1. **Introduction**

These lecture notes will develop “intermediate” mathematical analysis. What does this mean? Well, first of all, the material will contain all of what’s typically covered in “one-variable advanced calculus”, which is the rigorous treatment of calculus on the real line. However, we’ll also do a lot of our analysis in the world of metric spaces, where “distance” has been abstracted to its essence. It’s just as easy to develop limits and continuity in the generality of metric spaces as in the real line or \( n \)-dimensional Euclidean space. However, for derivatives, integrals, and series we’ll go back to the real line. At the other end of the spectrum, “real analysis” often includes “measure theory”, or at least the Lebesgue integral; we won’t do any of that here — it’ll be the Riemann integral for us.

The style will be conversational, rather than formal. These lecture notes are written for you (the student), rather than for the instructor or other professional mathematicians. What does this mean? I emphasize the word *style* here — the mathematics will be correct, and the proofs complete. But I’ve tried to write the stuff in “plain language”, the way I would expect you to write it; there won’t be any definition or theorem I wouldn’t expect you to be able to state for me if I ask for it.

That being said, I should mention that the thing to remember about mathematics is that it’s abstract. This means that mathematics doesn’t say anything directly about the natural world, unlike physics and chemistry, for example. On the plus side, in mathematics we can prove that what we say is correct, while in fields like physics and chemistry the best that can be expected is that the theory can accurately predict natural phenomena. Of course, the language of mathematics is crucial in the formulation of scientific theories. Also, there are areas of “applied mathematics” where certain aspects of the natural world are “axiomatized” and theorems are proven about this axiomatic formulation. But again, this doesn’t mean we can prove things about nature itself, only about “idealized” nature.

These lecture notes won’t contain any applied mathematics, only “pure” mathematics. Actually, in recent times there has been a lot of discussion complaining about the “artificial” division of mathematics into “pure” and “applied”. Just think of “pure” mathematics as abstract mathematics which makes no attempt to connect with the natural world. The value of pure mathematics is that it serves as the foundation for applications, and we can at least have a high degree of confidence in the correctness of the results in pure mathematics.
All of modern mathematics can be based upon set theory. We won’t be very fussy about this, especially when we talk about the “prerequisite” material, namely sets and functions. Set theory is certainly not our main objective — we’re interested in the foundations of analysis — so we won’t feel the need to give anything remotely like the complete “axiomatic” development of set theory (to see why it’s a good idea for us to eschew this, you really ought to spend some time looking through a book on axiomatic set theory). Rather, we’ll only briefly recall the basics of sets and functions. We’ll be even a little lazier than usual here, allowing ourselves to “define” sets, and to give informal “definitions” of functions and ordered pairs. The only part of this which might be new for you is “families” of sets, so pay attention to that.

Then we’ll start talking about real numbers. Let’s agree that we all know about arithmetic and inequalities. The discussion will start to get serious with the Completeness Axiom, which is really what makes calculus possible. We’ll introduce the real numbers through their axioms. Actually, it’s possible to construct the real numbers using set theory, and prove the properties we’ll state in our axioms, but this would take a lot of time (and again, to see what is involved, you really ought to read a math book that contains a careful construction of the real numbers — for example, “Baby Rudin” (Rudin’s Principles of Mathematical Analysis)), and is not really our main concern, which is the analysis of functions.

We’ll need to know the difference between countable and uncountable sets, because, for example, the natural numbers are countable while the real numbers are uncountable. There’s a lot that could be said about this, but we’ll keep it simple and only say what we’ll need to use later. We’ll be on the precipice of the abyss of “cardinal numbers”, but we’ll hold onto the guard rail and not fall in.

Analysis is really the study of limits. We lay the groundwork for our study by looking pretty hard at limits of sequences, which will be the foundation of all our later theory. A lot of our work with sequences will be in the context of metric spaces, although we’ll have some special things to say about real sequences — that is, sequences of real numbers.

The theory of convergence of sequences leads naturally into the general theory of limits and continuity, which again we’ll study mostly in abstract metric spaces, but with some (very) special things for functions whose domain or range (or both) are in real line.
I assume you’ve had courses in calculus and linear algebra. I also assume you’ve had some experience with sets and functions, and doing proofs. In examples and exercises, the properties of the familiar functions from your calculus course will be used willy-nilly; most of the properties will be rigorously justified by the time we’re done with the course, and the formal development of the theory won’t depend upon your prior knowledge of calculus.

Advice for the student: These lecture notes are meant to be read carefully! You’ll have to read each bit several times, with pen and paper to write down what your brain is telling (asking?) you. There are lots of exercises. A small number will be formally assigned as homework to be turned in, but all of them are important; my advice is to try them all, but at least you should be familiar with what they all say. At any point in the lecture notes, anything that’s said in any exercise which comes before can be used. Also, for the solution of any exercise, anything that appears in any other exercise which comes before can be used.

There’s a fairly complete index — use it! Also, get a lot of practice writing down from memory the statements of definitions, results, and examples. And you should know the results by their names.

1although the calculus is certainly the more important for these lecture notes
2. Sets

The “formal” development of modern mathematics can be based completely on set theory. More precisely, any mathematical definition is “really” about certain kinds of sets, and all results and proofs can be reduced to arguments involving only sets.

