) we can inductively choose \( n_k > n_{k-1} \) such that \( x_{n_k} > k \). Then \( x_{n_k} \to \infty \) by the Squeeze Theorem.

Case 2. \( x = -\infty \). Given \( M \in \mathbb{R} \), we can choose \( k \in \mathbb{N} \) such that for all \( n \geq k \) we have \( x_n < M \). Thus \( x_n \to -\infty \).

Case 3. \( x \in \mathbb{R} \). We will show that there exists a subsequence \( (x_{n_k}) \) such that for all \( k \in \mathbb{N} \) we have

\[
x - \frac{1}{k} < x_{n_k} < x + \frac{1}{k},
\]

which will imply that \( x_{n_k} \to x \) by the Squeeze Theorem. First choose \( j_1 \in \mathbb{N} \) such that for all \( n \geq j_1 \) we have \( x_n < x + 1 \), and then choose \( n_1 \geq j_1 \) such that \( x_{n_1} > x - 1 \). For \( k > 1 \), first choose \( j_k \in \mathbb{N} \) such that for all \( n \geq j_k \) we have \( x_n < x + 1/k \), and then inductively choose \( n_k > \max\{n_{k-1}, j_k\} \) such that \( x_{n_k} > x - 1/k \).

Now let \( y > x \), and choose \( z \in (x, y) \). Then there exists \( k \in \mathbb{N} \) such that for all \( n \geq k \) we have \( x_n < z \). Thus no subsequence of \( (x_n) \) can converge (or diverge) to \( y \).

This proves the statements concerning the \( \lim \sup \), and the arguments for the \( \lim \inf \) are similar. \( \square \)

**Corollary 4.52.** Let \( (x_n) \) be a sequence and \( x \in \overline{\mathbb{R}} \). Then \( x_n \to x \) in \( \overline{\mathbb{R}} \) if and only if \( \lim \inf x_n = \lim \sup x_n = x \).

*Proof.* First assume \( x_n \to x \) in \( \overline{\mathbb{R}} \). Then every subsequence converges (or diverges) to \( x \), so \( \lim \inf x_n = \lim \sup x_n = x \).

Conversely, assume \( \lim \inf x_n = \lim \sup x_n = x \). If \( x = \pm \infty \), an argument from the preceding proof shows \( x_n \to x \). On the other hand, if \( x \in \mathbb{R} \), given \( \varepsilon > 0 \) choose \( k \in \mathbb{N} \) such that both \( x - \varepsilon < \inf_{n \geq k} x_n \) and \( \sup_{n \geq k} x_n < x + \varepsilon \). Then \( n \geq k \) implies \( x - \varepsilon < x_n < x + \varepsilon \), hence \( |x_n - x| < \varepsilon \). Thus \( x_n \to x \). \( \square \)

**Theorem 4.53** (Bolzano-Weierstrass Theorem). Every bounded sequence in \( \mathbb{R} \) has a convergent subsequence.

*Proof.* Let \( (x_n) \) be a bounded sequence in \( \mathbb{R} \). Then \( \lim \sup x_n \) is a real number. As we have seen, \( (x_n) \) has a subsequence converging to \( \lim \sup x_n \). \( \square \)
Corollary 4.54. Every bounded infinite subset of \( \mathbb{R} \) has a cluster point.

Proof. Let \( A \) be a bounded infinite subset of \( \mathbb{R} \). Then there is a 1-1 sequence \((x_n)\) in \( A \). Since \( A \) is bounded, so is the sequence \((x_n)\). By the Bolzano-Weierstrass Theorem, \((x_n)\) has a convergent subsequence, say with limit \( x \). Then either \( x \) is a cluster point of the range of \((x_n)\), hence of the superset \( A \), or \( x_n = x \) for infinitely many \( n \). Since \((x_n)\) is 1-1, we could only have \( x_n = x \) for at most 1 value of \( n \). Therefore, \( x \) must be a cluster point of \( A \).

Remark 4.55. It turns out that the above corollary is actually equivalent to the Bolzano-Weierstrass Theorem, and in fact the distinction between them is sometimes blurred.

Definition 4.56. A sequence \((x_n)\) is called Cauchy if for all \( \epsilon > 0 \) there exists \( k \in \mathbb{N} \) such that for all \( n, j \in \mathbb{N} \),

\[
\text{if } n, j \geq k \text{ then } |x_n - x_j| < \epsilon.
\]

Theorem 4.57. A sequence converges if and only if it is Cauchy.

Proof. Let \((x_n)\) be a sequence. First, assume \((x_n)\) converges, and let \( x = \lim x_n \). Given \( \epsilon > 0 \), choose \( k \in \mathbb{N} \) such that \( n \geq k \) implies \( |x_n - x| < \epsilon/2 \). Then \( n, j \geq k \) imply

\[
|x_n - x_j| \leq |x_n - x| + |x - x_j| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,
\]

so \((x_n)\) is Cauchy.

Conversely, assume \((x_n)\) is Cauchy. Claim: \((x_n)\) is bounded. For, we can choose \( k \in \mathbb{N} \) such that \( n, j \geq k \) imply \( |x_n - x_j| < 1 \). Put \( M = \max\{1, |x_1 - x_k|, \ldots, |x_{k-1} - x_k|\} \) (where without loss of generality \( k > 1 \)). Then \( |x_n - x_k| \leq M \) for all \( n \in \mathbb{N} \), proving the claim. So, by the Bolzano-Weierstrass Theorem \((x_n)\) has a convergent subsequence, say \( x_{n_i} \rightarrow x \). Given \( \epsilon > 0 \), choose \( k \in \mathbb{N} \) such that \( n, j \geq k \) imply \( |x_n - x_j| < \epsilon/2 \). Also choose \( l \in \mathbb{N} \) such that \( i \geq l \) implies \( |x_{n_i} - x| < \epsilon/2 \). Let \( n \geq k \), and put \( i = \max\{l, k\} \). Then \( n_i \geq i \geq k \), so

\[
|x_n - x| \leq |x_n - x_{n_i}| + |x_{n_i} - x| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\]

Thus \((x_n)\) itself converges.

Remark 4.58. Note that the above proof that every Cauchy sequence is bounded is (unsurprisingly) similar to the proof that every convergent sequence is bounded.

Proposition 4.59. For all sequences \((x_n)\) and \((y_n)\), if \((y_n)\) converges, then:

(i) \( \limsup(x_n + y_n) = \limsup x_n + \lim y_n \) (where \( \pm \infty + t \) is interpreted as \( \pm \infty \) for all \( t \in \mathbb{R} \));

(ii) \( \limsup(x_n y_n) = (\limsup x_n)(\lim y_n) \) if \( \lim y_n > 0 \) (where \( (\pm \infty)t \) is interpreted as \( \pm \infty \) for all \( t > 0 \)).
Proposition 4.60. For all sequences \((x_n)\) and \((y_n)\), if \(x_n \leq y_n\) for all \(n \in \mathbb{N}\) then \(\limsup x_n \leq \limsup y_n\) and \(\liminf x_n \leq \liminf y_n\).

Proof. For all \(k \in \mathbb{N}\), \(\sup_{n \geq k} x_n \leq \sup_{n \geq k} y_n\), so
\[
\limsup x_n = \inf \sup_{n \geq k} x_n \leq \inf \sup_{n \geq k} y_n = \limsup y_n.
\]
Similarly for \(\liminf\).

\[\square\]

5. Limits

Standing Hypothesis 5.1. From now on, all our functions \(f\) will satisfy \(\text{dom } f \subseteq \mathbb{R}\) unless otherwise specified.

Definition 5.2. Let \(f : A \to \mathbb{R}\), \(t \in A'\), and \(L \in \mathbb{R}\). We say \(f(x)\) goes to \(L\) as \(x\) goes to \(t\), written \(f(x) \to L\) as \(x \to t\), or \(f(x) \xrightarrow{x \to t} L\), if for all \(\epsilon > 0\) there exists \(\delta > 0\) such that for all \(x \in A\),
\[
0 < |x - t| < \delta \quad \text{then} \quad |f(x) - L| < \epsilon.
\]

Remark 5.3. With the above notation, a routine Triangle Inequality argument shows that \(L\) is uniquely determined by \(f\) and \(t\), so it makes sense to formalize this relationship:

Definition 5.4. Let \(f : A \to \mathbb{R}\), \(t \in A'\), and \(L \in \mathbb{R}\). If \(f(x) \to L\) as \(x \to t\), we also say \(L\) is the limit of \(f\) at \(t\), written \(L = \lim_{x \to t} f(x)\).

Example 5.5. (i) \(\lim_{x \to t} x = t\).

(ii) \(\lim_{t \to t} L = L\).

Proposition 5.6 (Sequential Characterization of Limits). Let \(f : A \to \mathbb{R}\) and \(t \in A'\). Then \(\lim_{x \to t} f(x)\) exists if and only if for every sequence \((x_n)\) in \(A \setminus \{t\}\) converging to \(t\), the sequence \((f(x_n))\) converges (in which case \(\lim_{n \to \infty} f(x_n) = \lim_{x \to t} f(x)\)).

Proof. First assume \(\lim_{x \to t} f(x) = L\), and let \((x_n)\) be a sequence in \(A \setminus \{t\}\) converging to \(t\). Given \(\epsilon > 0\), choose \(\delta > 0\) such that \(0 < |x - t| < \delta\) implies \(|f(x) - L| < \epsilon\). Now choose \(k \in \mathbb{N}\) such that \(n \geq k\) implies \(|x_n - t| < \delta\). Then \(n \geq k\) implies \(|f(x_n) - L| < \epsilon\). Thus \(f(x_n) \to L\).

Conversely, assume the condition regarding sequences. First note that if \((x_n)\) and \((x'_n)\) are two sequences in \(A \setminus \{t\}\) converging to \(t\), then
\[
\lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} f(x'_n).
\]
To see this, define
\[
z_k := \begin{cases} x_n & \text{if } k = 2n - 1 \\ x'_n & \text{if } k = 2n. \end{cases}
\]
Then \(z_k \to t\), so \(\lim_{k \to \infty} f(z_k)\) exists by hypothesis. Since \((f(x_n))\) and \((f(x'_n))\) are both subsequences of \((f(z_k))\), we must have \(\lim f(x_n) = \lim f(x'_n)\).

Consequently, we can define \(L\) to be the common limit of all the sequences \((f(x_n))\) for \((x_n)\) in \(A \setminus \{t\}\) converging to \(t\). We will show \(\lim_{x \to t} f(x) = L\).
MAT 371 LECTURES 35

Suppose not. Then there exists \( \epsilon > 0 \) such that for all \( \delta > 0 \) there exists \( x \in A \) such that

\[
0 < |x - t| < \delta \quad \text{and} \quad |f(x) - L| \geq \epsilon.
\]

In particular, for all \( n \in \mathbb{N} \) there exists \( x_n \in A \) such that

\[
0 < |x_n - t| < \frac{1}{n} \quad \text{and} \quad |f(x_n) - L| \geq \epsilon.
\]

But then \( x_n \to t \) and \( f(x_n) \not\to L \), a contradiction. \( \square \)

**Definition 5.7.** Given \( f, g : A \to \mathbb{R} \), we say \( f \leq g \) if \( f(x) \leq g(x) \) for all \( x \in A \), and similarly for \( \geq, < \), and \( > \).

**Proposition 5.8** (Squeeze Theorem for Limits). Let \( f, g, h : A \to \mathbb{R} \) and \( t \in A' \). If \( f \leq g \leq h \) and

\[
\lim_{x \to t} f(x) = \lim_{x \to t} h(x) = L,
\]

then \( \lim_{x \to t} g(x) = L \).

**Proof.** Immediate from the Sequential Characterization of Limits and the Squeeze Theorem for Sequences. More precisely, suppose \( (x_n) \) is any sequence in \( A \setminus \{t\} \) converging to \( t \). Then the sequences \( (f(x_n)), (g(x_n)), \) and \( (h(x_n)) \) satisfy the hypothesis of the Squeeze Theorem for Sequences, and the Sequential Characterization of Limits tells us \( \lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} h(x_n) = L \), so we have \( g(x_n) \to L \). Again by the Sequential Characterization of Limits, we conclude that \( \lim_{x \to t} g(x) = L \). \( \square \)

**Remark 5.9.** In the above proof, the application of the Sequential Characterization of Limits and the corresponding result for sequences is so routine that in similar situations we will omit the details.

**Proposition 5.10.** Let \( f, g : A \to \mathbb{R} \) and \( t \in A' \). If \( \lim_{x \to t} f(x) = 0 \) and \( g \) is bounded, then \( \lim_{x \to t} f(x)g(x) = 0 \).

**Proof.** Immediate from the Sequential Characterization of Limits and the corresponding result for sequences. \( \square \)

**Proposition 5.11.** Let \( f, g : A \to \mathbb{R} \) and \( t \in A' \). If \( f \) and \( g \) both have limits at \( t \), then:

(i) \( \lim_{x \to t} (f(x) + g(x)) = \lim_{x \to t} f(x) + \lim_{x \to t} g(x) \);

(ii) \( \lim_{x \to t} cf(x) = c \lim_{x \to t} f(x) \) if \( c \in \mathbb{R} \);

(iii) \( \lim_{x \to t} f(x)g(x) = (\lim_{x \to t} f(x))(\lim_{x \to t} g(x)) \);

(iv) \( \lim_{x \to t} \frac{f(x)}{g(x)} = \frac{\lim_{x \to t} f(x)}{\lim_{x \to t} g(x)} \) if \( \lim_{x \to t} g(x) \neq 0 \) and \( 0 \not\in \text{ran } g \).

**Proof.** Immediate from the Sequential Characterization of Limits and the corresponding result for sequences. \( \square \)
**Proposition 5.12.** Let $A, B \subseteq \mathbb{R}$, $t \in A'$, $L \in B'$, $M \in \mathbb{R}$, $f: A \to B \setminus \{L\}$, and $g: B \to \mathbb{R}$. If $\lim_{x \to t} f(x) = L$ and $\lim_{y \to L} g(y) = M$, then

$$\lim_{x \to t} g \circ f(x) = M.$$

**Proof.** Let $(x_n)$ be a sequence in $A \setminus \{t\}$ converging to $t$. Then $(f(x_n))$ is a sequence in $B \setminus \{L\}$ converging to $L$, by the Sequential Characterization of Limits. Hence $g \circ f(x_n) = g(f(x_n)) \to M$, again by the Sequential Characterization of Limits. \hfill \Box

**Proposition 5.13.** Let $f, g: A \to \mathbb{R}$ and $t \in A'$, and suppose $f$ and $g$ both have limits at $t$. If $f \leq g$, then $\lim_{x \to t} f(x) \leq \lim_{x \to t} g(x)$.

**Proof.** Immediate from the Sequential Characterization of Limits and the corresponding result for sequences. \hfill \Box

**Definition 5.14 (Restricted Limits).** Let $f: A \to \mathbb{R}$, $B \subseteq A$, and $t \in B'$. The limit of $f$ at $t$ through $B$ is defined as

$$\lim_{x \to t} f(x) := \lim_{x \in B} f(x).$$

**Definition 5.15 (One-Sided Limits).** Let $f: A \to \mathbb{R}$ and $t \in \mathbb{R}$.

(i) Put $B := A \cap (t, \infty)$. If $t \in B'$ then the right-hand limit of $f$ at $t$ is defined as

$$f(t+) := \lim_{x \to t^+} f(x).$$

(ii) We also write

$$\lim_{x \uparrow t} f(x) = \lim_{x \to t^+} f(x) = f(t+).$$

(iii) Similarly for left-hand limit.

**Lemma 5.16.** Let $f: A \to \mathbb{R}$ and $t \in \mathbb{R}$, and suppose

$$t \in (A \cap (t, \infty))' \cap (A \cap (-\infty, t))'.$$  

Then $f$ has a limit at $t$ if and only if both one-sided limits exist and are equal, in which case

$$\lim_{x \to t} f(x) = f(t-) = f(t+).$$

**Proof.** One direction is trivial, so assume $f(t-) = f(t+) = L$. Given $\epsilon > 0$, choose $\delta > 0$ such that for all $x \in A$, $t - \delta < x < t$ implies $|f(x) - L| < \epsilon$ and $t < x < t + \delta$ implies $|f(x) - L| < \epsilon$. Then $0 < |x - t| < \delta$ implies $|f(x) - L| < \epsilon$. \hfill \Box

**Definition 5.17 (Monotone Functions).** A function $f: A \to \mathbb{R}$ is called:

(i) increasing if for all $x, y \in A$ we have $x < y$ implies $f(x) \leq f(y)$;
(ii) decreasing if for all $x, y \in A$ we have $x < y$ implies $f(x) \geq f(y)$;
(iii) monotone if it is increasing or decreasing;
(iv) strictly increasing if for all \( x, y \in A \) we have \( x < y \) implies \( f(x) < f(y) \);
(v) strictly decreasing if for all \( x, y \in A \) we have \( x < y \) implies \( f(x) > f(y) \);
(vi) strictly monotone if it is strictly increasing or strictly decreasing.

**Proposition 5.18.** Let \( f : [a, b] \to \mathbb{R} \) be increasing. Then:

(i) \( a < t \leq b \) implies \( f(t-) \) exists and equals \( \sup_{x < t} f(x) \), and moreover \( f(t-) \leq f(t) \);
(ii) \( a \leq t < b \) implies \( f(t+) \) exists and equals \( \inf_{x > t} f(x) \), and moreover \( f(t+) \geq f(t) \);
(iii) \( a \leq s < t \leq b \) implies \( f(s+) \leq f(t-) \).

Similarly if \( f \) is decreasing.

**Proof.** (i) Put \( L = \sup_{x < t} f(x) \). Then \( L \leq f(t) \). Given \( \epsilon > 0 \), choose \( c < t \) such that \( f(c) > L - \epsilon \). Then

\[
c < x < t \implies L - \epsilon < f(c) \leq f(x) \leq L \implies |f(x) - L| < \epsilon.
\]

Thus \( f(t-) = L \).

(ii) Similar to (i).

(iii) Choose \( c \in (s, t) \). Then

\[
f(s+) = \inf_{x > s} f(x) \leq f(c) \leq \sup_{x < t} f(x) = f(t-).
\]

\[\square\]

**Definition 5.19 (Limits at Infinity).** (i) Let \( f : A \to \mathbb{R} \) and \( L \in \mathbb{R} \), and suppose \( A \) is unbounded above. We say \( f(x) \) goes to \( L \) as \( x \) goes to \( \infty \), written \( f(x) \to L \) as \( x \to \infty \), if for all \( \epsilon > 0 \) there exists \( M \in \mathbb{R} \) such that for all \( x \in A \),

\[
\text{if } x > M \text{ then } |f(x) - L| < \epsilon.
\]

(ii) With the notation of (i), \( L \) is uniquely determined by \( f \), and is called the limit of \( f \) at \( \infty \), written \( \lim_{x \to \infty} f(x) \).

(iii) Similarly for \( \lim_{x \to -\infty} f(x) \).

**Definition 5.20 (Infinite Limits).** (i) Let \( f : A \to \mathbb{R} \) and \( t \in A' \). We say \( f(x) \) goes to \( \infty \) as \( x \to t \), written \( f(x) \to \infty \) as \( x \to t \), or \( \lim_{x \to t} f(x) = \infty \), if for all \( M \in \mathbb{R} \) there exists \( \delta > 0 \) such that for all \( x \in A \),

\[
\text{if } 0 < |x - t| < \delta \text{ then } f(x) > M.
\]

(ii) Similarly for \( f(x) \to -\infty \), and also for \( f(x) \to \pm \infty \) as \( x \to \pm \infty \).

**Remark 5.21.** (i) The Sequential Characterization of Limits extends appropriately to limits at infinity and to infinite limits. For example, if \( \text{dom } f \) is not bounded above, then \( \lim_{x \to \infty} f(x) = L \) if and only if for every sequence \( (x_n) \) in \( \text{dom } f \), \( x_n \to \infty \) implies \( f(x_n) \to L \).

(ii) The basic properties of limits extend to limits at infinity.
(iii) The basic properties of infinite limits of sequences extend to infinite limits of functions.

6. Continuity

**Standing Hypothesis 6.1.** Recall that we are assuming that all our functions \( f \) satisfy \( \text{dom } f, \text{ran } f \subseteq \mathbb{R} \) unless otherwise specified.

**Definition 6.2.** Let \( f : A \to \mathbb{R} \) and \( t \in A \).

(i) \( f \) is called **continuous at** \( t \) if for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that for all \( x \in A \),

\[
| x - t | < \delta \quad \text{then} \quad | f(x) - f(t) | < \epsilon.
\]

(ii) Otherwise we say \( f \) is **discontinuous at** \( t \), or has a **discontinuity** at \( t \).

(iii) \( f \) is called **continuous** if it is continuous at each element of \( A \). More generally, if \( B \subseteq A \), we say \( f \) is **continuous on** \( B \) if \( f \) is continuous at each element of \( B \).

**Theorem 6.3** (Sequential Characterization of Continuity). Let \( f : A \to \mathbb{R} \) and \( t \in A \). Then \( f \) is continuous at \( t \) if and only if for every sequence \( (x_n) \) in \( A \), if \( x_n \to t \) then \( f(x_n) \to f(t) \).

**Proof.** First assume \( f \) is continuous at \( t \), and let \( (x_n) \) be a sequence in \( A \) converging to \( t \). Given \( \epsilon > 0 \), choose \( \delta > 0 \) such that for all \( x \in A \), if \( | x - t | < \delta \) then \( | f(x) - f(t) | < \epsilon \). Now choose \( k \in \mathbb{N} \) such that \( n \geq k \) implies \( | x_n - t | < \delta \). Then \( n \geq k \) implies \( | f(x_n) - f(t) | < \epsilon \). Thus \( f(x_n) \to f(t) \).

Conversely, assume \( x_n \to t \) implies \( f(x_n) \to f(t) \). Suppose \( f \) is not continuous at \( t \). Choose \( \epsilon > 0 \) such that for all \( \delta > 0 \) there exists \( x \in A \) such that \( | x - t | < \delta \) and \( | f(x) - f(t) | \geq \epsilon \). In particular, for all \( n \in \mathbb{N} \) we can choose \( x_n \in A \) such that

\[
| x_n - t | < \frac{1}{n} \quad \text{and} \quad | f(x_n) - f(t) | \geq \epsilon.
\]

Then \( x_n \to t \) by the Squeeze Theorem, but \( f(x_n) \not\to f(t) \), giving a contradiction.

\( \square \)

**Proposition 6.4.** Let \( f, g : A \to \mathbb{R} \) and \( t \in A \). If \( f \) and \( g \) are both continuous at \( t \), then so are:

(i) \( f + g \);

(ii) \( cf \) if \( c \in \mathbb{R} \);

(iii) \( fg \);

(iv) \( \frac{f}{g} \) if \( 0 \not\in \text{ran } g \).

**Proof.** If \( x_n \to t \), then \( f(x_n) \to f(t) \) and \( g(x_n) \to g(t) \), so:

(i) \( (f + g)(x_n) = f(x_n) + g(x_n) \to f(t) + g(t) = (f + g)(t) \);

(ii) \( (cf)(x_n) = cf(x_n) \to cf(t) = (cf)(t) \);

(iii) \( (fg)(x_n) = f(x_n)g(x_n) \to f(t)g(t) = (fg)(t) \);

(iv) \( \left( \frac{f}{g} \right)(x_n) = \frac{f(x_n)}{g(x_n)} \to \frac{f(t)}{g(t)} = \left( \frac{f}{g} \right)(t) \).

\( \square \)
The result now follows from the Sequential Characterization of Continuity. \hfill \Box

**Theorem 6.5** (Limit Characterization of Continuity). Let \( f : A \to \mathbb{R} \) and \( t \in A \).

(i) If \( t \not\in A' \), then \( f \) is continuous at \( t \).

(ii) If \( t \in A' \), then \( f \) is continuous at \( t \) if and only if

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

**Proof.** (i) Since \( t \not\in A' \), we can choose \( \delta > 0 \) such that for all \( x \in A \setminus \{t\} \) we have \( |x - t| \geq \delta \). Then for any \( \epsilon > 0 \) and \( x \in A \), if \( |x - t| < \delta \) then \( x = t \), so \( f(x) = f(t) \), hence \( |f(x) - f(t)| < \epsilon \).

(ii) First assume \( f \) is continuous at \( t \). Given \( \epsilon > 0 \), choose \( \delta > 0 \) such that for all \( x \in A \), \( 0 < |x - t| < \delta \) implies \( |f(x) - f(t)| < \epsilon \). Then trivially \( 0 < |x - t| < \delta \) implies \( |f(x) - f(t)| < \epsilon \). Thus \( \lim_{x \to t} f(x) = f(t) \).

Conversely, assume \( \lim_{x \to t} f(x) = f(t) \). Given \( \epsilon > 0 \), choose \( \delta > 0 \) such that for all \( x \in A \), \( 0 < |x - t| < \delta \) implies \( |f(x) - f(t)| < \epsilon \). Trivially, \( |f(t) - f(t)| < \epsilon \). Thus, for all \( x \in A \), \( |x - t| < \delta \) implies \( |f(x) - f(t)| < \epsilon \). Therefore \( f \) is continuous at \( t \). \hfill \Box

**Corollary 6.6** (Continuity Characterization of Limits). Let \( f : A \to \mathbb{R} \), \( t \in A' \), and \( L \in \mathbb{R} \). Then \( \lim_{x \to t} f(x) = L \) if and only if the function \( g : A \cup \{t\} \to \mathbb{R} \) defined by

\[
g(x) = \begin{cases} 
  f(x) & \text{if } x \neq t \\
  L & \text{if } x = t 
\end{cases}
\]

is continuous at \( t \).

**Proof.** Since \( g(x) = f(x) \) for all \( x \in A \setminus \{t\} \), we have \( \lim_{x \to t} g(x) = g(t) \) if and only if \( \lim_{x \to t} f(x) = L \). By the Limit Characterization of Continuity, \( g \) is continuous at \( t \) if and only if \( \lim_{x \to t} g(x) = g(t) \). The result follows. \hfill \Box

**Proposition 6.7.** Let \( A, B \subseteq \mathbb{R} \), \( f : A \to B \), and \( g : B \to \mathbb{R} \).

(i) If \( t \in A' \), \( \lim_{x \to t} f(x) = L \), and \( g \) is continuous at \( L \), then

\[
\lim_{x \to t} g \circ f(x) = g(L).
\]

(ii) If \( t \in A \), \( f \) is continuous at \( t \), and \( g \) is continuous at \( f(t) \), then \( g \circ f \) is continuous at \( t \).

**Proof.** (i) If \( x_n \in A \setminus \{t\} \) and \( x_n \to t \), then \( f(x_n) \to L \) by the Sequential Characterization of Limits, so \( g \circ f(x_n) \to g(L) \) by the Sequential Characterization of Continuity.

(ii) If \( x_n \to t \), then \( f(x_n) \to f(t) \) by continuity of \( f \), so \( g \circ f(x_n) \to g \circ f(t) \) by continuity of \( g \). \hfill \Box

**Definition 6.8.** Let \( f : A \to \mathbb{R} \) and \( x \in A \).
(i) We say $f$ has a maximum at $x$ if $f(x) = \max \text{ran } f$, in which case we write $\max f$ for this maximum value. Similarly, we write $\sup f$ for $\sup \text{ran } f$.

(ii) Similarly for minimum, $\min f$, and $\inf f$.

**Theorem 6.9** (Extreme Value Theorem). Every continuous function on a closed bounded interval has both a maximum and a minimum.

**Proof.** Let $f: [a, b] \to \mathbb{R}$ be continuous. We only give the argument for the maximum; the minimum is similar. Choose a sequence $(y_n)$ in $\text{ran } f$ such that $y_n \to \sup f$, and for each $n$ choose $x_n \in [a, b]$ such that $y_n = f(x_n)$. By the Bolzano-Weierstrass Theorem $(x_n)$ has a convergent subsequence $(x_{n_k})$. Put $x = \lim x_{n_k}$. Since $a \leq x_n \leq b$ for all $n$, $x \in [a, b]$. By continuity,

$$\sup f = \lim f(x_{n_k}) = f(x).$$

Since $f(x) \in \text{ran } f$, $f(x) = \max f$. \hfill \Box

**Theorem 6.10** (Intermediate Value Theorem). If $f: [a, b] \to \mathbb{R}$ is continuous, then for all $d$ between $f(a)$ and $f(b)$ there exists $c \in [a, b]$ such that $f(c) = d$.

**Proof.** Without loss of generality $f(a) \leq d \leq f(b)$. Put

$$A := \{x \in [a, b] : f(x) \leq d\}.$$  

Then $a \in A$, so $A \neq \emptyset$. Let $c = \sup A$. Choose a sequence $(x_n)$ in $A$ such that $x_n \to c$. Since $a \leq x_n \leq b$ for all $n$, $c \in [a, b]$. Then $f(x_n) \to f(c)$ by continuity. Since $f(x_n) \leq d$ for all $n$, $f(c) \leq d$. Suppose $f(c) < d$. Then $c < b$ since $f(b) \geq d$. Choose $\delta > 0$ such that for all $x \in [a, b]$, $|x - c| < \delta$ implies $|f(x) - f(c)| < d - f(c)$. Pick $x$ such that

$$c < x < \min\{c + \delta, b\}.$$  

Then $f(x) - f(c) < d - f(c)$, so $f(x) < d$. But then $x \in A$ and $x > c$, a contradiction. \hfill \Box

**Remark 6.11.** The Intermediate Value Theorem is equivalent to the following statement: if $f: A \to \mathbb{R}$ is continuous and $I \subseteq A$ is an interval, then $f(I)$ is also an interval.

**Definition 6.12.** $f: A \to \mathbb{R}$ is called uniformly continuous if for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in A$,

$$\text{if } |x - y| < \delta \text{ then } |f(x) - f(y)| < \epsilon.$$  

**Remark 6.13.** Note that $f$ is continuous if and only if for all $x \in A$ and $\epsilon > 0$ there exists $\delta > 0$ such that for all $y \in A$ we have $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$. The definition of uniform continuity moves the universally quantified $x$ across the existentially quantified $\delta$, producing a stronger condition in that now a single $\delta$ has to work for all $x$ simultaneously. Therefore, basic logic tells us that uniform continuity implies pointwise continuity.
Theorem 6.14. Every continuous function on a closed bounded interval is in fact uniformly continuous.

Proof. Let \( f : [a, b] \to \mathbb{R} \) be continuous. Suppose \( f \) is not uniformly continuous. Then there exists \( \epsilon > 0 \) such that for all \( \delta > 0 \) there exist \( x, y \in [a, b] \) such that \( |x - y| < \delta \) and \( |f(x) - f(y)| \geq \epsilon \). In particular, for all \( n \in \mathbb{N} \) there exist \( x_n, y_n \in [a, b] \) such that

\[
|x_n - y_n| < \frac{1}{n} \quad \text{and} \quad |f(x_n) - f(y_n)| \geq \epsilon.
\]

Choose a convergent subsequence \((x_{n_k})\) of \((x_n)\), and put \( x := \lim x_{n_k} \). Since \( |x_{n_k} - y_{n_k}| \to 0 \), we also have \( y_{n_k} \to x \). But then

\[
\epsilon \leq |f(x_{n_k}) - f(y_{n_k})| \to |f(x) - f(x)| = 0,
\]

a contradiction. \( \square \)

Theorem 6.15. If \( f : [a, b] \to \mathbb{R} \) is continuous and 1-1, then:

1. \( f \) is strictly monotone;
2. \( f([a, b]) \) is a closed bounded interval \([c, d]\);
3. \( f^{-1} : [c, d] \to [a, b] \) is strictly monotone (the same way \( f \) is);
4. \( f^{-1} \) is continuous.

Proof. Without loss of generality \( f(a) < f(b) \); for the other case just multiply by \(-1\).

(i) We will show \( f \) is strictly increasing on \([a, b]\). Let \( a \leq x < y \leq b \), and suppose \( f(a) > f(x) \). Then by the Intermediate Value Theorem there exists \( z \in [x, b] \) such that \( f(z) = f(a) \), a contradiction. Thus \( f(a) \leq f(x) \). Note that if \( a < x \) then this reasoning shows \( f(a) < f(x) \). Similarly, \( f(x) < f(b) \).

Applying this reasoning to \([x, b]\), we get \( f(x) < f(y) \).

(ii) By the Extreme Value Theorem, \( f \) has a maximum \( d \) and a minimum \( c \) on \([a, b]\). Then \( f([a, b]) \subseteq [c, d] \). By the Intermediate Value Theorem, \([c, d] \subseteq \text{ran } f \). Thus \( f([a, b]) = [c, d] \).

(iii) Let \( c \leq y < z \leq d \), and suppose \( f^{-1}(y) > f^{-1}(z) \). Since \( f \) is strictly increasing,

\[
y = f(f^{-1}(y)) > f(f^{-1}(z)) = z,
\]

a contradiction.

(iv) Let \( t \in [c, d] \) and \( \epsilon > 0 \). Without loss of generality it suffices to show that if \( t < d \) then there exists \( \delta > 0 \) such that \( t \leq y < t + \delta \) implies \( |f^{-1}(y) - f^{-1}(t)| < \epsilon \), or equivalently \( f^{-1}(y) < f^{-1}(t) + \epsilon \), since \( f^{-1} \) is increasing. Without loss of generality \( f^{-1}(t) + \epsilon \leq b \). Then

\[
t \leq y < f^{-1}(t) + \epsilon \quad \text{implies} \quad f^{-1}(t) \leq f^{-1}(y) < f^{-1}(t) + \epsilon,
\]

so we can take \( \delta = f(f^{-1}(t) + \epsilon) - t \). \( \square \)

Remark 6.16. The above result (with obvious modifications) remains true if the closed bounded interval \([a, b]\) is replaced by any type of interval. The arguments are routine, and we only indicate the idea for an open interval.
(a, b); so, assume \( f: (a, b) \to \mathbb{R} \) is continuous and 1-1. Note that \( a \) and \( b \) are allowed to be infinite here. The conclusion is that:

(i) \( f \) is strictly monotone;
(ii) \( f(a, b) \) is an open interval \((c, d)\);
(iii) \( f^{-1}: (c, d) \to (a, b) \) is strictly monotone (the same way \( f \) is);
(iv) \( f^{-1} \) is continuous.

The argument from the above theorem can be copied almost verbatim, except for (ii), where the relevant facts are: \( f(a, b) \) is an interval by the Intermediate Value Theorem, and \( f \) can have neither a maximum nor a minimum, since it is strictly monotone on an open interval.

Remark 6.17. The above results can be used to prove the existence of roots (see the discussion at the end of Section 2).

7. DIFFERENTIATION

Standing Hypothesis 7.1. Recall that we are assuming that all our functions \( f \) satisfy \( \text{dom } f, \text{ran } f \subseteq \mathbb{R} \) unless otherwise specified.