But let me emphasize that we will not do it this way here. In this section we’ll just recall a few of the most important facts concerning sets. Virtually everything in this section, except perhaps for the general concept of families of sets, is prerequisite material for the course.

Main definitions

A **set** is a collection of objects, called the **elements** of the set.

If $A$ and $B$ are sets, a **function** from $A$ to $B$ is a rule which associates to each element $x$ of $A$ a unique element of $B$, called the **value** of the function at $x$.

A **family of sets** is a set whose elements are sets. An **indexed family of sets** is a function whose values are sets.

Notation and terminology

Sets:

$x \in A$ means the object $x$ is an element of the set $A$, and we also say $x$ is in $A$.

$x \notin A$ means $x$ is not in $A$.

Functions:

$f: A \to B$ means $f$ is a function from $A$ to $B$, and the value of $f$ at $x \in A$ is denoted $f(x)$.

The set $A$ is called the **domain** of $f$, written $A = \text{dom } f$.

The **range** of $f$ is

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

A professional mathematician would find our definition of sets unacceptable, saying “What is a collection?” and “What is an object?” in a formal development, “set” and “element” would be undefined terms.

Again, professional mathematicians would reject this definition of function, asking “What is a rule?” But it suits our purposes.

In this definition, we use the notation “:=” to indicate that the equals sign is being used not to assert that two things we already know about happen to be equal, rather := means that the right-hand side is the definition of the notation on the left. It’s really sort of a matter of taste whether to use “:=”; my taste leans toward using it sometimes, but I try not to go overboard with it.
Occasionally it’s convenient to introduce a function but not give it a name, in which case we use the notation

\[ x \mapsto (\text{some formula involving } x). \]

**Families 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 \( \{A_i\}_{i \in I} \), and \( I \) is called the *index set* of the family.

For special index sets we have special notation:

\[
\begin{align*}
\{A_i\}_{i=1}^\infty &= (A_1, A_2, \ldots) & \text{when } I = 1, 2, \ldots \\
\{A_i\}_{i=1}^n &= (A_1, \ldots, A_n) & \text{when } I = 1, 2, \ldots, n
\end{align*}
\]

We use parentheses on the right-hand side\(^5\) to emphasize that an indexed family is a function rather than just a set of sets. There is an important notational aspect to consider: it’s common to use “index” notation such as \( \{A_i \mid i \in I\} \) 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 \( F \), to give a clue that something out of the ordinary is afoot.

**Discussion of above definitions**

**Sets:**

All a set knows about is what its elements are. If \( x \) is an object and \( A \) is a set, the sentence “\( x \in A \)” is a (mathematical) proposition — that is, it’s either true or false. Roughly speaking, all a set is good for is to allow the formation of such propositions.\(^6\)

Many definitions and results in set theory are merely translations of concepts from logic; for example, the logical *biconditional* (“if and only if”) becomes set equality:

**Observation.**\(^7\) *If \( A \) and \( B \) are sets, then \( A = B \) if and only if if for any object \( x, x \in A \) if and only if \( x \in B \).*

\(^5\) and I’d really like to use them on the left also, but I’ll use the more established notation.

\(^6\) 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.

\(^7\) When we present something as an “observation”, we mean to imply that the assertion is so easy to prove that it’s completely routine, and so we feel justified in omitting the proof. On the other hand, we also mean to imply that the assertion could be useful, and so we are officially recording it so we have it available later.
Again: all a set cares about is what it’s elements are.

**Functions:**
All a function knows about is what it’s values are — the particular form of the rule used to determine the value \( f(x) \) is irrelevant:

**Observation.** Let \( f \) and \( g \) be functions. Then \( f = g \) if and only if both of the following are satisfied:

(i) \( \text{dom } f = \text{dom } g \), and
(ii) \( f(x) = g(x) \) for all \( x \in \text{dom } f \).

If \( f: A \to B \), then 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 \( \text{ran } f \subset C \), then we can equally well regard \( f: A \to C \); it is still the *same* function.

In examples and exercises we’ll freely use functions you saw in your calculus course, although their formal properties won’t necessarily be established until later. Here they are:

**Polynomial:** Of the form \( c_0 + c_1x + c_2x^2 + \cdots + c_nx^n \).

**Rational:** Of the form \( f/g \), where \( f \) and \( g \) are polynomials. By default, the domain is all real numbers where the denominator \( g \) is nonzero.

**Algebraic:** Has a formula involving only arithmetic operations, including \( n \)th roots for \( n = 2, 3, 4, \ldots \). By default, the domain is all real numbers where the formula makes sense.

**Logarithm:** The “natural” \( \log \) with base \( e \).

**Exponential:** \( e^x \).

**Trigonometric:** \( \sin, \cos, \text{ and } \tan \) — we won’t have any use for the others.

**Inverse trig:** \( \tan^{-1} \) is the only one we’ll use, and our convention is that it’s range will be \((-\pi/2, \pi/2)\).

**Families of sets:**
Families of sets come in two flavors: unindexed and indexed. These really are distinct things, but we’ll blur this distinction, because in

---

8Actually, we have made a choice here — in some areas of advanced mathematics it’s important to regard the set \( B \) as an official part of the function (the “codomain”).

9Actually, this is only an informal definition, which doesn’t include all the “official” algebraic functions, but it’s good enough for our purposes.

10We won’t have any use for the “common” log base 10, or for the silly notation \( \ln \).