**Definition 7.2.** Let \( f \) be defined on some open interval containing \( t \).

(i) The **derivative of \( f \) at \( t \)** is defined as

\[
f'(t) := \lim_{x \to t} \frac{f(x) - f(t)}{x - t},
\]

provided this limit exists.

(ii) \( f \) is called **differentiable at \( t \)** if \( f'(t) \) exists.

(iii) \( f \) is called **differentiable** if it is differentiable at each element of its domain. More generally, if \( B \subseteq \text{dom } f \), we say \( f \) is differentiable on \( B \) if \( f \) is differentiable at every element of \( B \).

(iv) We also write

\[
\frac{d}{dx} f(x) = f'(x).
\]

**Example 7.3.**

(i) If \( c \in \mathbb{R} \) then \( \frac{d}{dx} c = 0 \).

(ii) \( \frac{d}{dx} x = 1 \).

**Proposition 7.4.** If \( f \) is differentiable at \( t \), then \( f \) is continuous at \( t \).

**Proof.** If \( x \in \text{dom } f \setminus \{t\} \), then

\[
f(x) - f(t) = \frac{f(x) - f(t)}{x - t} (x - t) \xrightarrow{x \to t} f'(t)0 = 0,
\]

so \( \lim_{x \to t} f(x) = f(t) \). Thus \( f \) is continuous at \( t \).

**Proposition 7.5.** If \( f \) and \( g \) are both differentiable at \( t \), then:

(i) \( (f + g)'(t) = f'(t) + g'(t) \);
(ii) (Product Rule) \( (fg)'(t) = f'(t)g(t) + f(t)g'(t) \);
(iii) (cf)'(t) = cf'(t) if \( c \in \mathbb{R} \);
(iv) (Quotient Rule) \( \left( \frac{f}{g} \right)'(t) = \frac{f'(t)g(t) - f(t)g'(t)}{g(t)^2} \) if \( 0 \not\in \text{ran } g \).
MAT 371 LECTURES

Proof. (i) We have
\[
\frac{(f + g)(x) - (f + g)(t)}{x - t} = \frac{f(x) - f(t)}{x - t} + \frac{g(x) - g(t)}{x - t}
\]
\[
x \to t \to f'(t) + g'(t).
\]
(ii) We have
\[
\frac{(fg)(x) - (fg)(t)}{x - t} = \frac{f(x)g(x) - f(t)g(t)}{x - t}
\]
\[
= \frac{f(x)g(x) - g(t)}{x - t} + \frac{f(t)g(t)}{x - t} - \frac{f(t)g(t)}{x - t}
\]
\[
x \to t \to f(t)g'(t) + f'(t)g(t),
\]
by continuity of \( f \) at \( t \).
(iii) Immediate from (ii), since the derivative of a constant is 0.
(iv) By (ii), it suffices to show \( \frac{1}{g'}(t) = \frac{-g''(t)}{g(t)^2} \):
\[
\frac{\frac{1}{g}(x) - \frac{1}{g}(t)}{x - t} = \frac{g(t) - g(x)}{(x - t)g(x)g(t)} \xrightarrow{x \to t} \frac{-g'(t)}{g(t)^2},
\]
by continuity of \( g \) at \( t \). \( \square \)

Corollary 7.6 (Power Rule). \( \frac{d}{dx}x^n = nx^{n-1} \) for all \( n \in \mathbb{Z} \).

Proof. Case 1. \( n = 0 \). We have
\[
\frac{d}{dx}x^0 = \frac{d}{dx}1 = 0 = 0x^{-1}.
\]
Case 2. \( n = 1 \). We have
\[
\frac{d}{dx}x^1 = \frac{d}{dx}x = 1 = 1x^0.
\]
Case 3. \( n > 1 \). Inductively, if we assume \( \frac{d}{dx}x^{n-1} = (n - 1)x^{n-2} \), then
\[
\frac{d}{dx}x^n = \frac{d}{dx}(x^{n-1}x) = (n - 1)x^{n-2}x + x^{n-1}1 = nx^{n-1}.
\]
Case 3. \( n < 0 \). Then \( n = -k \) with \( k > 0 \), so
\[
\frac{d}{dx}x^n = \frac{d}{dx} \frac{1}{x^k} = \frac{-kx^{k-1}}{x^{2k}} = -kx^{-k-1} = nx^{n-1}.
\]
\( \square \)

Lemma 7.7. \( f \) is differentiable at \( t \) if and only if there exists a function \( q \) which is continuous at \( t \) and satisfies
\[
(7.1) \quad f(x) = f(t) + q(x)(x - t) \quad \text{for all } x \in \text{dom } f,
\]
in which case \( f'(t) = q(t) \).
Proof. Note that for \( x \in \text{dom} \ f \setminus \{t\} \), Equation (7.1) is equivalent to

\[
q(x) = \frac{f(x) - f(t)}{x - t}.
\]

Thus, by the Continuity Characterization of Limits, if \( L \in \mathbb{R} \) then \( f'(t) \) exists and equals \( L \) if and only if the function \( q : \text{dom} \ f \rightarrow \mathbb{R} \) defined by

\[
q(x) = \begin{cases} 
\frac{f(x) - f(t)}{x - t} & \text{if } x \neq t \\
L & \text{if } x = t
\end{cases}
\]

is continuous at \( t \). Of course, in this case Equation (7.1) holds for all \( x \in \text{dom} \ f \), and we also have \( q(t) = f'(t) \). \( \square \)

Remark 7.8. To show \( f \) is differentiable at \( t \), it suffices to verify Equation (7.1) for \( x \neq t \).

Theorem 7.9 (Chain Rule). If \( f \) is differentiable at \( t \) and \( g \) is differentiable at \( f(t) \), then \( g \circ f \) is differentiable at \( t \) and

\[
(g \circ f)'(t) = g'(f(t))f'(t).
\]

Proof. Write

\[
f(x) = f(t) + q(x)(x - t)
\]

\[
g(y) = g(f(t)) + r(y)(y - f(t)),
\]

with \( q \) continuous at \( t \) and \( r \) continuous at \( f(t) \). Then

\[
g \circ f(x) = g \circ f(t) + r(f(x))(f(x) - f(t))
\]

\[
= g \circ f(t) + r(f(x))q(x)(x - t).
\]

Since \( (r \circ f)q \) is continuous at \( t \), \( g \circ f \) is differentiable at \( t \) and

\[
(g \circ f)'(t) = r \circ f(t)q(t) = g'(f(t))f'(t).
\]

\( \square \)

Lemma 7.10 (Critical Point Lemma). Let \( f : (a, b) \rightarrow \mathbb{R} \) and \( t \in (a, b) \). If \( f \) has a maximum or minimum at \( t \) and is differentiable at \( t \), then \( f'(t) = 0 \).

Proof. Without loss of generality assume \( f \) has a maximum at \( t \). For all \( x \in (a, t) \) we have

\[
\frac{f(x) - f(t)}{x - t} \geq 0,
\]

and letting \( x \uparrow t \) we get \( f'(t) \geq 0 \). On the other hand, for all \( x \in (t, b) \) we have

\[
\frac{f(x) - f(t)}{x - t} \leq 0,
\]

and letting \( x \downarrow t \) we get \( f'(t) \leq 0 \). Therefore, we must have \( f'(t) = 0 \). \( \square \)

Lemma 7.11 (Rolle’s Theorem). Let \( f \) be continuous on \([a, b]\) and differentiable on \((a, b)\). If \( f(a) = f(b) \), then there exists \( c \in (a, b) \) such that \( f'(c) = 0 \).
Proof. By the Extreme Value Theorem, \( f \) has a maximum and a minimum on \([a, b]\). Since \( f(a) = f(b) \), at least one of the maximum or minimum must occur at some \( c \in (a, b) \) (even if \( f \) is constant). By the Critical Point Lemma, \( f'(c) = 0 \).

Remark 7.12. As an immediate corollary of Rolle’s Theorem (but which is not really worth recording as one of our official results, although we will use it once), is the following: if \( f \) and \( g \) are continuous on \([a, b]\) and differentiable on \((a, b)\), with \( f(a) = g(a) \) and \( f(b) = g(b) \), then there exists \( c \in (a, b) \) such that \( f'(c) = g'(c) \). To see this, just apply Rolle’s Theorem to \( f - g \). Whenever this result is used, it can (by slight abuse of terminology) be referred to simply as Rolle’s Theorem.

**Theorem 7.13** (Cauchy’s Mean Value Theorem). If \( f \) and \( g \) are continuous on \([a, b]\) and differentiable on \((a, b)\), then there exists \( c \in (a, b) \) such that

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

Proof. Define \( h : [a, b] \to \mathbb{R} \) by

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

Then \( h \) is continuous, is differentiable on \((a, b)\), and we further have \( h(a) = h(b) \). By Rolle’s Theorem, there exists \( c \in (a, b) \) such that

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

and this implies the desired equation.

**Corollary 7.14** (Mean Value Theorem). If \( f \) is continuous on \([a, b]\) and differentiable on \((a, b)\), then there exists \( c \in (a, b) \) such that

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

Proof. Just apply Cauchy’s Mean Value Theorem with \( g(x) = x \).

**Proposition 7.15.** If \( f : [a, b] \to \mathbb{R} \) is continuous, then:

(i) \( f' > 0 \) on \((a, b)\) implies \( f \) is strictly increasing on \([a, b]\);
(ii) \( f' < 0 \) on \((a, b)\) implies \( f \) is strictly decreasing on \([a, b]\);
(iii) \( f' \geq 0 \) on \((a, b)\) implies \( f \) is increasing on \([a, b]\);
(iv) \( f' \leq 0 \) on \((a, b)\) implies \( f \) is decreasing on \([a, b]\);
(v) \( f' = 0 \) on \((a, b)\) implies \( f \) is constant on \([a, b]\).

Proof. In each case the desired behavior follows immediately from the Mean Value Theorem.

**Definition 7.16** (Higher Order Derivatives). (i) For \( n = 0, 1, \ldots \) the \( n \)th derivative of \( f \) is defined inductively as

\[
f^{(n)} := \begin{cases} f & \text{if } n = 0 \\ (f^{(n-1)})' & \text{if } n > 0. \end{cases}
\]
(ii) We also write
\[ \frac{d^n}{dx^n} f(x) = f^{(n)}(x). \]

Remark 7.17. We have never allowed ourselves to consider differentiating a function at an endpoint of its domain. However, it makes sense to restrict a differentiable function to a closed interval \([a, b]\) contained in its domain. It is occasionally convenient, as in the following theorem, to speak of a derivative on a closed interval \([a, b]\), and whenever we do this we have in mind that the function is differentiable on some open interval containing \([a, b]\).

It is interesting to note, however, that if we are considering a function \(f\) which is differentiable on \([a, b]\), we don’t need any information about the values of \(f\) outside \([a, b]\), even to compute the derivative at the endpoints, since we can use one-sided limits; for example, \(f'(a)\) is the right-hand limit of \((f(x) - f(a))/(x - a)\) as \(x \downarrow a\).

The following result is a generalization of the Mean Value Theorem to higher order derivatives:

**Theorem 7.18 (Taylor’s Theorem).** Let \(I\) be a closed interval with one endpoint \(a\), let \(J\) be the open interval with the same endpoints, and let \(n \in \mathbb{N}\). Assume \(f^{(n-1)}\) is continuous on \(I\) and differentiable on \(J\). Then for all \(x \in I\) there exists \(c\) between \(a\) and \(x\) such that
\[ f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)(x - a)^k}{k!} + \frac{f^{(n)}(c)(x - a)^n}{n!}. \]

**Proof.** First of all, if \(n = 1\) the result is just the Mean Value Theorem, so without loss of generality assume that \(n > 1\). Fix \(x \in I\), and define \(g: I \to \mathbb{R}\) by
\[ g(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)(t - a)^k}{k!} + M(t - a)^n, \]
where \(M \in \mathbb{R}\) is chosen such that
\[ f(x) = g(x). \]

Note that \(f^{(k)}(a) = g^{(k)}(a)\) for \(k = 0, \ldots, n - 1\). Since \(f(x) = g(x)\), by Rolle’s Theorem there exists \(c_1\) between \(a\) and \(x\) such that
\[ f'(c_1) = g'(c_1). \]

Then since \(f'(a) = g'(a)\), again by Rolle’s Theorem there exists \(c_2\) between \(a\) and \(c_1\) such that
\[ f''(c_2) = g''(c_2). \]

Continuing inductively, after \(n\) steps we get \(c := c_n\) between \(a\) and \(c_{n-1}\) such that \(f^{(n)}(c) = g^{(n)}(c)\). But \(g^{(n)}\) is identically \(Mn!\). Thus we get
\[ M = \frac{f^{(n)}(c)}{n!}. \]

Coupled with \(f(x) = g(x)\), this gives the desired result. \(\square\)
Theorem 7.19 (L’Hôpital’s Rule). Let \( a \) and \( L \) be extended real numbers. Assume \( \lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0 \) or \( \lim_{x \to a} g(x) = \pm \infty \). If

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

then

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

Proof. Without loss of generality it suffices to show that if \( L < M \) then there exists \( b > a \) such that for all \( x \in (a, b) \) we have

\[
\frac{f(x)}{g(x)} < M.
\]

Pick \( K \in (L, M) \), and choose \( b_1 > a \) such that for all \( x \in (a, b_1) \),

\[
\frac{f'(x)}{g'(x)} < K.
\]

Then whenever \( a < x < y < b_1 \), Cauchy’s Mean Value Theorem tells us there exists \( t \in (x, y) \) such that

\[
(7.2) \quad \frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(t)}{g'(t)} < K.
\]

Case 1. \( \lim f = \lim g = 0 \). Letting \( x \downarrow a \) in (7.2) gives

\[
\frac{f(y)}{g(y)} \leq K < M \quad \text{for all } y \in (a, b_1).
\]

Case 2. \( \lim g = \infty \). Choose \( b_2 \in (a, y) \) such that for all \( x \in (a, b_2) \),

\[
g(x), g(x) - g(y) > 0.
\]

Multiplying (7.2) by \( \frac{g(x) - g(y)}{g(x)} \) gives

\[
\frac{f(x) - f(y)}{g(x)} < K \frac{g(x) - g(y)}{g(x)},
\]

so

\[
\frac{f(x)}{g(x)} < \frac{f(y)}{g(x)} + K \left( 1 - \frac{g(y)}{g(x)} \right) \xrightarrow{x \downarrow a} K.
\]

Thus there exists \( b \in (a, b_2) \) such that for all \( x \in (a, b) \), \( \frac{f(x)}{g(x)} < M \).

Similarly for \( \lim g = -\infty \), \( \lim_{x \uparrow a} \), or \( M < L \). \( \square \)

Remark 7.20. It is interesting to note that l’Hôpital’s Rule in the “infinite” case only requires the denominator to have an infinite limit; however, if the numerator does not also have an infinite limit, then presumably l’Hôpital’s Rule is not needed.
Theorem 7.21 (Inverse Function Theorem). Let $f$ be 1-1 and continuous on an open interval containing $x$, and differentiable at $x$ with $f'(x) \neq 0$. Then $f^{-1}$ is differentiable at $f(x)$ and

$$(f^{-1})'(f(x)) = \frac{1}{f'(x)}.$$ 

Proof. By Theorem 6.15, $f^{-1}$ is continuous on an open interval $I$ containing $f(x)$. For all $y \in I \setminus \{ f(x) \}$,

$$\frac{f^{-1}(y) - f^{-1}(f(x))}{y - f(x)} = \frac{1}{f'(y)} \frac{y - f(x)}{f(y) - f(x)} \to \frac{1}{f'(x)},$$

since $y \to f(x)$ implies $f^{-1}(y) \to x$ by continuity of $f^{-1}$ at $f(x)$. \hfill $\square$

Remark 7.22. The above result can be used to extend the Power Rule to rational exponents.

8. Integration

Standing Hypothesis 8.1. Recall that we are assuming that all our functions $f$ satisfy $\text{dom } f, \text{ran } f \subseteq \mathbb{R}$ unless otherwise specified.

Definition 8.2. Let $-\infty < a < b < \infty$.

(i) A partition of $[a, b]$ is defined to be a finite set $P = \{ x_i \}_{i=0}^n$ such that

$$a = x_0 < x_1 < \cdots < x_n = b,$$

and $\mathcal{P}[a, b]$ denotes the set of all partitions of $[a, b]$.

(ii) Given $P = \{ x_i \}_{i=0}^n \in \mathcal{P}[a, b]$, the norm of $P$ is defined as

$$\|P\| := \max_i \Delta x_i,$$

where

$$\Delta x_i := x_i - x_{i-1} \quad \text{for } i = 1, \ldots, n.$$ 

(iii) If $P, Q \in \mathcal{P}[a, b]$, we say $Q$ refines $P$ if $Q \supseteq P$.

Remark 8.3. Thus, a partition of an interval is just a finite subset which includes the endpoints. Actually, in other areas of mathematics, the words “partition” and “norm” have completely different meanings.