11which are only trivial combinations of the 3 we’ll use, anyway
practice there won’t be much danger of confusion. Basically, we use whatever seems more convenient in context. In most cases it’s a matter of taste; since “family of sets” is simpler linguistically, conceptually, and notationally than “indexed family of sets”, this is what I prefer, but many people prefer indexed families. In a few cases — for example, Cartesian product (see below) — indexed families are necessary. In some cases it would seem artificial to use an indexing for a family — for example, how would you index the family of all subsets of a given set?

To every indexed family \( \{A_i\}_{i \in I} \) of sets, there is an associated family of sets, namely the range \( \{A_i \mid i \in I\} \) (and please note the subtle change of notation!) of the indexed family. The awkward similarity of these two notations is one of the reasons why the distinction between families and unindexed families is blurred. Conversely, to any family \( \mathcal{F} \) of sets we could associate an indexed family using the identity function on \( \mathcal{F} \) — this is awkward, and not usually done.

There is no logical need for the term “family of sets” (at least in the unindexed case); nevertheless it’s handy because it would sometimes be confusing to say “set of sets”.

**Theory associated to above definitions**

**Sets:**

**Definition.** The empty set is the unique set \( \emptyset \) such that for all \( x \), \( x \not\in \emptyset \).

There is only one empty set: suppose \( A \) and \( B \) are sets which are empty, i.e., have no elements. Then the sentence “for all \( x \), \( x \in A \) if and only if \( x \in B \)” is true for a very stupid reason: no matter what \( x \) is, both of the propositions \( x \in A \) and \( x \in B \) are guaranteed to be false. This is an example of something that is what we call “vacuously true”. Anyway, since the empty set is unique it makes sense to give it a special symbol — “\( \emptyset \)” in this case.

The empty set can be quite confusing. We need it because manipulations with sets can easily produce the empty set. In terms of logic, the empty set serves the following purpose: it gives us a set \( A \) for which the proposition “\( x \in A \)” is false for every \( x \). It’s almost paradoxical that the empty set is very trivial, yet often leads to blunders if it’s not handled carefully. For example, frequently it’s all too easy to overlook the possibility of a set being empty, which can cause a gap in a proof. Also, sometimes the special case of the empty set can be trickier to handle than the other cases.
There is no set $A$ such that $x \in A$ is true for every $x$; 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 — if you’re curious, look it up or ask someone — we won’t need to concern ourselves with it.

The set-theoretic translation of the logical conditional (“implies”) is the subset relation:

**Definition.** $A$ is a **subset** of $B$ if for all $x$, $x \in A$ implies $x \in B$.

**Notation and Terminology.** $A \subset B$ means $A$ is a subset of $B$, and we also say $B$ is a **superset** of $A$, written $B \supset A$.

A subset $A$ of $B$ is called **proper** if $A \neq B$.

In the above, we used a phrase of the form “we also say . . .”. It’s 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 “$B \supset A$” is the same as that of “$A \subset B$”, namely, “$B \supset A$ if for all $x$, $x \in A$ implies $x \in B$”.

**Example.** Let’s prove $A \subset A$. This is so trivially true that when we write the proof it’s going to seem like we didn’t really do anything. Here it is: the conditional “$x \in A = \Rightarrow x \in A$” is trivially true, therefore $A \subset A$.

**Example.** Let’s prove $\emptyset \subset A$. The conditional “$x \in \emptyset = \Rightarrow x \in A$” is trivially true, because the hypothesis $x \in \emptyset$ is false. Therefore $\emptyset \subset A$.

**Lemma 2.1.** Let $A$ and $B$ be sets. Then $A = B$ if and only if both $A \subset B$ and $B \subset A$.

**Proof.** The biconditional “$x \in A \iff x \in B$” is equivalent to the following conjunction of conditionals: “$x \in A \Rightarrow x \in B$ and $x \in B \Rightarrow x \in A$”. Therefore $A = B$ if and only if $A \subset B$ and $B \subset A$. $\square$

---

12 It’s a *tautology.*

13 We called this result a “lemma” — the alternatives would be “theorem”, “proposition”, or “corollary”. It’s certainly not a corollary, because it doesn’t follow easily from any result we had previously proven. It’s not usually considered good form to call such an elementary result a theorem; that’s usually reserved for main, or at least significant, results. It’s sort of a matter of taste whether to call it a lemma or a proposition; both words indicate a result of lesser importance — in fact, the word lemma sometimes connotes something truly minor (although there are counterexamples to this, for example Urysohn’s Lemma — you can find out what that is by reading a topology book). But the real difference is that the word lemma is used for a result whose main purpose is to be used to prove other things. This is certainly true for this result: it gives the method which is most often used to prove two sets are equal.
Here is the set-theoretic version of \textit{negation} (“not”): 

\textbf{Definition.} If the set $U$ is understood, for any subset $A \subset U$ the \textit{complement} of $A$ is 

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

The complement of $A$ is not just $\{x \mid x \not\in A\}$, rather it’s $\{x \in U \mid x \not\in A\}$. 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.

\textbf{Functions:}

\textbf{Definition.} Let $f: A \to B$.

(i) If $C \subset A$, the \textit{image} of $C$ \textit{under} $f$ is 

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

(ii) If $D \subset B$, the \textit{pre-image} of $D$ \textit{under} $f$ is 

$$f^{-1}(D) := \{x \in A \mid f(x) \in D\}.$$ 

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. To help keep things straight, we usually write elements of the domain as lower-case letters like $x, y, \ldots$, but subsets of the domain as upper-case letters like $C, D, \ldots$. 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).

\textbf{Definition.} Let $f: A \to B$ and $C \subset A$.

(i) The \textit{restriction} of $f$ to $C$ is the function $f|_C: C \to B$ defined by 

$$f|_C(x) = f(x) \quad \text{for} \ x \in C.$$ 

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

A function has \textit{only one} restriction to any given subset of its domain, but in general has \textit{many} extensions to a given superset of its domain.

\textbf{Example.} Let $A = B = \{-1, 0, 1\}$ and $C = \{0\}$. Define functions $f, g: A \to B$ by 

$$f(x) = x \quad \text{and} \quad g(x) = -x \quad \text{for} \ x \in A,$$

and define $h: C \to B$ by 

$$h(0) = 0.$$
Then $h$ is the restriction of both $f$ and $g$ to $C$, and both $f$ and $g$ are extensions of $h$ to $A$.

There’s nothing useful to say about forward images of complements, but inverse images behave nicely:

**Corollary 2.2.** Let $f: A \to B$ and $C \subset B$. Then

$$f^{-1}(C^c) = f^{-1}(C)^c.$$ 

In the above statement, from the context it’s clear that the complements are taken with respect to $B$ and $A$, respectively.

**Proof.** We have:

$$x \in f^{-1}(C^c) \iff f(x) \in C^c$$
$$\iff f(x) \notin C$$
$$\iff x \notin f^{-1}(C)$$
$$\iff x \in f^{-1}(C)^c \quad \square$$

Here’s what happens when you “take images” both ways:

**Proposition 2.3.** Let $f: A \to B$, and let $C \subset A$ and $D \subset B$. Then:

(i) $f^{-1}(f(C)) \supset C$

(ii) $f(f^{-1}(D)) \subset D$

**Proof.**

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

(ii). Let $y \in f(f^{-1}(D))$. We must show $y \in D$. We can choose $x \in f^{-1}(D)$ such that $f(x) = y$. Then $f(x) \in D$, so $y \in D$. \quad \square

**Exercise 2.4.** Referring to the preceding proposition, in each part find an example where the containment or inclusion is proper. That is:

(a) Find an example where $f^{-1}(f(C)) \neq C$;

(b) Find an example where $f(f^{-1}(D)) \neq D$.

Functions respect the subset relation:

**Proposition 2.5.** Let $f: A \to B$, and let $C, D \subset A$ and $E, F \subset B$. Then:

(i) If $C \subset D$ then $f(C) \subset f(D)$;

(ii) If $E \subset F$ then $f^{-1}(E) \subset f^{-1}(F)$.

**Proof.**

(i). Let $C \subset D$. Then $x \in C$ implies $x \in D$, so $f(C) \subset f(D)$ by definition of image.

(ii). Let $E \subset F$. Then $f(x) \in E$ implies $f(x) \in F$, hence $x \in f^{-1}(E)$ implies $x \in f^{-1}(F)$, therefore $f^{-1}(E) \subset f^{-1}(F)$.
Definition. If $f: A \to B$ and $g: B \to C$, the \textit{composition} \footnote{The Tragedy of Mathematics is that function composition is written backwards, as in the above definition. We read from left to right, and even draw pictures of functions going from left to right, but we write compositions of functions from right to left. Unfortunately, we can’t change this now; some mathematicians have tried to correct this, but it’s too ingrained. This is why there is no universal convention regarding whether to call $g \circ f$ the composition of $f$ and $g$ or of $g$ and $f$ — there are two kinds of mathematicians in the world: those who say it’s the first, and those who say it’s the second.} of $f$ and $g$ (or of $g$ and $f$) is the function $g \circ f: A \to C$ defined by

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

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

Definition. (i) The \textit{identity function} on a set $A$ is defined by

$$\text{id}_A(x) := x \quad \text{for } x \in A.$$  

(ii) If $A$ and $B$ are sets and $c \in B$, the \textit{constant function} from $A$ to $B$ with value $c$ is the function $f: A \to B$ defined by

$$f(x) = c \quad \text{for } x \in A.$$  

We also say $f$ is \textit{identically} $c$.

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

(i) $f$ is \textit{one-to-one}, written 1-1, if

$$f(x) = f(y) \implies x = y \quad \text{for all } x, y \in A.$$  

(ii) $f$ is \textit{onto $B$} if $\text{ran } f = B$.

(iii) If $f$ is 1-1 and onto $B$, the \textit{inverse of $f$} is the unique function $f^{-1}: B \to A$ such that $f^{-1} \circ f = \text{id}_A$ and $f \circ f^{-1} = \text{id}_B$.

Thus, if $f: A \to B$ and $g: B \to A$ we have $g = f^{-1}$ if and only if both $g(f(x)) = x$ for all $x \in A$ and $f(g(y)) = y$ for all $y \in B$.

Exercise 2.6. Find an example of functions $f: A \to B$ and $g: B \to A$ such that $f \circ g = \text{id}_B$ but $g \circ f \neq \text{id}_A$.

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”.