Definition 8.4. Let $f$ be bounded on $[a, b]$ and $P = \{ x_i \}_{i=0}^n \in \mathcal{P}[a, b]$. The upper sum of $f$ associated to $P$ is defined as

$$U(P) = U(f, P) := \sum_{i=1}^n M_i \Delta x_i,$$

where

$$M_i = M_i(f) := \sup_{[x_{i-1}, x_i]} f,$$
and the lower sum of $f$ associated to $P$ is
\[ L(P) = L(f, P) := \sum_{i=1}^{n} m_i \Delta x_i, \]
where
\[ m_i = m_i(f) := \inf_{[x_{i-1}, x_i]} f. \]

**Lemma 8.5.**
(i) If $Q$ refines $P$ then
\[ L(P) \leq L(Q) \leq U(Q) \leq U(P). \]
(ii) For all $P, Q \in \mathcal{P}[a, b],$
\[ L(P) \leq U(Q). \]

**Proof.** (i) By induction, it suffices to assume $Q = P \cup \{s\}$ with $x_{i-1} < s < x_i$ (say). Put
\[ M'_i := \sup_{[x_{i-1}, s]} f, \quad M''_i := \sup_{[s, x_i]} f, \quad \Delta x'_i := s - x_{i-1}, \quad \text{and} \quad \Delta x''_i := x_i - s. \]
Then $M'_i, M''_i \leq M_i$ and $\Delta x_i = \Delta x'_i + \Delta x''_i$, so
\[ U(P) - U(Q) = M_i \Delta x_i - (M'_i \Delta x'_i + M''_i \Delta x''_i) \]
\[ = (M_i - M'_i) \Delta x'_i + (M_i - M''_i) \Delta x''_i \]
\[ \geq 0. \]
Similarly for $L$.

(ii) Let $R = P \cup Q$. Then $R$ refines both $P$ and $Q$, so by (i) we have
\[ L(P) \leq L(R) \leq U(R) \leq U(Q). \]

\[\square\]

**Definition 8.6.** Let $f$ be bounded on $[a, b]$. We say $f$ is **Riemann integrable** on $[a, b]$ if $\sup_{P \in \mathcal{P}[a, b]} L(P) = \inf_{P \in \mathcal{P}[a, b]} U(P)$, in which case this common value is called the **Riemann integral** of $f$ on $[a, b]$, written $\int_a^b f$ or $\int_a^b f(x) \, dx$.

**Example 8.7.** $\int_a^b 1 = b - a$.

**Remark 8.8.** It follows immediately from Lemma 8.5 that
\[ -\infty < \sup_P L(P) \leq \inf_P U(P) < \infty. \]

**Definition 8.9.** Let $f$ be bounded on $[a, b]$ and $P = \{x_i\}_{i=0}^{n} \in \mathcal{P}[a, b]$. Whenever $t_i \in [x_{i-1}, x_i]$ for $i = 1, \ldots, n$, the number
\[ \sum_{i=1}^{n} f(t_i) \Delta x_i \]
is called a **Riemann sum for $f$ associated to $P$**, and $\mathcal{R}(P) = \mathcal{R}(f, P)$ denotes the set of all Riemann sums for $f$ associated to $P$. 
Remark 8.10. For all $S \in \mathcal{R}(P)$ we have $L(P) \leq S \leq U(P)$. Thus, the set $\mathcal{R}(P)$ is a subset of the closed interval $[L(P), U(P)]$; it usually does not coincide with the interval, in fact it may or may not include the endpoints. However, it gets arbitrarily close to those endpoints:

Lemma 8.11. Let $f$ be bounded on $[a, b]$ and $P = \{x_i\}_{i=0}^n \in \mathcal{P}[a, b]$. Then:

(i) $U(P) = \sup \mathcal{R}(P)$;

(ii) $L(P) = \inf \mathcal{R}(P)$;

(iii) $U(P) - L(P) = \sup_{S,T \in \mathcal{R}(P)} |S - T|$.

Proof. (i) Clearly $U(P)$ is an upper bound for $\mathcal{R}(P)$. Let $\epsilon > 0$. For each $i = 1, \ldots, n$ choose $t_i \in [x_{i-1}, x_i]$ such that

$$M_i - f(t_i) < \frac{\epsilon}{b - a}.$$ 

Then

$$U(P) - \sum f(t_i) \Delta x_i = \sum (M_i - f(t_i)) \Delta x_i$$

$$< \sum \frac{\epsilon}{b - a} \Delta x_i$$

$$= \epsilon.$$ 

Thus $U(P) - \epsilon$ is not an upper bound for $\mathcal{R}(P)$. This proves (i), and (ii) is similar.

(iii) We have

$$\sup_{S,T \in \mathcal{R}(P)} |S - T| = \sup_{S,T \in \mathcal{R}(P)} (S - T)$$

$$= \sup (\mathcal{R}(P) - \mathcal{R}(P))$$

$$= \sup \mathcal{R}(P) + \sup (-\mathcal{R}(P))$$

$$= \sup \mathcal{R}(P) - \inf \mathcal{R}(P)$$

$$= U(P) - L(P),$$

by Parts (i) and (ii).

\[\square\]

Theorem 8.12. If $f$ is bounded on $[a, b]$, then the following are equivalent:

(i) $f$ is integrable on $[a, b]$.

(ii) There exists $I \in \mathbb{R}$ such that for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $P \in \mathcal{P}[a, b]$, if $\|P\| < \delta$ then for all $S \in \mathcal{R}(P)$ we have

$$|S - I| < \epsilon.$$ 

(iii) There exists $I \in \mathbb{R}$ such that for all $\epsilon > 0$ there exists a partition $P$ such that for all $S \in \mathcal{R}(P)$ we have

$$|S - I| < \epsilon.$$ 

(iv) For all $\epsilon > 0$ there exists $P \in \mathcal{P}[a, b]$ such that $U(P) - L(P) < \epsilon$. 
Proof. (i) ⇒ (ii). Assume $f$ is integrable, and let $\epsilon > 0$. We will show that there exists $\delta > 0$ such that if $P \in \mathcal{P}[a, b]$ with $\|P\| < \delta$ then $U(P) < \int_a^b f + \epsilon$.

This suffices, for then a similar argument would show we could decrease $\delta$ if necessary so that also $L(P) > \int_a^b f - \epsilon$, and then we would have $|S - \int_a^b f| < \epsilon$ for any $S \in \mathcal{R}(P)$. Claim: if $M = \sup_{[a, b]} |f|$, $P = \{x_i\}_0^n \in \mathcal{P}[a, b]$, $s \in [a, b]$, and $Q = P \cup \{s\}$, then

$$U(P) - U(Q) \leq 2M \|P\|.$$

Let $x_{i-1} < s < x_i$, and put

$$M'_i := \sup_{[x_{i-1}, s]} f, \quad M''_i := \sup_{[s, x_i]} f, \quad \Delta x'_i := s - x_{i-1}, \quad \text{and} \quad \Delta x''_i := x_i - s.$$

Then

$$U(P) - U(Q) = (M_i - M'_i) \Delta x'_i + (M_i - M''_i) \Delta x''_i$$

$$\leq 2M \Delta x'_i + 2M \Delta x''_i$$

$$= 2M \Delta x_i \leq 2M \|P\|,$$

and this verifies the claim. By induction, if $Q$ refines $P$ and has at most $k$ more points than $P$, then

$$U(P) - U(Q) \leq 2kM \|P\|.$$

Now, we can choose $R \in \mathcal{P}[a, b]$ such that $U(R) < \int_a^b f + \epsilon/2$. Let $R$ have $k$ points, and choose $\delta > 0$ such that $2kM \delta < \epsilon/2$. Let $P \in \mathcal{P}[a, b]$ with $\|P\| < \delta$, and put $Q = P \cup R$. Then by the above we have

$$U(P) \leq U(Q) + 2kM \|P\| \leq U(R) + 2kM \delta$$

$$< \int_a^b f + \frac{\epsilon}{2} + \frac{\epsilon}{2} = \int_a^b f + \epsilon,$$

as desired.

(ii) trivially implies (iii).

(iii) ⇒ (iv). Given $\epsilon > 0$ choose $P \in \mathcal{P}[a, b]$ such that $|S - T| < \epsilon/3$ for all $S \in \mathcal{R}(P)$. Then $|S - T| < 2\epsilon/3$ for all $S, T \in \mathcal{R}(P)$. Taking the sup over $S, T$ gives $U(P) - L(P) \leq 2\epsilon/3 < \epsilon$.

(iv) ⇒ (i). Assume $f$ is not integrable. Then $\epsilon := \inf_P U(P) - \inf_P L(P) > 0$, and for every $P \in \mathcal{P}[a, b]$ we have

$$U(P) - L(P) \geq \epsilon.$$

We have shown that (iv) if false if (i) is, so we are done.

\[
\square
\]

Remark 8.13. (i) The equivalence (i) ⇔ (ii) in the above theorem is called Darboux’s Theorem.

(ii) The condition (iv) in the above theorem is called Riemann’s Criterion.

(iii) In (ii) and (iii) of the above theorem, of course we have $I = \int_a^b f$.

**Theorem 8.14.** Every continuous function on $[a, b]$ is integrable on $[a, b]$. 
Proof. Let \( f : [a, b] \to \mathbb{R} \) be continuous. Since the interval \([a, b]\) is closed and bounded, \( f \) is uniformly continuous. Given \( \epsilon > 0 \), choose \( \delta > 0 \) such that for all \( x, y \in [a, b] \),

\[
|x - y| < \delta \quad \text{implies} \quad |f(x) - f(y)| < \frac{\epsilon}{2(b - a)}.
\]

Now choose \( P = \{x_i\}_0^n \in \mathcal{P}[a, b] \) such that \( \|P\| < \delta \). Then

\[
M_i - m_i \leq \frac{\epsilon}{2(b - a)} \quad \text{for all } i,
\]

so

\[
U(P) - L(P) = \sum (M_i - m_i) \Delta x_i \leq \frac{\epsilon}{2(b - a)} \sum \Delta x_i = \frac{\epsilon(b - a)}{2(b - a)} = \frac{\epsilon}{2} < \epsilon.
\]

\[ \square \]

Example 8.15. The Dirichlet function \( f \) on \([0, 1]\), defined by

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

is not integrable.

Proposition 8.16. If \( f \) and \( g \) are both integrable on \([a, b]\), then:

(i) \( \int_a^b (f + g) = \int_a^b f + \int_a^b g \);

(ii) \( \int_a^b cf = c \int_a^b f \) if \( c \in \mathbb{R} \).

Proof. (i) Given \( \epsilon > 0 \), choose \( P \in \mathcal{P}[a, b] \) such that if \( t_i \in [x_{i-1}, x_i] \) for all \( i \) then

\[
\left| \sum f(t_i) \Delta x_i - \int_a^b f \right|, \left| \sum g(t_i) \Delta x_i - \int_a^b g \right| < \frac{\epsilon}{2}.
\]

To see that this is possible, choose \( \delta_1, \delta_2 > 0 \) such that for all \( P \in \mathcal{P}[a, b] \), if \( \|P\| < \delta_1 \) then for all \( S \in \mathcal{R}(f, P) \) we have \( |S - \int_a^b f| < \epsilon/2 \), and if \( \|P\| < \delta_2 \) then for all \( T \in \mathcal{R}(g, P) \) we have \( |T - \int_a^b g| < \epsilon/2 \), and then we can take any \( P \) with \( \|P\| < \min\{\delta_1, \delta_2\} \). Then

\[
\left| \sum (f + g)(t_i) \Delta x_i - \left( \int_a^b f + \int_a^b g \right) \right| \leq \left| \sum f(t_i) \Delta x_i - \int_a^b f \right| + \left| \sum g(t_i) \Delta x_i - \int_a^b g \right| < \epsilon.
\]

(ii) Given \( \epsilon > 0 \), choose \( P \in \mathcal{P}[a, b] \) such that if \( t_i \in [x_{i-1}, x_i] \) for all \( i \) then

\[
\left| \sum f(t_i) \Delta x_i - \int_a^b f \right| < \frac{\epsilon}{|c| + 1}.
\]
Then
\[
\left| \sum c f(t_i) \Delta x_i - c \int_a^b f \right| = |c| \left| \sum f(t_i) \Delta x_i - \int_a^b f \right| \\
\leq \frac{|c| \epsilon}{|c| + 1} < \epsilon.
\]
\[\square\]

**Proposition 8.17.** Let \( a < b < c \). Then \( f \) is integrable on \([a, c]\) if and only if it is integrable on both \([a, b]\) and \([b, c]\), in which case
\[
\int_a^c f = \int_a^b f + \int_b^c f.
\]

*Proof.* First assume \( f \) is integrable on \([a, c]\). Given \( \epsilon > 0 \), choose \( P \in \mathcal{P}[a, c] \) such that \( U(P) - L(P) < \epsilon \). Without loss of generality \( b \in P \), since throwing \( b \) into \( P \) can only decrease \( U(P) - L(P) \). Put \( P' = P \cap [a, b] \) and \( P'' = P \cap [b, c] \). Then \( P' \in \mathcal{P}[a, b] \) and \( P'' \in \mathcal{P}[b, c] \), and we have
\[
U(P) - L(P) = U(P') - L(P') + U(P'') - L(P''),
\]
so
\[
U(P') - L(P'), U(P'') - L(P'') < \epsilon.
\]

Conversely, assume \( f \) is integrable on both \([a, b]\) and \([b, c]\). Given \( \epsilon > 0 \), choose \( P' \in \mathcal{P}[a, b] \) and \( P'' \in \mathcal{P}[b, c] \) such that
\[
\left| S' - \int_a^b f \right|, \left| S'' - \int_b^c f \right| < \frac{\epsilon}{2}
\]
for all \( S' \in \mathcal{R}(P') \) and \( S'' \in \mathcal{R}(P'') \). Put \( P = P' \cup P'' \in \mathcal{P}[a, c] \), and let \( S \in \mathcal{R}(P) \). Then \( S = S' + S'' \), with \( S' \in \mathcal{R}(P') \) and \( S'' \in \mathcal{R}(P'') \), respectively, so
\[
\left| S - \left( \int_a^b f + \int_b^c f \right) \right| \leq \left| S' - \int_a^b f \right| + \left| S'' - \int_b^c f \right| < \epsilon.
\]
Therefore, Theorem 8.12 tells us \( f \) is integrable on \([a, c]\), with \( \int_a^c f = \int_a^b f + \int_b^c f \). \[\square\]

**Definition 8.18.** Define
\[
\int_a^b f := \begin{cases} 
-\int_a^b f & \text{if } a > b \\
0 & \text{if } a = b.
\end{cases}
\]

**Corollary 8.19.** Let \( f \) be integrable on a closed interval \( I \). Then for all \( a, b, c \in I \),
\[
\int_a^c f = \int_a^b f + \int_b^c f.
\]

*Proof.* This follows from considering the cases determined by the possible relative positions of \( a, b, c \). \[\square\]
Proposition 8.20. Let $f$ and $g$ be integrable on $[a, b]$. If $f \leq g$, then $\int_a^b f \leq \int_a^b g$. In particular, if $m \leq f \leq M$ on $[a, b]$, then $m(b-a) \leq \int_a^b f \leq M(b-a)$.

Proof. For every $P \in \mathcal{P}[a, b]$ we have $\int_a^b f \leq U(f, P) \leq U(g, P)$. Thus $\int_a^b f \leq \int_a^b g$. □

Proposition 8.21. If $f$ is integrable on $[a, b]$, then so is $|f|$, and

$$\left|\int_a^b f\right| \leq \int_a^b |f|.$$  

Proof. Given $\epsilon > 0$, choose $P = \{x_i\}_{i=0}^n \in \mathcal{P}[a, b]$ such that $U(f, P) - L(f, P) < \epsilon$. For each $i = 1, \ldots, n$, if $s, t \in [x_{i-1}, x_i]$ then

$$|f(s)| - |f(t)| \leq |f(s) - f(t)| \leq M_i(f) - m_i(f),$$

so

$$M_i(|f|) - m_i(|f|) \leq M_i(f) - m_i(f).$$

Hence

$$U(|f|, P) - L(|f|, P) \leq U(f, P) - L(f, P) < \epsilon.$$  

Thus $|f|$ is integrable.

For the other part, $-|f| \leq f \leq |f|$ on $[a, b]$, so

$$-\int_a^b |f| \leq \int_a^b f \leq \int_a^b |f|.$$  

Hence

$$\left|\int_a^b f\right| \leq \int_a^b |f|.$$  

□

Proposition 8.22. If $f$ and $g$ are integrable on $[a, b]$, then so is $fg$.

Proof. Case 1. $f = g$. Put $M = \sup |f|$, and without loss of generality $M > 0$. Given $\epsilon > 0$, choose $P = \{x_i\}_{i=0}^n \in \mathcal{P}[a, b]$ such that $U(f, P) - L(f, P) < \epsilon/(2M)$. For each $i = 1, \ldots, n$, if $s, t \in [x_{i-1}, x_i]$ then

$$f^2(s) - f^2(t) = (f(s) + f(t))(f(s) - f(t)) \leq 2M(M_i(f) - m_i(f)).$$

Hence

$$M_i(f^2) - m_i(f^2) \leq 2M(M_i(f) - m_i(f)),$$

so

$$U(f^2, P) - L(f^2, P) \leq 2M(U(f, P) - L(f, P)) < \epsilon.$$  

Case 2. $f, g$ arbitrary. Then

$$fg = \frac{1}{4}((f + g)^2 - (f - g)^2),$$

which is integrable by the above. □
**Theorem 8.23** (Mean Value Theorem for Integrals). Let \( f, g : [a, b] \to \mathbb{R} \). If \( f \) is continuous and \( g \) is integrable and nonnegative, then there exists \( c \in [a, b] \) such that

\[
\int_a^b f g = f(c) \int_a^b g.
\]

**Proof.** Put \( m = \min f \) and \( M = \max f \). Then \( mg \leq fg \leq Mg \), so

\[
m \int_a^b g \leq \int_a^b fg \leq M \int_a^b g.
\]

If \( \int_a^b g = 0 \), then \( \int_a^b fg = 0 \) and the conclusion holds. On the other hand, if \( \int_a^b g > 0 \) then

\[
m \leq \frac{\int_a^b fg}{\int_a^b g} \leq M.
\]

Since \( f \) is continuous, by the Intermediate Value Theorem there exists \( c \in [a, b] \) such that

\[
f(c) = \frac{\int_a^b fg}{\int_a^b g},
\]

which implies the desired equality. \( \square \)

**Theorem 8.24.** If \( f \) is integrable on \([a, b]\), then the function \( x \mapsto \int_a^x f \) is continuous on \([a, b]\).

**Proof.** Put \( M = \sup |f| \). Given \( \epsilon > 0 \), choose \( \delta > 0 \) such that \( M\delta < \epsilon \). Then \( a \leq y \leq x \leq b \) and \( x - y < \delta \) imply

\[
\left| \int_a^x f - \int_a^y f \right| = \left| \int_y^x f \right| \leq \int_y^x |f| \leq M(x - y) < \epsilon.
\]

\( \square \)

**Theorem 8.25** (Fundamental Theorem of Calculus). Let \( f \) be integrable on \([a, b]\).

(i) Define \( F : [a, b] \to \mathbb{R} \) by \( F(x) = \int_a^x f \). For all \( t \in (a, b) \), if \( f \) is continuous at \( t \) then

\[
F'(t) = f(t).
\]

(ii) If \( G \) is continuous on \([a, b]\) and \( G' = f \) on \((a, b)\), then

\[
\int_a^b f = G(b) - G(a).
\]
Proof. (i) If \( x \in [a, b] \setminus \{t\} \) then
\[
\left| \frac{F(x) - F(t)}{x - t} - f(t) \right| = \left| \frac{\int_a^x f - \int_a^t f}{x - t} - f(t) \right|
\]
\[
= \left| \frac{\int_t^x f(s) \, ds}{x - t} - \frac{\int_t^t f(t) \, ds}{x - t} \right|
\]
\[
= \left| \frac{\int_t^x (f(s) - f(t)) \, ds}{x - t} \right|
\]
\[
\leq \sup \left\{ |f(s) - f(t)| : s \text{ between } t \text{ and } x \right\} \frac{|x - t|}{|x - t|}
\]
\[
= \sup \left\{ |f(s) - f(t)| : s \text{ between } t \text{ and } x \right\}
\]
\[
\xrightarrow{x \to t} 0.
\]

(ii) Given \( \epsilon > 0 \), choose \( P = \{x_i\}_0^n \subset \mathcal{P}[a, b] \) such that for all \( S \in \mathcal{R}(P) \) we have \( |S - \int_a^b f| < \epsilon \). By the Mean Value Theorem, for each \( i = 1, \ldots, n \) there exists \( t_i \in (x_{i-1}, x_i) \) such that
\[
G(b) - G(a) = \sum_{i=1}^n (G(x_i) - G(x_{i-1})) = \sum f(t_i) \Delta x_i.
\]
Hence
\[
\left| G(b) - G(a) - \int_a^b f \right| < \epsilon.
\]
Letting \( \epsilon \to 0 \) gives the conclusion. \( \square \)

Remark 8.26. We stated Part (ii) of the Fundamental Theorem of Calculus in a somewhat fussy way; in most cases, the (only slightly less general) statement “if \( G' \) is integrable on \([a, b]\), then \( \int_a^b G' = G(b) - G(a)\)” is good enough. However, note that this requires considering a derivative on a closed interval, and that our convention is that this tacitly means \( G \) extends to a differentiable function on an open interval containing \([a, b]\). In the next two theorems we use this less fussy (and more convenient) way of stating the hypotheses.

Theorem 8.27 (Integration by Parts). If \( f' \) and \( g' \) are integrable on \([a, b]\), then
\[
\int_a^b f'g = f(b)g(b) - f(a)g(a) - \int_a^b fg'.
\]
Proof. Since \( f \) and \( g \) are differentiable, they are continuous, hence integrable. Thus \( fg' \) and \( f'g \) are integrable. By the Fundamental Theorem of Calculus,
\[
f(b)g(b) - f(a)g(a) = \int_a^b (fg)' = \int_a^b f'g + \int_a^b f'g,
\]
implying the desired equality. \( \square \)
Theorem 8.28 (Change of Variables Theorem). If \( \phi' \) is integrable on \([a, b]\) and \( f \) is continuous on \( \phi([a, b]) \) then

\[
\int_a^b f(\phi(x))\phi'(x) \, dx = \int_{\phi(a)}^{\phi(b)} f(u) \, du.
\]

Proof. Define \( F: \phi([a, b]) \to \mathbb{R} \) by \( F(x) = \int_a^x f \). Then by the Fundamental Theorem of Calculus we have \( F' = f \) on \( \phi([a, b]) \) and \( (F \circ \phi)' = (f \circ \phi)\phi' \) on \([a, b]\) (where we used the Chain Rule for the latter). Since \( \phi \) is differentiable, it is continuous, and since \( f \) is also continuous, the composition \( f \circ \phi \) is continuous, hence integrable. Since \( \phi' \) is also integrable, the product \((f \circ \phi)\phi'\) is integrable. We apply the (other part of the) Fundamental Theorem of Calculus (twice) in the following computation to finish:

\[
\int_{\phi(a)}^{\phi(b)} f = F \circ \phi(b) - F \circ \phi(a) = \int_a^b (F \circ \phi)' = \int_a^b (f \circ \phi)\phi'.
\]

\[ \square \]

Remark 8.29. In the above proof, it might seem that we violated our rule prohibiting differentiating at an endpoint: the hypothesis only gave us \( f \) on the closed interval \( \phi([a, b]) \) (which might not have endpoints \( \phi(a) \) and \( \phi(b) \), by the way), and so the Fundamental Theorem of Calculus seems to only give us \( F' = f \) on the corresponding open interval (that is, the open interval with the same endpoints). However, everything is ok, because \( f \) can be extended to a continuous function on an open interval containing \( \phi([a, b]) \) (for example, just make \( f \) constant to the right and to the left).

Definition 8.30. (i) Let \( f \) be Riemann integrable on \([a, t]\) for every \( t \in (a, b) \). If either \( f \) is unbounded on \([a, b]\) or \( b = \infty \) we define

\[
\int_a^b f := \lim_{t \to b} \int_a^t f,
\]

provided this limit exists, in which case we say \( \int_a^b f \) exists as an improper integral. Note: throughout this definition, we require all limits to exist as real numbers, not extended real numbers.

(ii) Similarly, if \( f \) is Riemann integrable on \([t, b]\) for every \( t \in (a, b) \), and either \( f \) is unbounded on \((a, b]\) or \( a = -\infty \), we define

\[
\int_a^b f := \lim_{t \to a} \int_t^b f,
\]

provided this limit exists, in which case we again say \( \int_a^b f \) exists as an improper integral.
(iii) If \( f \) is Riemann integrable on \([a, b]\) whenever \( a < s < t < b \), and
either \( f \) is unbounded on \((a, b)\) or \( a \) or \( b \) is infinite, we pick any
\( c \in (a, b) \), and define
\[
\int_a^b f := \int_a^c f + \int_c^b f,
\]
provided each of \( \int_a^c f \) and \( \int_c^b f \) exists as either a Riemann or an
improper integral, in which case we again say \( \int_a^b f \) exists as an
improper integral.

Remark 8.31.  
(i) In case (i) of the above definition, if \( f \) is bounded,
\( b \in \mathbb{R} \), and \( f \) is defined at \( b \), then in fact the hypothesis guarantees
that \( f \) must be Riemann integrable on \([a, b]\), and then, by continuity
of the integral as a function of the upper limit of integration, we
have \( \int_a^b f = \lim_{\tau\uparrow b} \int_a^\tau f \). Similarly in cases (ii) and (iii).

(ii) On the other hand, again in case (i) above, if \( f \) is bounded and
\( b \in \mathbb{R} \), but \( f \) is not defined at \( b \), then we extend \( f \) to \( g: [a, b] \to \mathbb{R} \)
by assigning the value \( g(b) \) arbitrarily, and say (by slight abuse of
terminology) the original function \( f \) is Riemann integrable on \([a, b]\),
with \( \int_a^b f := \int_a^b g \). This latter value does not depend upon the
choice of \( g(b) \). Similarly in cases (ii) and (iii).

(iii) In case (i) above, if \( c \in (a, b) \) then \( \int_a^c f \) exists if and only if \( \int_c^b f \)
does, in which case \( \int_a^c f = \int_a^b f + \int_c^b f \). Similarly in case (ii).

(iv) In case (iii) above, the choice of \( c \in (a, b) \) does not matter.

Proposition 8.32.  
If each of \( \int_a^b f \) and \( \int_a^b g \) exists as a Riemann or improper
integral, then

(i) \( \int_a^b (f + g) = \int_a^b f + \int_a^b g \);

(ii) \( \int_a^b cf = c \int_a^b f \) if \( c \in \mathbb{R} \).

Proof.  
If necessary, split the interval \([a, b]\) into two, so that without loss of
generality \( f \) and \( g \) are integrable on \([a, t]\) for all \( t \in (a, b) \). For (i), we have
\[
\int_a^t (f + g) = \lim_{t \uparrow b} \int_a^t (f + g) = \lim_{t \uparrow b} \left( \int_a^t f + \int_a^t g \right)
= \lim_{t \uparrow b} \int_a^t f + \lim_{t \uparrow b} \int_a^t g
= \int_a^b f + \int_a^b g.
\]

For (ii), we have
\[
\int_a^b cf = \lim_{t \uparrow b} \int_a^t cf = \lim_{t \uparrow b} c \int_a^t f = c \lim_{t \uparrow b} \int_a^t f = c \int_a^b f.
\]

\( \square \)
**Theorem 8.33** (Comparison Theorem for Improper Integrals). Let \( f \) be Riemann integrable on \([s, t]\) whenever \( a < s < t < b \), and suppose \( \int_a^b g \) exists and \( 0 \leq f \leq g \). Then \( \int_a^b f \) exists, and

\[
\int_a^b f \leq \int_a^b g.
\]

*Proof.* If necessary, split the interval \([a, b]\) into two, so that without loss of generality \( f \) and \( g \) are integrable on \([a, t]\) for all \( t \in (a, b) \). Since \( 0 \leq f \leq g \), for all \( t \in (a, b) \) we have

\[
\int_a^t f \leq \int_a^t g \leq \int_a^b g.
\]

Since \( f \geq 0 \), the function \( t \mapsto \int_a^t f \) is increasing and bounded above, so the left hand limit \( \lim_{t \uparrow b} \int_a^t f \) exists, and we get

\[
\int_a^b f = \lim_{t \uparrow b} \int_a^t f \leq \int_a^b g.
\]

\( \square \)

**Example 8.34.** \( \int_1^\infty \frac{1}{x^p} \, dx \) exists if and only if \( p > 1 \). To verify this, first let \( p > 1 \). Then

\[
\lim_{t \to \infty} \frac{x^{1-p}}{1-p} \bigg|_1^t = \frac{1}{p-1},
\]

since \( \lim_{t \to \infty} t^a = 0 \) when \( a < 0 \). For \( p = 1 \),

\[
\lim_{t \to \infty} \log x \bigg|_1^t = \infty,
\]

so \( \int_1^\infty \frac{1}{x} \, dx \) does not exist. Finally, for \( p < 1 \) we have \( 1/x^p \geq 1/x \) for all \( x \geq 1 \), so by the Comparison Theorem \( \int_1^\infty \frac{1}{x^p} \, dx \) fails to exist.

**Corollary 8.35.** If \( f \) is Riemann integrable on \([s, t]\) whenever \( a < s < t < b \), and if \( \int_a^b |f| \) exists, then so does \( \int_a^b f \), and

\[
\left| \int_a^b f \right| \leq \int_a^b |f|.
\]

*Proof.* Since

\[
0 \leq |f| + f \leq 2|f|
\]

and \( \int_a^b 2|f| \) exists, by the Comparison Theorem so does \( \int_a^b (|f| + f) \), hence so does

\[
\int_a^b f = \int_a^b ((|f| + f) - |f|).
\]

For the other part, if \( a < s < t < b \) then

\[
\left| \int_s^t f \right| \leq \int_s^t |f| \leq \int_a^b |f|,
\]
so letting \( s \downarrow a \) and \( t \uparrow b \) we get
\[
\left| \int_a^b f \right| \leq \int_a^b |f|.
\]
}\]

\[\square\]

**Remark 8.36.** Combining the above corollary with the Comparison Theorem, we see that if \( f \) is Riemann integrable on \([s, t]\) whenever \( a < s < t < b \), and if \( f_a^b g \) exists and \( |f| \leq g \), then \( f_a^b f \) exists.

9. **Log and Exp**

**Definition 9.1** (The “Natural” Logarithm Function). Define \( \log \colon (0, \infty) \to \mathbb{R} \) by
\[
\log x = \int_1^x \frac{dt}{t}.
\]

**Proposition 9.2.**

(i) \( \log'(x) = \frac{1}{x} \);
(ii) \( \log 1 = 0 \);
(iii) \( \log xy = \log x + \log y \);
(iv) \( \log \frac{x}{y} = \log x - \log y \);
(v) \( \log x^n = n \log x \) for all \( n \in \mathbb{Z} \);
(vi) \( \log \) is strictly increasing;
(vii) \( \lim_{x \to 0^+} \log x = -\infty \) and \( \lim_{x \to \infty} \log x = \infty \);
(viii) \( \log \) is onto \( \mathbb{R} \).

**Proof.**

(i) This follows from the Fundamental Theorem of Calculus.

(ii) We have
\[
\log 1 = \int_1^1 \frac{dt}{t} = 0.
\]

(iii) Fix \( y > 0 \). Then
\[
\frac{d}{dx} \log xy = \frac{1}{xy} \frac{d}{dx} (xy) = \frac{1}{x} = \frac{d}{dx} \log x.
\]
Hence there exists \( c \in \mathbb{R} \) such that \( \log xy = \log x + c \) for all \( x > 0 \). Letting \( x = 1 \), we find \( c = \log y \).

(iv) This follows from (iii), since
\[
\log \frac{x}{y} + \log y = \log \left( \frac{x}{y} \right) = \log x.
\]

(v) This follows from (iii) and induction for \( n > 0 \), from (ii) for \( n = 0 \), and from (ii) and (iv) for \( n < 0 \).

(vi) Since \( \log' > 0 \), \( \log \) is strictly increasing by the Mean Value Theorem.

(vii) Since \( \log \) is increasing, it is either bounded below or \( \lim_{x \to 0^+} \log x = -\infty \). Since \( \log 2 > \log 1 = 0 \),
\[
\log 2^{-n} = -n \log 2 \xrightarrow{n \to \infty} -\infty,
\]
hence \( \log \) is not bounded below. Therefore, we must have \( \lim_{x \to 0^+} \log x = -\infty \).
Similarly, to show \( \lim_{x \to \infty} \log x = \infty \), it suffices to notice

\[
\log 2^n = n \log 2 \xrightarrow{n \to \infty} \infty.
\]

(viii) This follows from (vii) and the Intermediate Value Theorem, since \( \log \) is continuous. \( \square \)

**Remark 9.3.** In the proof of (vii) above, we used the letter \( n \) to (lazily) signal that we were considering a *sequence*.

**Definition 9.4 (The Exponential Function).** Define \( \exp : \mathbb{R} \to (0, \infty) \) by \( \exp = \log^{-1} \).

**Proposition 9.5.**

(i) \( \exp'(x) = \exp(x) \);

(ii) \( \exp(0) = 1 \);

(iii) \( \exp(x) \exp(y) = \exp(x + y) \);

(iv) \( \frac{\exp(x)}{\exp(y)} = \exp(x - y) \);

(v) \( \exp \) is strictly increasing;

(vi) \( \lim_{x \to -\infty} \exp(x) = 0 \) and \( \lim_{x \to \infty} \exp(x) = \infty \);

(vii) \( \exp \) is onto \( (0, \infty) \).

**Proof.**

(i) By the Inverse Function Theorem,

\[
\exp'(x) = \frac{1}{\log'(\exp(x))} = \exp(x).
\]

(ii) \( \exp(0) = \exp(\log 1) = 1 \).

(iii) We have

\[
\exp(x) \exp(y) = \exp\left(\log(\exp(x) \exp(y))\right)
= \exp\left(\log \exp(x) + \log \exp(y)\right)
= \exp(x + y).
\]

(iv) We have

\[
\frac{\exp(x)}{\exp(y)} = \exp\left(\log\left(\frac{\exp(x)}{\exp(y)}\right)\right)
= \exp\left(\log \exp(x) - \log \exp(y)\right)
= \exp(x - y).
\]

(v) This follows from Theorem 6.15, since \( \exp \) is the inverse of the strictly increasing function \( \log \).

(vi) This follows from the corresponding limits involving \( \log \).

(vii) The range of \( \exp \) is the domain of \( \log \), which is \( (0, \infty) \). \( \square \)

**Definition 9.6 (Arbitrary Powers).** For each \( x > 0 \) and \( t \in \mathbb{R} \) define

\[
x^t = \exp(t \log x).
\]

**Proposition 9.7.** For all \( x, y > 0 \) and \( t, s \in \mathbb{R} \):

(i) \( x^t x^s = x^{t+s} \);
(ii) \( \frac{x^t}{x^s} = x^{t-s} \);

(iii) \( \log x^t = t \log x \);

(iv) \( (x^t)^s = x^{ts} \);

(v) \( (xy)^t = x^t y^t \);

(vi) (Power Rule for Arbitrary Exponents) \( \frac{d}{dx} x^t = tx^{t-1} \).