Exercise 2.7. Let $f: A \to B$. Prove:
(a) \( f \) is 1-1 if and only if 
\[
f^{-1}(f(C)) = C \quad \text{for all } C \subset A;
\]
(b) \( f \) is onto if and only if 
\[
f(f^{-1}(D)) = D \quad \text{for all } D \subset B.
\]

The notation \( f^{-1} \) must be used with care: for any function \( f : A \rightarrow B \) it makes sense to consider \( f^{-1}(C) \) for any \( C \subset B \). 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 \( f^{-1} : B \rightarrow A \) does exist and \( C \subset B \), 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 \( f^{-1} \).

**Families of sets:**

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

**Definition.** Let \( \mathcal{F} \) be a family of sets.

(i) The **union** of \( \mathcal{F} \) is 
\[
\bigcup \mathcal{F} := \{ x \mid \text{there exists } A \in \mathcal{F} \text{ such that } x \in A \}.
\]

(ii) The **intersection** of \( \mathcal{F} \) is 
\[
\bigcap \mathcal{F} := \{ x \mid \text{for all } A \in \mathcal{F}, x \in A \}.
\]

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

(i) The **union** of \( \{A_i\}_{i \in I} \) is 
\[
\bigcup_{i \in I} A_i := \{ x \mid \text{there exists } i \in I \text{ such that } x \in A_i \}.
\]

(ii) The **intersection** of \( \{A_i\}_{i \in I} \) is 
\[
\bigcap_{i \in I} A_i := \{ x \mid \text{for all } i \in I, x \in A_i \}.
\]

(iii) The **Cartesian product** of \( \{A_i\}_{i \in I} \) is the set \( \prod_{i \in I} A_i \) of functions 
\[
i \mapsto x_i \text{ with domain } I \text{ such that } x_i \in A_i \text{ for all } i \in I.
\]

We do **not** try to form a Cartesian product of an unindexed family of sets!

**Notation and Terminology.** For special index sets \( I \) we have special notation:
(i) $I = \{1, 2, \ldots \}$:

\[
\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots \\
\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \cdots \\
(x_1, x_2, \ldots) \in \prod_{i=1}^{\infty} A_i = A_1 \times A_2 \times \cdots
\]

(ii) $I = \{1, 2, \ldots, n\}$:

\[
\bigcup_{i=1}^{n} A_i = A_1 \cup \cdots \cup A_n \\
\bigcap_{i=1}^{n} A_i = A_1 \cap \cdots \cap A_n \\
(x_1, \ldots, x_n) \in \prod_{i=1}^{n} A_i = A_1 \times \cdots \times A_n \\
A^n = A \times \cdots \times A
\]

Sometimes the notation involved in a particular situation makes it clearer to write $\bigcup\{A_i \mid i \in I\}$ rather than $\bigcup_{i \in I} A_i$, and similarly for intersections and Cartesian products.

It’s occasionally convenient to rewrite a union:

**Exercise 2.8.** Let $\{A_i\}_{i \in I}$ be an indexed family of sets, and let $j \mapsto i_j : J \to I$ be an onto function. Prove that

\[
\bigcup_{i \in I} A_i = \bigcup_{j \in J} A_{i_j}.
\]

Also, it’s occasionally convenient to write $\bigcup_{A \in \mathcal{F}} A$ for $\bigcup \mathcal{F}$, and similarly for intersections.

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 $\{A_i \mid i \in I\}$; it’s easy to see that the union is the same for both. Similarly for intersection.

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 a function.

For two sets we can write:
union  \( A \cup B = \{ x \mid x \in A \text{ or } x \in B \} \)

intersection  \( A \cap B = \{ x \mid x \in A \text{ and } x \in B \} \)

product  \( A \times B = \{ (x, y) \mid x \in A \text{ and } y \in B \} \)

Note that for two sets, union becomes the set-theoretic version of the logical \textit{disjunction} (“or”), and intersection the set-theoretic version of \textit{conjunction} (“and”).

**Exercise 2.9.** Prove that \( A \times (B \cup C) = (A \times B) \cup (A \times C) \).

**Definition.** \( A \) is \textit{disjoint} from \( B \) if \( A \cap B = \emptyset \), otherwise \( A \) \textit{intersects} \( B \).

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

How should we define what it means for an indexed family \( \{A_i\}_{i \in I} \) of sets to be \textit{pairwise disjoint}? It seems simplest to let it mean the same as for the associated family \( \{A_i \mid i \in I\} \). 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 \( A_i \cap A_j = \emptyset \). This is \textit{not} the same as “either \( i = j \) or \( A_i \cap A_j = \emptyset \).” Thus, according to this definition, a pairwise disjoint indexed family of sets might not be 1-1: we can have \( i \neq j \) and \( A_i = A_j \), even for a pairwise disjoint family \( \{A_i\}_{i \in I} \). However, frequently a pairwise disjoint indexed family of sets will be 1-1, and in such a case it’s a little easier to verify pairwise disjointness: assume \( i \neq j \) and deduce \( A_i \cap A_j = \emptyset \).

**Definition.** The \textit{difference} of \( A \) and \( B \) is

\[
A \setminus B := A \cap B^c
\]

In the above definition, what “universe” was used? It doesn’t matter — no matter what universe we choose (as long as it contains all the sets in question), the difference (!) is the same.