**Proof.** (i) We have

\[
x^t x^s = \exp(t \log x) \exp(s \log x) = \exp(t \log x + s \log x) = \exp((t + s) \log x) = x^{t+s}.
\]

(ii) We have

\[
\frac{x^t}{x^s} = \frac{\exp(t \log x)}{\exp(s \log x)} = \exp(t \log x - s \log x) = \exp((t - s) \log x) = x^{t-s}.
\]

(iii) \( \log x^t = \log \exp(t \log x) = t \log x \).

(iv) \( (x^t)^s = \exp(s \log x^t) = \exp(st \log x) = x^{st} = x^{ts} \).

(v) We have

\[
(xy)^t = \exp(t \log xy) = \exp(t(\log x + \log y)) = \exp(t \log x + t \log y) = \exp(t \log x) \exp(t \log y) = x^t y^t.
\]

(vi) We have

\[
\frac{d}{dx} x^t = \frac{d}{dx} \exp(t \log x) = \exp(t \log x) \frac{t}{x} = tx^{t-1}.
\]

\( \Box \)

**Definition 9.8** (The Number \( e \)). \( e := \exp(1) \).

**Proposition 9.9.** \( e^x = \exp(x) \) for all \( x \in \mathbb{R} \).

**Proof.** \( e^x = \exp(x \log e) = \exp(x) \). \( \Box \)

**Example 9.10.** The exponential and logarithm functions, together with l'Hôpital’s Rule, allow easy computation of many limits which are in certain “exponential indeterminate” forms, for example:

(i) \( \lim_{x \to 0} x^x = 1 \);

(ii) \( \lim_{x \to \infty} x^{1/x} = 1 \);

(iii) \( \lim_{x \to \infty} \left(1 + \frac{t}{x}\right)^x = e^t \) if \( t \in \mathbb{R} \).
10. SERIES

Definition 10.1. (i) Given a sequence \( (a_n) \) in \( \mathbb{R} \), define a new sequence \( (s_k) \) by

\[
s_k = \sum_{n=1}^{k} a_n.
\]

The series \( \sum_{n=1}^{\infty} a_n \) is defined to be the sequence \( (s_k) \).

(ii) We say the series \( \sum_{n=1}^{\infty} a_n \) has \( n \)th term \( a_n \), and \( k \)th partial sum \( s_k \).

(iii) If the series \( \sum_{n=1}^{\infty} a_n \) converges (that is, if the sequence \( (s_k) \) of partial sums converges), the limit \( \lim_{k \to \infty} s_k \) is (ambiguously) denoted \( \sum_{n=1}^{\infty} a_n \), and called the sum of the series.

Remark 10.2. Series and sequences are just two ways of looking at the same thing. More precisely, not only does every series uniquely determine the sequence of partial sums (as in the above definition), but conversely every sequence of real numbers is the sequence of partial sums of a unique series: given the sequence \( (s_k) \), define the series \( \sum a_n \) by

\[
a_n = \begin{cases} 
  s_1 & \text{if } n = 1 \\
  s_n - s_{n-1} & \text{if } n > 1.
\end{cases}
\]

Remark 10.3. Just as with sequences, it is often convenient to allow a series to “start somewhere other than 1”, for example we frequently encounter series of the form \( \sum_{n=0}^{\infty} a_n \). Also, it is frequently convenient to refer to a series using the abbreviated notation “\( \sum a_n \)”, especially when discussing general properties. This abbreviated notation allows the starting point of the series to be anything, but in the development of the general theory we usually assume the starting point of the series is 1.

Example 10.4. If \( |x| < 1 \), the geometric series \( \sum_{n=0}^{\infty} x^n \) converges and

\[
\sum_{n=0}^{\infty} x^n = \frac{1}{1 - x}.
\]

To see this, note that an easy induction argument shows the identity

\[
\sum_{n=0}^{k} x^n = \frac{1 - x^{k+1}}{1 - x},
\]

and \( x^{k+1} \to 0 \) since \( |x| < 1 \). This is one of the rare instances where we can get a closed form expression for the partial sums. If \( c \in \mathbb{R} \) it is sometimes convenient to abuse the terminology by referring to a series of the form \( \sum cx^n \) (no matter what the starting point is) as a geometric series as well.
Example 10.5. The series $\sum_{n=1}^{\infty} 1/(n^2 + n)$ converges, because the $k$th partial sum is

$$s_k = \sum_{n=1}^{k} \frac{1}{n^2 + n} = \sum_{n=1}^{k} \left( \frac{1}{n} - \frac{1}{n + 1} \right) = 1 - \frac{1}{k + 1} \quad k \to \infty \to 1.$$ 

Again, we got lucky enough to find a closed form for $s_k$. In this case we say we have a telescoping sum.

Example 10.6. The harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges, because

$$s_k = \sum_{n=1}^{k} \frac{1}{n} \geq \int_{1}^{k+1} \frac{dx}{x} = \log(k + 1) \to \infty.$$ 

The technique of this example is generalized in the Integral Test below.

Remark 10.7. The following result could be phrased as “a series converges if and only if it is Cauchy”:

**Theorem 10.8.** $\sum a_n$ converges if and only if for all $\epsilon > 0$ there exists $l \in \mathbb{N}$ such that if $k \geq j \geq l$ then $|\sum_{j}^{k} a_n| < \epsilon$.

**Proof.** This follows from the corresponding result for sequences, since

$$|s_k - s_{j-1}| = \left| \sum_{j}^{k} a_n \right|.$$ 

\[ \square \]

Remark 10.9. The above theorem shows immediately that changing (or removing) finitely many terms of a series does not affect whether it converges (although of course it can affect the sum if the series converges).

**Corollary 10.10.** If $\sum a_n$ converges, then $a_n \to 0$.

**Proof.** This follows immediately from the preceding theorem, since

$$a_n = \sum_{n}^{\infty} a_k.$$ 

\[ \square \]

Remark 10.11. The above corollary only goes one way; it can happen that $a_n \to 0$ but $\sum a_n$ diverges: the most elementary example of this phenomenon is the harmonic series $\sum 1/n$.

Example 10.12. If $|x| \geq 1$, the geometric series $\sum_{n=0}^{\infty} x^n$ diverges. This follows immediately from the above corollary, since $|x| \geq 1$ implies $x^n \not\to 0$.

Remark 10.13. The following theorem says (roughly) that a series can sometimes be viewed as a sort of “infinite Riemann sum” for an improper integral:

**Theorem 10.14** (Integral Test). Let $f : [1, \infty) \to \mathbb{R}$ be decreasing and non-negative. Then the series $\sum_{n=1}^{\infty} f(n)$ converges if and only if the improper integral $\int_{1}^{\infty} f(x) \, dx$ exists.
Proof. Note that, since \( f \) is monotone, \( \int_1^t f \) exists for all \( t \in (1, \infty) \). Since \( f(n) \geq 0 \) for all \( n \), the sequence of partial sums
\[
s_k = \sum_{1}^{k} f(n)
\]
is increasing, hence converges if and only if it is bounded above. For \( k \geq 2 \),
\[
\sum_{2}^{k} f(n) \leq \int_1^{k} f \leq \sum_{1}^{k-1} f(n),
\]
so the sequence \( \left( \sum_{1}^{k} f(n) \right)_{k=1}^{\infty} \) is bounded above if and only if the sequence \( \left( \int_1^{k} f \right)_{k=1}^{\infty} \) is. Since \( f \geq 0 \), the function \( x \mapsto \int_1^{x} f \) is increasing, so \( \lim_{x \to \infty} \int_1^{x} f \) exists (in the real numbers) if and only if the sequence \( \left( \int_1^{k} f \right) \) is bounded above. The result follows. \( \square \)

Remark 10.15. In the Integral Test, the left endpoint of the interval \([1, \infty)\) can be replaced by any positive number; thus it is enough for \( f \) to be eventually decreasing and nonnegative (meaning there exists \( a > 0 \) such that the desired property holds on \([a, \infty)\)).

Example 10.16. The \( p \)-series \( \sum_{1}^{\infty} \frac{1}{n^p} \) converges if and only \( p > 1 \). This is most easily seen by considering the cases \( p > 0 \) and \( p \leq 0 \). When \( p > 0 \), the Integral Test applies with the nonnegative decreasing function \( x \mapsto 1/x^p \), and we know \( \int_1^{\infty} 1/x^p \, dx \) exists if and only if \( p > 1 \). When \( p \leq 0 \), the terms \( 1/n^p \) do not go to 0, so the series diverges.

Note that the harmonic series is the special case \( p = 1 \).

Proposition 10.17. If \( \sum a_n \) and \( \sum b_n \) both converge, then:

(i) \( \sum (a_n + b_n) = \sum a_n + \sum b_n \);
(ii) \( \sum ca_n = c \sum a_n \) if \( c \in \mathbb{R} \).

Proof. (i) \( \sum_{1}^{k} (a_n + b_n) = \sum_{1}^{k} a_n + \sum_{1}^{k} b_n \to \sum_{1}^{\infty} a_n + \sum_{1}^{\infty} b_n \).
(ii) \( \sum_{1}^{k} ca_n = c \sum_{1}^{k} a_n \to c \sum_{1}^{\infty} a_n \). \( \square \)

Theorem 10.18 (Comparison Test). If \( |a_n| \leq b_n \) for all \( n \) and \( \sum b_n \) converges, then \( \sum a_n \) converges and \( |\sum a_n| \leq \sum b_n \).

Proof. Given \( \epsilon > 0 \), choose \( l \in \mathbb{N} \) such that \( k \geq j \geq l \) implies \( \sum_{j}^{k} b_n < \epsilon \). Then \( k \geq j \geq l \) implies
\[
\left| \sum_{j}^{k} a_n \right| \leq \sum_{j}^{k} |a_n| \leq \sum_{j}^{k} b_n < \epsilon.
\]
This shows \( \sum a_n \) converges.

For the other part, for all \( k \in \mathbb{N} \) we have
\[
\left| \sum_{1}^{k} a_n \right| \leq \sum_{1}^{k} b_n .
\]
Letting \( k \to \infty \), we get
\[
\left| \sum_{1}^{\infty} a_n \right| \leq \sum_{1}^{\infty} b_n.
\]
\[\square\]

**Definition 10.19.** We say a series \( \sum a_n \) **converges absolutely** if \( \sum |a_n| \) converges.

**Corollary 10.20.** If \( \sum a_n \) converges absolutely, then \( \sum a_n \) converges and \( |\sum a_n| \leq \sum |a_n| \).

**Proof.** This follows immediately from the Comparison Test. \( \square \)

**Remark 10.21.** Thus, in the Comparison Test we can conclude that \( \sum a_n \) converges absolutely. Moreover, if we only have \( |a_n| \leq b_n \) for large \( n \) (meaning there exists \( k \in \mathbb{N} \) such that for all \( n \geq k \) the desired property holds), then we can still conclude \( \sum a_n \) converges absolutely, although we no longer have the inequality \( |\sum a_n| \leq \sum b_n \).

**Remark 10.22.** If both series have nonnegative terms, there is a useful contrapositive of the Comparison Test: if \( 0 \leq a_n \leq b_n \) for large \( n \) and \( \sum a_n \) diverges, then \( \sum b_n \) also diverges.

**Corollary 10.23 (Limit Comparison Test).** If the sequence of fractions \( (a_n/b_n) \) converges to a nonzero real number, then the series \( \sum a_n \) converges absolutely if and only if \( \sum b_n \) does. In particular, if \( a_n, b_n > 0 \) for all \( n \in \mathbb{N} \) then \( \sum a_n \) converges if and only if \( \sum b_n \) does.

**Proof.** Letting \( L = \lim a_n/b_n \), we have
\[
\frac{|L||b_n|}{2} < |a_n| < \frac{3|L||b_n|}{2}
\]
for large \( n \), so the result follows from the Comparison Test. \( \square \)

**Definition 10.24.** We say a series **converges conditionally** if it converges but not absolutely.

**Remark 10.25.** Conditionally convergent series are **delicate** (for example, see the remark following the Rearrangement Theorem below); the only general test we will prove which has the power to recognize conditional convergence is the following:

**Theorem 10.26 (Alternating Series Test).** If \( a_n \downarrow 0 \), then the alternating series \( \sum_{0}^{\infty} (-1)^n a_n \) converges, and moreover the sum satisfies the inequalities
\[
0 \leq \sum_{0}^{\infty} (-1)^n a_n \leq a_0.
\]

Moreover, these inequalities are strict if \( (a_n) \) is strictly decreasing.
MAT 371 LECTURES

Proof. For any $k \in \mathbb{N}$,

$$
\sum_{0}^{k} (-1)^{n}a_{n} = (a_{0} - a_{1}) + (a_{2} - a_{3}) + \cdots + \begin{cases} 
    a_{k-1} - a_{k} & \text{if } k \text{ odd} \\
    a_{k} & \text{if } k \text{ even}
\end{cases} \geq 0.
$$

On the other hand,

$$
\sum_{0}^{k} (-1)^{n}a_{n} = a_{0} - (a_{1} - a_{2}) - \cdots - \begin{cases} 
    a_{k-1} - a_{k} & \text{if } k \text{ even} \\
    a_{k} & \text{if } k \text{ odd}
\end{cases} \leq a_{0}.
$$

Given $\epsilon > 0$, choose $p \in \mathbb{N}$ such that $a_{p} < \epsilon$. Then $k \geq j \geq p$ implies

$$
\left| \sum_{j}^{k} (-1)^{n}a_{n} \right| \leq a_{j} \leq a_{p} < \epsilon,
$$

by the above reasoning. Therefore, $\sum (-1)^{n}a_{n}$ converges, and then the inequality

$$
0 \leq \sum_{0}^{k} (-1)^{n}a_{n} \leq a_{0} \quad \text{for all } k
$$

gives

$$
0 \leq \sum_{0}^{\infty} (-1)^{n}a_{n} \leq a_{0}.
$$

Moreover, if $(a_{n})$ is strictly decreasing, the even partial sums $s_{2k}$ are strictly decreasing and the odd partial sums $s_{2k+1}$ are strictly increasing, so the sum is strictly between 0 and $a_{0}$.

□

Remark 10.27. More generally, a series is called alternating if its signs alternate (although this allows for some 0 terms), and the Alternating Series Test says that if the absolute values of the terms of an alternating series decrease to 0 then the series converges and its sum is between 0 and the first term of the series.

Example 10.28. The alternating harmonic series $\sum_{1}^{\infty} (-1)^{n+1}/n$ converges by the Alternating Series Test, but the convergence is only conditional, since the series of absolute values is the harmonic series, which diverges. The Alternating Series Test also says the sum of the alternating harmonic series is strictly between 0 and 1; actually, we'll see later that the sum is log 2.

Theorem 10.29 (Rearrangement Theorem). If $\sum_{1}^{\infty} a_{n}$ converges absolutely and $f: \mathbb{N} \to \mathbb{N}$ is 1-1 onto, then $\sum_{1}^{\infty} a_{f(n)}$ converges absolutely and

$$
\sum_{1}^{\infty} a_{f(n)} = \sum_{1}^{\infty} a_{n}.
$$
Proof. Since $\sum |a_{f(n)}|$ is a rearrangement of $\sum |a_n|$, it suffices to show $\sum a_{f(n)}$ converges and $\sum a_{f(n)} = \sum a_n$. Given $\epsilon > 0$, choose $j \in \mathbb{N}$ such that $\sum_{j+1}^{\infty} |a_n| < \epsilon/2$. Now choose $l \in \mathbb{N}$ such that
\[
f\{1, \ldots, l\} \supseteq \{1, \ldots, j\}.
\]
Then $k \geq l$ implies
\[
\left| \sum_{1}^{k} a_{f(n)} - \sum_{1}^{\infty} a_n \right| \leq \left| \sum_{1}^{k} a_{f(n)} - \sum_{1}^{j} a_n \right| + \left| \sum_{1}^{j} a_n - \sum_{1}^{\infty} a_n \right|
\]
\[
= \left| \sum_{n \leq k}^{\infty} a_{f(n)} \right| + \left| \sum_{j+1}^{\infty} a_n \right|
\]
\[
\leq \sum_{n \leq k}^{\infty} |a_{f(n)}| + \sum_{j+1}^{\infty} |a_n|
\]
\[
\leq 2 \sum_{j+1}^{\infty} |a_n| < \epsilon.
\]

Remark 10.30. What if $\sum a_n$ converges conditionally? Then it turns out that there exist divergent rearrangements, and moreover for any $s \in \mathbb{R}$ there exists a rearrangement with sum $s$. This is called Riemann’s Rearrangement Theorem; since we do not need it, we do not include the proof. As an example of this phenomenon, let $s$ denote the sum of the alternating harmonic series $\sum_{1}^{\infty} (-1)^{n+1}/n$. The rearrangement formed by alternately taking two positive terms and one negative term converges to $3s/2$.

Theorem 10.31 (Ratio Test). Assume $a_n \neq 0$ for all $n$.

(i) If $\lim \sup \left| \frac{a_{n+1}}{a_n} \right| < 1$ then $\sum a_n$ converges absolutely, while

(ii) if $\lim \inf \left| \frac{a_{n+1}}{a_n} \right| > 1$ then $\sum a_n$ diverges.

Proof. (i) If $\lim \sup \left| \frac{a_{n+1}}{a_n} \right| < t < 1$, then there exists $k \in \mathbb{N}$ such that $n \geq k$ implies
\[
\left| \frac{a_{n+1}}{a_n} \right| < t.
\]
Then for all $j \in \mathbb{N}$,
\[
|a_{k+j}| \leq t|a_{k+j-1}| \leq t^2|a_{k+j-2}| \leq \cdots \leq t^j|a_k|.
\]
Thus the tail $\sum_{k+1}^{\infty} a_n$ converges absolutely by comparison to the geometric series $\sum t^n |a_n|$, so $\sum_{k=1}^{\infty} a_n$ converges absolutely also.

(ii) In this case there exists $k \in \mathbb{N}$ such that $n \geq k$ implies

$$\left| \frac{a_{n+1}}{a_n} \right| > 1,$$

so

$$0 < |a_k| \leq |a_{k+1}| \leq \cdots.$$

Hence $a_n \not\to 0$, thus $\sum a_n$ diverges.

\[ \square \]

Remark 10.32. It often happens that $L := \lim |a_{n+1}/a_n|$ exists, and then the Ratio Test says $\sum a_n$ converges absolutely if $L < 1$ and diverges if $L > 1$. However, if $L = 1$ the series could either converge or diverge. For example, the harmonic series $\sum 1/n$ diverges, while the $p$-series $\sum 1/n^p$ converges, but in both cases the ratios $a_{n+1}/a_n$ converge to 1.

Theorem 10.33 (Root Test). Put $\rho = \limsup |a_n|^{1/n}$.

(i) If $\rho < 1$ then $\sum a_n$ converges absolutely, while

(ii) if $\rho > 1$ then $\sum a_n$ diverges.

Proof. (i) If $\rho < t < 1$, then there exists $k \in \mathbb{N}$ such that $n \geq k$ implies $|a_n|^{1/n} < t$, so $|a_n| < t^n$. Hence $\sum a_n$ converges absolutely by comparison to the geometric series $\sum t^n$.

(ii) In this case for all $k \in \mathbb{N}$ there exists $n \geq k$ such that $|a_n|^{1/n} > 1$, so $|a_n| > 1$. Hence $a_n \not\to 0$, so $\sum a_n$ diverges.

\[ \square \]

Remark 10.34. Similarly to the Ratio Test, if $\rho = 1$ the series could either converge or diverge, and the same examples $\sum 1/n$ and $\sum 1/n^2$ apply.

11. SEQUENCES OF FUNCTIONS

Standing Hypothesis 11.1. Recall that we are assuming that all our functions $f$ satisfy $\text{dom } f$, ran $f \subseteq \mathbb{R}$ unless otherwise specified.

Definition 11.2. Let $f, f_1, f_2, \ldots : A \to \mathbb{R}$. We say the sequence $(f_n)$ of functions:

(i) **converges pointwise** to $f$ if for every $x \in A$ we have $f_n(x) \to f(x)$.

We also write $f_n \to f$ or $f = \lim f_n$;

(ii) **converges uniformly** to $f$ if for all $\epsilon > 0$ there exists $k \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ and $x \in A$,

$$\text{if } n \geq k \text{ then } |f_n(x) - f(x)| < \epsilon.$$

Remark 11.3. It follows almost immediately from the definition that $f_n \to f$ uniformly if and only if $\sup_{x \in A} |f_n(x) - f(x)| \to 0$.

Remark 11.4. Note that $f_n \to f$ pointwise if and only if for all $x \in A$ and $\epsilon > 0$ there exists $k \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ we have $n \geq k$ implies $|f_n(x) - f(x)| < \epsilon$. The definition of uniform convergence moves the universally quantified $x$ across the existentially quantified $k$, producing a stronger condition in that now a single $k$ has to work for all $x$ simultaneously.
Therefore, basic logic tells us that uniform convergence implies pointwise convergence.

**Theorem 11.5 (Uniform Cauchy Criterion).** Let \( f_1, f_2, \ldots : A \to \mathbb{R} \). Then \((f_n)\) converges uniformly if and only if for all \( \epsilon > 0 \) there exists \( k \in \mathbb{N} \) such that for all \( x \in A \) and \( n, j \in \mathbb{N} \),

\[
\text{if} \quad n, j \geq k \quad \text{then} \quad |f_n(x) - f_j(x)| < \epsilon.
\]

**Proof.** First assume \( f_n \to f \) uniformly. Given \( \epsilon > 0 \), choose \( k \in \mathbb{N} \) such that for all \( x \in A \) and \( n \geq k \) implies \( |f_n(x) - f(x)| < \epsilon/2 \). Then for all \( x \in A \), if \( n, j \geq k \) then

\[
|f_n(x) - f_j(x)| \leq |f_n(x) - f(x)| + |f(x) - f_j(x)| < \epsilon.
\]

Conversely, assume \((f_n)\) satisfies the Uniform Cauchy Criterion. Then for all \( x \in A \), the sequence \((f_n(x))\) is Cauchy, so there exists \( f(x) \in \mathbb{R} \) such that \( f_n(x) \to f(x) \). Given \( \epsilon > 0 \), choose \( k \in \mathbb{N} \) such that for all \( x \in A \) and \( n, j \in \mathbb{N} \), if \( n, j \geq k \) then \( |f_n(x) - f_j(x)| < \epsilon/2 \). Fix \( n \geq k \) and \( x \in A \), and let \( j \to \infty \) to get

\[
|f_n(x) - f(x)| \leq \frac{\epsilon}{2} < \epsilon.
\]

\[\square\]

**Theorem 11.6.** Let \( f, f_1, f_2, \ldots : A \to \mathbb{R} \) and \( t \in A \). If \( f_n \to f \) uniformly and each \( f_n \) is continuous at \( t \), then \( f \) is also continuous at \( t \).

**Proof.** Given \( \epsilon > 0 \), choose \( k \in \mathbb{N} \) such that \( |f_k(x) - f(x)| < \epsilon/3 \) for all \( x \in A \). Now choose \( \delta > 0 \) such that \( |x - t| < \delta \) implies \( |f_k(x) - f_k(t)| < \epsilon/3 \). Then \( |x - t| < \delta \) implies

\[
|f(x) - f(t)| \leq |f(x) - f_k(x)| + |f_k(x) - f_k(t)| + |f_k(t) - f(t)| < \epsilon.
\]

\[\square\]

**Theorem 11.7.** Let \( f, f_1, f_2, \ldots : [a, b] \to \mathbb{R} \). If \( f_n \to f \) uniformly and each \( f_n \) is integrable, then \( f \) is integrable and

\[
\int_a^b f_n \to \int_a^b f.
\]

**Proof.** We first show the sequence \((\int_a^b f_n)\) converges. Given \( \epsilon > 0 \), choose \( k \in \mathbb{N} \) such that \( n, j \geq k \) imply

\[
|f_n(x) - f_j(x)| < \frac{\epsilon}{2(b - a)} \quad \text{for all} \quad x \in [a, b].
\]

Then \( n, j \geq k \) imply

\[
\left| \int_a^b f_n - \int_a^b f_j \right| \leq \int_a^b |f_n - f_j| \leq \frac{\epsilon}{2(b - a)} (b - a) = \frac{\epsilon}{2} < \epsilon.
\]

Thus \( I := \lim_n \int_a^b f_n \) exists.
Now we show $f$ is integrable and $\int_a^b f = I$. Given $\epsilon > 0$, choose $k \in \mathbb{N}$ such that both

$$|f_k(x) - f(x)| < \frac{\epsilon}{3(b-a)} \quad \text{for all } x \in [a,b]$$

and

$$\left| \int_a^b f_k - I \right| < \frac{\epsilon}{3}.$$ 

Choose a partition $P = \{x_i\}^n_0$ such that $t_i \in [x_{i-1}, x_i]$ for all $i$ implies

$$\left| \sum_i f_k(t_i) \Delta x_i - \int_a^b f_k \right| < \frac{\epsilon}{3}.$$ 

Then

$$\left| \sum f(t_i) \Delta x_i - I \right| \leq \left| \sum f(t_i) \Delta x_i - \sum f_k(t_i) \Delta x_i \right|$$

$$+ \left| \sum f_k(t_i) \Delta x_i - \int_a^b f_k \right| + \left| \int_a^b f_k - I \right|$$

$$< \sum |f(t_i) - f_k(t_i)| \Delta x_i + \frac{2\epsilon}{3}$$

$$< \frac{\epsilon}{3(b-a)}(b-a) + \frac{2\epsilon}{3}$$

$$= \epsilon.

\square

**Theorem 11.8.** Let $(f_n)$ be a sequence of differentiable functions on $(a,b)$. If $(f'_n)$ converges uniformly on $(a,b)$ and $(f_n(c))$ converges for some $c \in (a,b)$, then $(f_n)$ converges pointwise on $(a,b)$ to a differentiable function $f$, and

$$f'_n \to f'$$

pointwise on $(a,b)$. Moreover, if $(a,b)$ is bounded then $f_n \to f$ uniformly.

**Proof.** For each $n \in \mathbb{N}$ and $t \in (a,b)$ define $g_n^t : (a,b) \to \mathbb{R}$ by

$$g_n^t(x) = \begin{cases} 
  f_n(x) - f_n(t) & \text{if } x \neq t \\
  \frac{f'_n(t)}{x-t} & \text{if } x = t.
\end{cases}$$

Claim: the sequence $(g_n^t)_{n=1}^\infty$ converges uniformly. Given $\epsilon > 0$, choose $k \in \mathbb{N}$ such that $n, j \geq k$ and $x \in (a,b)$ imply

$$|f'_n(x) - f'_j(x)| < \epsilon.$$
Then \( n, j \geq k \) and \( x \neq t \) imply
\[
|g_n'(x) - g_j'(x)| = \left| \frac{(f_n - f_j)(x) - (f_n - f_j)(t)}{x - t} \right|
= \left| (f_n - f_j)'(s) \right| \quad \text{for some } s \text{ between } x \text{ and } t
< \epsilon.
\]
Also,
\[
|g_n'(t) - g_j'(t)| = |f_n'(t) - f_j'(t)| < \epsilon,
\]
and the claim is verified. Put \( g^t := \lim g_n^t \). Each \( g_n^t \) is continuous at \( t \), hence so is \( g^t \). Note that for all \( x, t \in (a, b) \),
\[
(*) \quad f_n(x) = f_n(t) + g_n^t(x)(x - t).
\]
Letting \( t = c \), we see that \( (f_n) \) converges at \( x \). Put \( f := \lim f_n \). Letting \( n \to \infty \) in \( (*) \), we get
\[
f(x) = f(t) + g^t(x)(x - t).
\]
Since \( g^t \) is continuous at \( t \), \( f \) is differentiable at \( t \) and
\[
f'(t) = g^t(t) = \lim f_n'(t) = \lim f_n'(t).
\]
For the other part, just note that if \( (a, b) \) is bounded then \( |x - t| \leq b - a \) in \( (*) \), which implies the convergence of \( f_n \) to \( f \) is uniform.

\section*{12. Series of Functions}

\textbf{Standing Hypothesis 12.1.} Recall that we are assuming that all our functions \( f \) satisfy \( \text{dom } f, \text{ran } f \subseteq \mathbb{R} \) unless otherwise specified.

\textbf{Definition 12.2.} \( \text{(i) Given a sequence } (f_n) \text{ of real-valued functions on } A, \text{the series } \sum_{n=1}^{\infty} f_n \text{ is defined to be the sequence } \left( \sum_{n=1}^{k} f_n \right)_{k=1}^{\infty} \text{ of partial sums.} \)
\( \text{(ii) If the series } \sum_{n=1}^{\infty} f_n \text{ converges pointwise (that is, if the sequence}\n\text{of partial sums converges pointwise), the limit } \lim_{k \to \infty} \sum_{n=1}^{k} f_n \text{ is } \)\n(ambiguously) denoted \( \sum_{n=1}^{\infty} f_n \), and called the \textit{sum} of the series.
\( \text{(iii) Of course, we say } \sum_{n=1}^{\infty} f_n \text{ \textit{converges uniformly} if the sequence}\n\text{of partial sums converges uniformly.} \)
\( \text{(iv) We say } \sum_{n=1}^{\infty} f_n \text{ \textit{converges absolutely} if for every } x \in A \text{ the series } \sum_{n=1}^{\infty} f_n(x) \text{ of real numbers converges absolutely.} \)

\textbf{Proposition 12.3} (Uniform Cauchy Criterion for Series). \( \sum f_n \) converges uniformly on \( A \) if and only if for all \( \epsilon > 0 \) there exists \( l \in \mathbb{N} \) such that for all \( x \in A \) and \( k, j \in \mathbb{N} \),
\[
\text{if } k \geq j \geq l \quad \text{then } \left| \sum_{j}^{k} f_n(x) \right| < \epsilon.
\]

\textit{Proof.} This follows immediately from the Uniform Cauchy Criterion for sequences of functions. \( \square \)
Theorem 12.4 (Weierstrass M-Test). Assume there exists a nonnegative sequence \((M_n)\) in \(\mathbb{R}\) such that both

(i) \(\sum M_n < \infty\), and

(ii) for all \(n \in \mathbb{N}\) and \(x \in A\) we have \(|f_n(x)| \leq M_n\).

Then \(\sum f_n\) converges uniformly and absolutely on \(A\).

Proof. Given \(\epsilon > 0\), choose \(l \in \mathbb{N}\) such that \(k \geq j \geq l\) implies \(\sum_j^k M_n < \epsilon\). Then \(k \geq j \geq l\) and \(x \in A\) imply

\[
\left| \sum_j^k f_n(x) \right| \leq \sum_j^k |f_n(x)| \leq \sum_j^k M_n < \epsilon.
\]

\(\square\)

Example 12.5. \(\sum x^n\) converges absolutely on \((-1, 1)\), and uniformly on every interval of the form \([-s, s]\) with \(0 < s < 1\).

Theorem 12.6. If \(\sum f_n\) converges uniformly and each \(f_n\) is continuous at \(t\), then the sum \(\sum f_n\) is continuous at \(t\).

Proof. Immediate from Theorem 11.6.

\(\square\)

Theorem 12.7. If \(\sum f_n\) is a uniformly convergent series of integrable functions on \([a, b]\), then the sum \(\sum f_n\) is integrable and

\[
\int_a^b \sum f_n = \sum \int_a^b f_n.
\]

Proof. Immediate from Theorem 11.7.

\(\square\)

Theorem 12.8. Let \((\sum f_n)\) be a series of differentiable functions on \((a, b)\). If \(\sum f_n\) converges uniformly on \((a, b)\) and \(\sum f_n(c)\) converges for some \(c \in (a, b)\), then \(\sum f_n\) converges pointwise on \((a, b)\), the sum \(\sum f_n\) is differentiable, and

\[
\left( \sum f_n \right)' = \sum f'_n
\]

on \((a, b)\). Moreover, if \((a, b)\) is bounded then \(\sum f_n\) converges uniformly.

Proof. Immediate from Theorem 11.8.

\(\square\)

13. Power Series

Definition 13.1. Let \((c_n)_{n=0}^\infty\) be a real-valued sequence and let \(a \in \mathbb{R}\). The series \(\sum_{n=0}^\infty c_n (x - a)^n\) is called the power series with coefficients \(c_n\) and center \(a\).

Theorem 13.2 (Cauchy-Hadamard Theorem). Let \(\sum_{n=0}^\infty c_n (x - a)^n\) be a power series, and put

\[
r := \limsup_{n \to \infty} |c_n|^{1/n}.
\]
interpreted as 0 if the lim sup is \( \infty \), and \( \infty \) if the lim sup is 0. Then for all \( x \in \mathbb{R} \),

(i) if \( |x - a| < r \) then \( \sum c_n (x - a)^n \) converges absolutely, while

(ii) if \( |x - a| > r \) then \( \sum c_n (x - a)^n \) diverges.

Moreover, \( \sum c_n (x - a)^n \) converges uniformly on every interval of the form \( [a - s, a + s] \) with \( 0 < s < r \).

**Proof.** We have

\[
\limsup |c_n (x-a)^n|^{1/n} = |x-a| \limsup |c_n|^{1/n} = \frac{|x-a|}{r},
\]

so the first part follows from the Root Test.

For the other part, let \( 0 < s < r \), and pick \( t \in (s, r) \). Then

\[
\frac{1}{t} > \frac{1}{r} = \limsup |c_n|^{1/n},
\]

so there exists \( k \in \mathbb{N} \) such that \( n \geq k \) implies

\[
|c_n|^{1/n} < \frac{1}{t}.
\]

Thus \( n \geq k \) and \( |x-a| \leq s \) imply

\[
|c_n (x-a)^n| \leq \left( \frac{s}{t} \right)^n.
\]

Since \( \sum \left( \frac{s}{t} \right)^n \) is a convergent geometric series, the power series \( \sum c_n (x-a)^n \) converges uniformly on \( [a - s, a + s] \) by the Weierstrass M-Test.

**Definition 13.3.** With the above notation, \( r \) is called the **radius of convergence** of the power series \( \sum c_n (x-a)^n \).

**Example 13.4.**

(i) The power series \( \sum n^n x^n \) has radius of convergence 0.

(ii) The power series \( \sum x^n/n^n \) has radius of convergence \( \infty \).

(iii) For every real number \( p \) the power series \( \sum x^n/n^p \) has radius of convergence 1 (because \( (n^p)^{1/n} = (n^{1/n})^p \to 1^p = 1 \)).