Now that we have Cartesian products, we can consider:

**Definition.** An \textit{ordered pair} is an element of a Cartesian product of two sets.

If \( (x, y) \) is an ordered pair, the \textit{first coordinate} is \( x \), and the \textit{second coordinate} is \( y \).

**Observation.** \( (x, y) = (z, w) \) if and only if \( x = z \) and \( y = w \).

\[15\] We made a choice here — it’s really just a matter of taste, and there are two kinds of mathematicians in the world...
Definition. The graph of a function \( f : A \to B \) is

\[ \{(x, f(x)) \mid x \in A\}. \]

The graph of \( f \) is a subset of \( A \times B \). It wouldn’t do any harm to identify \( f \) with its graph, because the graph gives us a rule: to \( x \in A \) associate the value \( y \), where \( y \) is the second coordinate of the unique ordered pair in the graph with first coordinate \( x \).

Exercise 2.10. For each \( n = 1, 2, \ldots \) define

\[ A_n = \{1, \ldots, n\}. \]

Find \( \bigcup_{n=1}^{\infty} A_n \) and \( \bigcap_{n=1}^{\infty} A_n \), and prove you are right.

Exercise 2.11. Let \( A \) be a set, and let \( \mathcal{F} \) be the family of finite subsets of \( A \). Prove that \( A = \bigcup_{F \in \mathcal{F}} F \). (See Lecture 4 for a precise definition of finite if necessary.)

Distributive Laws. Let \( A \) be a set, and let \( \mathcal{F} \) be a family of sets. Then:

(i)

\[ A \cap \bigcup \mathcal{F} = \bigcup_{B \in \mathcal{F}} (A \cap B); \]

(ii)

\[ A \cup \bigcap \mathcal{F} = \bigcap_{B \in \mathcal{F}} (A \cup B). \]

Proof. (i). We have

\[ x \in A \cap \bigcup \mathcal{F} \]

\[ \iff x \in A \text{ and } x \in \bigcup \mathcal{F} \]

\[ \iff x \in A \text{ and there exists } B \in \mathcal{F} \text{ such that } x \in B \]

\[ \iff \text{there exists } B \in \mathcal{F} \text{ such that both } x \in A \text{ and } x \in B \]

\[ \iff \text{there exists } B \in \mathcal{F} \text{ such that } x \in A \cap B \]

\[ \iff x \in \bigcup_{B \in \mathcal{F}} (A \cap B) \]
(ii). We have

\[ x \in A \cup \bigcap \mathcal{F} \]

\[ \iff \quad x \in A \text{ or } x \in \bigcap \mathcal{F} \]

\[ \iff \quad x \in A \text{ or for all } B \in \mathcal{F}, \ x \in B \]

\[ \iff \quad \text{for all } B \in \mathcal{F}, \text{ either } x \in A \text{ or } x \in B \]

\[ \iff \quad \text{for all } B \in \mathcal{F}, \ x \in A \cup B \]

\[ \iff \quad x \in \bigcap_{B \in \mathcal{F}} (A \cup B) \quad \square \]

At both places labelled * in the above proof, we applied tautologies from logic involving quantifiers. In fact, the above Distributive Laws of set theory are just translations of the Distributive Laws of logic.

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

(i) \[ \left( \bigcup_{A \in \mathcal{F}} A \right)^c = \bigcap_{A \in \mathcal{F}} A^c \]

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

**Proof.** (i). We have

\[ x \in \left( \bigcup_{A \in \mathcal{F}} A \right)^c \]

\[ \iff \quad x \notin \bigcup_{A \in \mathcal{F}} A \]

\[ \iff \quad \text{it’s false that } x \in \bigcup_{A \in \mathcal{F}} A \]

\[ \iff \quad \text{it’s false that there exists } B \in \mathcal{F} \text{ such that } x \in B \]

\[ \iff \quad \text{for all } B \in \mathcal{F}, x \notin B \]

\[ \iff \quad \text{for all } B \in \mathcal{F}, x \in B^c \]

\[ \iff \quad x \in \bigcap_{B \in \mathcal{F}} B^c \]

(ii). From (i) we have

\[ \bigcup_{A \in \mathcal{F}} A^c = \left( \bigcup_{A \in \mathcal{F}} A \right)^{cc} = \left( \bigcap_{A \in \mathcal{F}} A^{cc} \right)^c = \left( \bigcap_{A \in \mathcal{F}} A^c \right)^c = \left( \bigcap_{A \in \mathcal{F}} A \right)^c = \left( \bigcup_{A \in \mathcal{F}} A \right)^c \quad \square \]
Again note that at the place labelled * we used a tautology from logic involving quantifiers.

Also, in (ii) we used the identity $A^{cc} = A$.