**Remark 13.5.** The following theorem tells us differentiating and integrating a power series term-by-term preserves the radius of convergence and gives a series whose sum is the derivative or integral of the original:

**Theorem 13.6.** If the power series \( \sum c_n (x-a)^n \) has radius of convergence \( r \), then both power series

\[
\sum n c_n (x-a)^{n-1} \quad \text{and} \quad \sum \frac{c_n}{n+1} (x-a)^{n+1}
\]

also have radius of convergence \( r \). Moreover, if \( r > 0 \) and we define \( f: (a-r, a+r) \to \mathbb{R} \) by \( f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n \), then for all \( x \in (a-r, a+r) \) we have

\[
f'(x) = \sum_{n=1}^{\infty} n c_n (x-a)^{n-1} \quad \text{and} \quad \int_a^x f = \sum_{n=0}^{\infty} \frac{c_n}{n+1} (x-a)^{n+1}.
\]
Proof. First of all, multiplying by $x - a$ term-by-term does not change the radius of convergence, so $\sum n c_n (x-a)^{n-1}$ has the same radius of convergence as $\sum n c_n (x-a)^n$. Since $n^{1/n} \to 1$, the Cauchy-Hadamard Theorem implies that $\sum n c_n (x-a)^n$ has the same radius of convergence as $\sum c_n (x-a)^n$. Thus the power series $\sum n c_n (x-a)^{n-1}$ has radius of convergence $r$. Since the series $\sum c_n (x-a)^n$ is obtained from $\sum \frac{c_n}{n+1} (x-a)^{n+1}$ by differentiating term-by-term, it follows that $\sum \frac{c_n}{n+1} (x-a)^{n+1}$ must also have radius of convergence $r$.

For the other part, fix $x$ in $(a-r,a+r)$, and choose $s \in (0,r)$ such that $x \in (a-s,a+s)$. Since the power series $\sum n c_n (x-a)^{n-1}$ converges uniformly on $(a-s,a+s)$ and the series $\sum c_n (x-a)^n$ converges at some point in $(a-s,a+s)$, Theorem 12.8 (on differentiating series) tells us $f$ is differentiable and

$$f'(x) = \sum_0^\infty \frac{d}{dx} c_n (x-a)^n = \sum_1^\infty n c_n (x-a)^{n-1}.$$ 

Similarly, since $\sum c_n (t-a)^n$ converges uniformly for $t$ in the closed interval with endpoints $a$ and $x$, Theorem 12.7 (on integrating series) tells us

$$\int_a^x f(t) \, dt = \sum_0^\infty \int_a^x c_n (t-a)^n \, dt = \sum_0^\infty \frac{c_n}{n+1} (x-a)^{n+1}.$$ 

\[\square\]

**Corollary 13.7.** If $f(x) = \sum c_n (x-a)^n$ for all $x$ in an open interval containing $a$, then the coefficients are given by

$$c_n = \frac{f^{(n)}(a)}{n!} \quad \text{for all } n = 0, 1, \ldots.$$

**Proof.** An induction argument shows

$$f^{(n)}(x) = \sum_0^\infty \frac{k!}{(k-n)!} c_k (x-a)^{k-n}.$$ 

Thus

$$f^{(n)}(a) = n! c_n.$$ 

\[\square\]

**Definition 13.8.** If $f$ is infinitely differentiable at $a$, the power series

$$\sum_{n=0}^\infty \frac{f^{(n)}(a)(x-a)^n}{n!}$$

is called the Taylor series of $f$ at $a$. 
Remark 13.9. The Taylor series of $f$ may have a nonzero radius of convergence but only converge to $f$ at $a$. For example, if
\[ f(x) = \begin{cases} 
\frac{e^{-1/x^2}}{x} & \text{if } x \neq 0 \\
0 & \text{if } x = 0 
\end{cases} \]
then $f^{(n)}(0) = 0$ for every $n$, so the Taylor series of $e^{-1/x^2}$ at 0 has radius of convergence $\infty$, but this Taylor series converges to $e^{-1/x^2}$ only at $x = 0$. To see this, note that an induction argument using l'Hôpital's Rule shows that for each $n = 0, 1, \ldots$ we have
\[ f^{(n)}(x) = \begin{cases} 
g(x)e^{-1/x^2} & \text{if } x \neq 0 \\
0 & \text{if } x = 0, \end{cases} \]
where $g$ is a rational function.

Example 13.10. The Taylor series of $\frac{1}{1-x}$ at 0 is $\sum_{0}^{\infty} x^n$, which converges to $\frac{1}{1-x}$ on the interval $(-1, 1)$. Integrating term-by-term, we find the Taylor series of $-\log(1-x)$ at 0. Multiplying by $-1$ and substituting $-x$ for $x$, we get the Taylor series for $\log(1+x)$ at 0, which converges to $\log(1+x)$ on $(-1, 1)$:
\[ \log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots. \]
We can also see that this is valid at 1, that is,
\[ \log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \cdots, \]
by using the Alternating Series Test to show that the series $\sum_{1}^{\infty} (-1)^{n+1}x^n/n$ converges uniformly on $[0, 1]$. To see the uniform convergence, given $\epsilon > 0$ choose $l \in \mathbb{N}$ such that $1/l < \epsilon$. If $x \in [0, 1]$ then $\sum(-1)^{n+1}x^n/n$ satisfies the hypotheses of the Alternating Series Test. Hence $k \geq j \geq l$ and $x \in [0, 1]$ imply
\[ \left| \sum_{j}^{k} (-1)^{n+1}x^n/n \right| \leq \frac{x^j}{j} \leq \frac{1}{l} < \epsilon, \]
so the series satisfies the Uniform Cauchy Criterion.

This also gives an example of a uniformly convergent series which does not satisfy the hypotheses of the Weierstrass $M$-Test (and this is most easily seen by noting that the series converges conditionally at $x = 1$). However, a more elementary example of this phenomenon is given by any series of constant functions whose constant values comprise the terms of a conditionally convergent series.

Example 13.11. The Taylor series of $e^x$ at 0 is $\sum_{0}^{\infty} x^n$. The radius of convergence is $\infty$, but to see that this series converges to $e^x$ requires analyzing the remainder term in Taylor's Theorem: for all $k \in \mathbb{N}$ there exists $c$ between 0
and $x$ such that

$$
|e^x - \sum_{n=0}^{k-1} \frac{x^n}{n!} - \frac{e^x|x|^k}{k!}| \leq \frac{e^x|x|^k}{k!} \quad k \to \infty \quad 0.
$$

**Example 13.12 (Trigonometry).** For all $c_1, c_2 \in \mathbb{R}$ the power series

$$
c_1 + c_2 x - \frac{c_1}{2!} x^2 - \frac{c_2}{3!} x^3 + \frac{c_1}{4!} x^4 + \cdots
$$

has radius of convergence $\infty$, since the $n$th term has absolute value at most $(|c_1| + |c_2|)|x|^n/n!$ and $\sum_{n=0}^{\infty} (|c_1| + |c_2|)|x|^n/n!$ converges (to $(|c_1| + |c_2|)e^x$) for all $x$.

Letting $c_1 = 1$ and $c_2 = 0$, we define $\cos : \mathbb{R} \to \mathbb{R}$ by

$$
\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!},
$$

while letting $c_1 = 0$ and $c_2 = 1$, we define $\sin : \mathbb{R} \to \mathbb{R}$ by

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

We will give a brief outline showing that much (all?) of the theory of trigonometry can be easily derived by analyzing the Taylor series of $\cos$ and $\sin$.

Differentiating the series for $\sin$ and $\cos$ gives

$$
\sin' = \cos \quad \text{and} \quad \cos' = -\sin,
$$

so both $\sin$ and $\cos$ satisfy the differential equation $f'' = -f$, hence so does any linear combination $c_1 \cos + c_2 \sin$.

Conversely, suppose we are given a real-valued function on an open interval $I$ containing 0, and assume $f'' = -f$ on $I$. Then by induction $f^{(n+2)} = -f^{(n)}$ for $n = 0, 1, \ldots$. Put $c_1 = f(0)$ and $c_2 = f'(0)$. Then the Taylor series of $f$ at 0 is

$$
c_1 + c_2 x - \frac{c_1}{2!} x^2 - \frac{c_2}{3!} x^3 + \frac{c_1}{4!} x^4 + \cdots.
$$

Fix $x \in I$, and choose an upper bound $M$ for $|f|$ and $|f'|$ on the closed interval with endpoints 0 and $x$. By Taylor’s Theorem, for all $n \in \mathbb{N}$ there exists $c$ between 0 and $x$ such that

$$
|f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(0)x^k}{k!}| = \left| f^{(n)}(c)x^n \right| \leq \frac{M|x|^n}{n!},
$$

since $|f^{(n)}|$ is $|f|$ or $|f'|$ each time. Thus the Taylor series converges to $f$ on $I$, Therefore we have shown that there exist unique $c_1, c_2 \in \mathbb{R}$ such that $f = c_1 \cos + c_2 \sin$.

Because the Taylor series of $\sin$ has only odd powers,

$$
\sin(-x) = -\sin x,
$$

and $\cos(-x) = \cos x$. Thus

$$
\sin x = \sin \left( -x \right), \quad \cos x = \cos \left( -x \right).
$$

Excluding the trivial case, $x = 0$, we use these relations to show that $\sin$ and $\cos$ are in some sense independent of each other.
and similarly (because the series for \( \cos \) has only even powers) 
\[
\cos(-x) = \cos x.
\]

Letting \( x = 0 \) in the Taylor series, we get 
\[
\sin 0 = 0 \quad \text{and} \quad \cos 0 = 1.
\]

We have 
\[
\frac{d}{dx}(\sin^2 x + \cos^2 x) = 2 \sin x \cos x - 2 \cos x \sin x = 0,
\]
so \( \sin^2 + \cos^2 \) is constant. Since \( \sin^2 0 + \cos^2 0 = 1 \), we get the identity 
\[
\sin^2 x + \cos^2 x = 1.
\]

For fixed \( y \in \mathbb{R} \), we have 
\[
\frac{d^2}{dx^2} \sin(x + y) = \frac{d}{dx} \cos(x + y) = -\sin(x + y),
\]
so there exist unique \( c_1, c_2 \in \mathbb{R} \) such that \( \sin(x + y) = c_1 \cos x + c_2 \sin x \).

Differentiating, we get \( \cos(x + y) = -c_1 \sin x + c_2 \cos x \). Letting \( x = 0 \), we find \( c_1 = \sin y \) and \( c_2 = \cos y \), so we get the addition formulas 
\[
\sin(x + y) = \sin x \cos y + \cos x \sin y
\]
\[
\cos(x + y) = \cos x \cos y - \sin x \sin y.
\]

If \( 0 < x \leq 2 \) and \( n \in \mathbb{N} \) then 
\[
0 < x^2 < 6 = 3 \cdot 2 \leq (n + 2)(n + 1),
\]
so 
\[
\frac{x^n}{n!} > \frac{x^{n+2}}{(n + 2)!} > 0,
\]
hence the Taylor series for \( \sin \) satisfies the (strict version of the) hypotheses of the Alternating Series Test. Thus 
\[
\sin x > 0 \quad \text{for all} \quad 0 < x \leq 2,
\]
so \( \cos \) is strictly decreasing on \([0, 2]\]. Now, 
\[
\cos 2 = \sum_{0}^{\infty} \frac{(-1)^n 2^n}{(2n)!} = -1 + \sum_{2}^{\infty} \frac{(-1)^n 2^n}{(2n)!},
\]
and by the Alternating Series Test 
\[
\sum_{2}^{\infty} \frac{(-1)^n 2^n}{(2n)!} < \frac{2^4}{4!} = \frac{2}{3},
\]
so \( \cos 2 < 0 \). Hence there exists a unique \( c \in (0, 2) \) such that \( \cos c = 0 \). Define \( \pi = 2c \). Then 
\[
\cos \frac{\pi}{2} = 0.
\]
Since \( 1 = \sin^2 \pi/2 + \cos^2 \pi/2 = \sin^2 \pi/2 \) and \( \sin \pi/2 > 0 \), we have 
\[
\sin \frac{\pi}{2} = 1.
Thus
\[ \sin \left( x + \frac{\pi}{2} \right) = \cos x \quad \text{and} \quad \cos \left( x + \frac{\pi}{2} \right) = -\sin x. \]

Therefore, \( \sin \) and \( \cos \) are both periodic with period \( 2\pi \), and we get the usual graphs.
INDEX

absolute convergence, 66, 72
alternate triangle inequality, 13
Alternating Series Test, 66
Approximation of Suprema, 17
Archimedean Principle, 17

Bernoulli’s Inequality, 16
between, 13
binomial coefficients, 15
Binomial Theorem, 15
Bolzano-Weierstrass Theorem, 33
bounded, 25
bounded above, 14
bounded below, 14

Cauchy
uniform, 70, 72
Cauchy sequence, 33
Cauchy’s Mean Value Theorem, 45
Cauchy-Hadamard Theorem, 73
Chain Rule, 44
Change of Variables Theorem, 57
closed form, 63
cluster point, 26
Comparison Test, 65
Comparison Theorem for Improper Integrals, 39
Completeness Axiom of R, 17
conditional convergence, 66
Continuity Characterization of Limits, 39
continuous, 38
at a point, 38
limit characterization of, 39
sequential characterization of, 38
uniformly, 40
converge
absolutely, 66, 72
conditionally, 66
pointwise, 69, 72
sequence, 24
series, 63
uniformly, 69, 72
converge pointwise
series, 72
cos, 77
countability of Q, 22
countable, 21
countably infinite, 21
Critical Point Lemma, 44
Darboux’s Theorem, 51
decimal expansions, 18
decreasing, 36
strictly, 37
Density of Irrationals, 20
Density of Rationals, 18
derivative, 42
higher order, 45
differentiable, 42
at a point, 42
Dirichlet function, 52
discontinuity, 38
discontinuous, 38
diverge
sequence, 25
to \( \infty \), 30
e, 62
element, 1
exponential function, 61
extended real numbers, 29
Extreme Value Theorem, 40
family, 7
Field Axioms, 11
finite, 21
floor, 18
function, 3
1-1, 6
arithmetic with, 6
bounded, 25
composition, 6
constant, 6
domain, 4
extension, 5
identity, 6
image, 4
inverse, 6
one-to-one, 6
onto, 6
pre-image, 4
range, 4
real-valued, 4
restriction, 5
Fundamental Theorem of Calculus, 55
improper integral, 57
increasing, 36
strictly, 37
indexed family, 7
inf, 16
infimum, 16
MAT 371 LECTURES  81

infinite, 21
integers, 12
integral, 49
Integral Test, 64
Integration by Parts, 56
Intermediate Value Theorem, 40
interval, 14
  bounded, 14
degenerate, 14
half-closed, 14
half-open, 14
nondegenerate, 14
open, 14
unbounded, 14
Inverse Function Theorem, 48
irrational numbers, 12

L'Hôpital's Rule, 47
lim inf, 31
lim sup, 31
limit
  at infinity, 37
  continuity characterization of, 39
  function, 34
  infinite, 30, 37
  left-hand, 36
  one-sided, 36
  restricted, 36
  right-hand, 36
  sequence, 25
  sequential characterization of, 34
Limit Characterization of Continuity, 39
Limit Comparison Test, 66
logarithm function, 60
lower bound, 14
lower sum, 48, 49

Mathematical Induction, 11
maximum, 14
function, 40
Mean Value Theorem, 45
Cauchy, 45
Mean Value Theorem for Integrals, 55
minimum, 14
function, 40
monotone, 37
strict, 37

N, 11
natural numbers, 11

Order Axioms, 12
ordered pair, 3

pairwise disjoint, 8
partial sum, 63
partition (for integrals), 48
Fermat's Triangle, 15
pointwise convergence, 69, 72
Power Rule, 43
power series, 73
  center, 73
  coefficients, 73
  powers
    arbitrary, 61
    integer, 12
    rational, 20
Product Rule, 42
Q, 12
Quotient Rule, 42

R, 3
radius of convergence, 74
Ratio Test, 68
rational numbers, 12
Rearrangement Theorem, 67
Riemann, 68
refines, 48
Riemann integrable, 49
Riemann integral, 49
Riemann sum, 49
Riemann's Criterion, 51
Riemann's Rearrangement Theorem, 68
Rolle's Theorem, 44
Root Test, 69
roots, 20

sequence, 23
  bounded, 25
  Cauchy, 33
  converge, 24
  diverge, 25
term, 23
Sequential Characterization of Continuity, 38
Sequential Characterization of Limits, 34
series, 63, 72
  alternating harmonic, 67
  converge, 63
  converge pointwise, 72
  geometric, 63
  harmonic, 64
  p-, 65
  partial sum, 63
  power, 73
  sum, 63, 72
  Taylor, 75
term, 63
set, 1
  bounded, 25
Cartesian product, 3, 8
complement, 3
countable, 21
countably infinite, 21
De Morgan's Laws, 9
difference, 3
disjoint, 3
Distributive Laws, 8
empty, 2
equal, 2
family, 7
finite, 21
indexed family, 7
infinite, 21
intersection, 3, 7
proper subset, 2
subset, 2
superset, 2
uncountable, 21
union, 3, 7

\sin, 77

Squeeze Theorem
  limits, 35
  sequences, 25
subsequence, 26
sup, 16
supremum, 16

Taylor series, 75
Taylor's Theorem, 46
telegraphing sum, 64
term
  sequence, 23
  series, 63
triangle inequality, 13
  alternate, 13

uncountability of \( \mathbb{R} \), 23
uncountable, 21
uniform Cauchy criterion, 70, 72
uniform convergence, 69, 72
uniformly continuous, 40
upper bound, 14
upper sum, 48

Weierstrass M-Test, 73
Well-Ordering Principle, 14

\( \mathbb{Z} \), 12