Any result about families of sets can be applied to a collection of two sets, for example:

**Example.**

(i) 

\[ A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \]

(ii) 

\[
(A \cup B) \setminus C = (A \cup B) \cap C^c
\]

\[
= (A \cap C^c) \cup (B \cap C^c)
\]

\[
= (A \setminus C) \cup (B \setminus C).
\]

(iii) 

\[
A \setminus (B \cap C) = A \cap (B \cap C)^c
\]

\[
= A \cap (B^c \cup C^c)
\]

\[
= (A \cap B^c) \cup (A \cap C^c)
\]

\[
= (A \setminus B) \cup (A \setminus C).
\]

**Proposition 2.12.** Let \( \{A_i\}_{i \in I} \) be a family of sets, and let \( k \in I \). Then

\[
\bigcap_{i \in I} A_i \subset A_k \subset \bigcup_{i \in I} A_i.
\]

**Proof.** We must prove two things: \( \bigcap_{i \in I} A_i \subset A_k \) and \( A_k \subset \bigcup_{i \in I} A_i \). The first follows from the tautology

\[
(\forall i \in I, x \in A_i) \implies x \in A_k
\]

and the second from the tautology

\[
x \in A_k \implies (\exists i \in I \text{ such that } A_i).
\]

**Observation.**

(i) \( A \subset \bigcap F \) if and only if \( A \subset B \) for all \( B \in F \);

(ii) \( \bigcup F \subset A \) if and only if \( B \subset A \) for all \( B \in F \).

**Example.** Let’s prove:

(i) \( A = (A \cap B) \) if and only if \( A \subset B \);

(ii) \( A = (A \cup B) \) if and only if \( B \subset A \).

For (i):

\[
A = (A \cap B) \iff A \subset A \cap B \quad (A \supset A \cap B \text{ always})
\]

\[
\iff A \subset A \text{ and } A \subset B
\]

\[
\iff A \subset B \quad (A \subset A \text{ always})
\]
For (ii):

\[ A = (A \cup B) \iff A \supset (A \cup B) \quad (A \subset A \cup B \text{ always}) \]
\[ \iff A \supset A \quad \text{and} \quad A \supset B \]
\[ \iff A \supset B \quad (A \supset A \text{ always}) \]

**Proposition 2.13.** 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) \[ f\left( \bigcup \mathcal{F} \right) = \bigcup_{C \in \mathcal{F}} f(C) \]

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

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

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

**Proof.**

(i). \[ y \in f\left( \bigcup \mathcal{F} \right) \]
\[ \iff \text{there exists } x \in \bigcup \mathcal{F} \text{ such that } y = f(x) \]
\[ \iff \text{there exists } C \in \mathcal{F} \text{ such that } \]
\[ \quad \text{there exists } x \in C \text{ such that } y = f(x) \]
\[ \iff \text{there exists } C \in \mathcal{F} \text{ such that } y \in f(C) \]
\[ \iff y \in \bigcup_{C \in \mathcal{F}} f(C) \]

(ii). We have

\[ x \in \bigcap \mathcal{F} \]
\[ \implies \text{for all } C \in \mathcal{F}, \ x \in C \]
\[ \implies \text{for all } C \in \mathcal{F}, \ f(x) \in f(C) \]
\[ \implies f(x) \in \bigcap_{C \in \mathcal{F}} f(C), \]

so \( f(\bigcap \mathcal{F}) \subset \bigcap_{C \in \mathcal{F}} f(C) \).
(iii). We have

\[ x \in f^{-1}\left( \bigcup \mathcal{G} \right) \]
\[ \iff f(x) \in \bigcup \mathcal{G} \]
\[ \iff \text{there exists } D \in \mathcal{G} \text{ such that } f(x) \in D \]
\[ \iff \text{there exists } D \in \mathcal{G} \text{ such that } x \in f^{-1}(D) \]
\[ \iff x \in \bigcup_{D \in \mathcal{G}} f^{-1}(D) \]

(iv). We have

\[ x \in f^{-1}\left( \bigcap \mathcal{G} \right) \]
\[ \iff f(x) \in \bigcap \mathcal{G} \]
\[ \iff \text{for all } D \in \mathcal{G}, f(x) \in D \]
\[ \iff \text{for all } D \in \mathcal{G}, x \in f^{-1}(D) \]
\[ \iff x \in \bigcap_{D \in \mathcal{G}} f^{-1}(D) \]

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 \( C \).

We could have deduced (iv) from (iii) using De Morgan’s Laws, but it would not have saved work; in fact, it would have been a little longer.

**Exercise 2.14.** Rewrite and prove part (iii) of the above proposition using indexed families of sets.

Here’s what functions do with set difference:

**Corollary 2.15.** Let \( f : A \to B \), and let \( C, D \subset A \) and \( E, F \subset B \). Then:

(i) \( f(C \setminus D) \supset f(C) \setminus f(D) \);
(ii) \( f^{-1}(E \setminus F) = f^{-1}(E) \setminus f^{-1}(F) \);

**Proof.** (i). Let \( y \in f(C) \setminus f(D) \). Then \( y \in f(C) \) and \( y \not\in f(D) \). Since \( y \in f(C) \), we can choose \( x \in C \) such that \( y = f(x) \). We have \( x \notin D \) since \( y \not\in f(D) \). Thus \( x \in C \setminus D \). Hence \( y \in f(C \setminus D) \).
(ii). We have:

\[ f^{-1}(E \setminus F) = f^{-1}(E \cap F^c) \]
\[ = f^{-1}(E) \cap f^{-1}(F^c) \]
\[ = f^{-1}(E) \cap f^{-1}(F)^c \]
\[ = f^{-1}(E) \setminus f^{-1}(F) \]

Again, inverse images are nicer than forward ones. Part (ii) was a corollary of other facts concerning inverse images, but we had to prove (i) directly.
3. THE REAL NUMBERS

The real numbers form the foundation of mathematical analysis. Although the real numbers can be constructed starting only from the counting numbers 1, 2, . . . , let me reassure you that we won’t do this; instead we’ll be satisfied with a careful listing of the properties of the real numbers.\textsuperscript{16}

These come in three types: (1) arithmetic, (2) inequalities, and (3) the all-important Completeness Axiom, guaranteeing that the “real number line” has no gaps.

Notation and Terminology. $\mathbb{R}$ denotes the set of real numbers.

Essential properties of arithmetic:

(i) $(x + y) + z = x + (y + z)$ and $(xy)z = x(yz)$;
(ii) $x + y = y + x$ and $xy = yx$;
(iii) $x(y + z) = xy + xz$;
(iv) $0 + x = x$;
(v) $1x = x$;
(vi) $x - y = x + (-y)$;
(vii) $x - x = 0$;
(viii) $xx^{-1} = 1$ if $x \neq 0$;
(ix) $x/y = xy^{-1}$ if $y \neq 0$.

The importance of the arithmetic properties is that they imply all other properties of arithmetic.\textsuperscript{17} For example:

(i) if $x + z = y + z$ then $x = y$;
(ii) if $xz = yz$ and $z \neq 0$ then $x = y$;
(iii) $0x = 0$;
(iv) $-x = (-1)x$;
(v) $(-1)^2 = 1$;
(vi) if $xy = 0$ then either $x = 0$ or $y = 0$.

Here are the other important sets of real numbers:

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

$\mathbb{N} = \{1, 2, 3, . . . \}$

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

\textsuperscript{16}However, it’s important for you 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.

\textsuperscript{17}This is the purpose of axioms: assume them, and everything else follows.
(a) $1 \in \mathbb{N}$, and
(b) $n + 1 \in \mathbb{N}$ for all $n \in \mathbb{N}$.

This property of $\mathbb{N}$ is the Principle of Mathematical Induction: any subset $A \subset \mathbb{N}$ satisfying both $1 \in A$ and $n + 1 \in A$ for all $n \in A$ 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
\[ \mathbb{Q} = \left\{ \frac{n}{k} \mid n, k \in \mathbb{Z}, k \neq 0 \right\} \]
is the smallest subset of $\mathbb{R}$ containing $\mathbb{Z}$ and closed under division.

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

The Principle of Mathematical Induction is often the appropriate weapon to prove a statement about the natural numbers, for example:

**Exercise 3.1.** Prove by induction that $n^3 + 5n$ is divisible by 6 for each $n \in \mathbb{N}$.

**Integer powers.** For each $x \in \mathbb{R}$ and $n \in \mathbb{Z}$,
\[ x^n := \begin{cases} 
\underbrace{x \cdots x}^n & \text{if } n > 0 \\
\frac{1}{x^{-n}} & \text{if } n < 0 \text{ and } x \neq 0 \\
1 & \text{if } n = 0 \text{ and } x \neq 0.
\end{cases} \]

All the usual properties of algebra follow from the above, for example:

**Example.**
(i) $0x = 0$;
(ii) $-x = (-1)x$;
(iii) $(-1)^2 = 1$;
(iv) if $xy = 0$ then either $x = 0$ or $y = 0$.

**Essential properties of inequalities:**
(i) exactly one of $x > 0$, $x < 0$, or $x = 0$ holds;
(ii) $x + y > 0$ and $xy > 0$ for all $x, y > 0$;
(iii) $x < y \iff y - x > 0$;
(iv) $x > y \iff y < x$;
(v) $x \leq y$ if and only if either $x < y$ or $x = y$;
(vi) $x \geq y \iff y \leq x$. 
Definition. Inequalities are classified in two types:

(i) **strict**: \(x < y \text{ and } x > y\);
(ii) **weak**: \(x \leq y \text{ and } x \geq y\);

The importance of the above properties is that they imply all other properties of inequalities, for example:

(i) exactly one of \(x < y\), \(y > x\), or \(x = y\) holds;
(ii) if \(x < y\) and \(z < w\) then \(x + z < y + w\);
(iii) if \(x < y\) and \(z > 0\) then \(xz < yz\);
(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 \(y > 0\),

and similarly for weak inequalities.

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).

Functions which interact nicely with respect to inequalities deserve special mention:

**Monotone Functions:**

**Definition.** A real-valued function \(f\) is:

(i) **increasing** if \(f(x) \leq f(y)\) for all \(x < y\);
(ii) **decreasing** if \(f(x) \geq f(y)\) for all \(x < y\);
(iii) **monotone** if it’s increasing or decreasing;
(iv) **strictly increasing** if \(f(x) < f(y)\) for all \(x < y\);
(v) **strictly decreasing** if \(f(x) > f(y)\) for all \(x < y\);
(vi) **strictly monotone** if it’s strictly increasing or strictly decreasing.

**Observation.** Strictly monotone functions are 1-1.

**Definition.** (i) If \(a, b, x \in \mathbb{R}\), 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 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) \([a, b] := \{x \mid a \leq x \leq b\}\) (closed, bounded);
(b) \([a, \infty) := \{x \mid a \leq x\}\) (left-half-closed, unbounded);
(c) \((-\infty, b]\) := \{x \mid x \leq b\} \text{ (right-half-closed, unbounded)};
(d) \((-\infty, \infty) := \mathbb{R} \text{ (open, unbounded)};
(e) (a, b) := \{x \mid a < x < b\} \text{ (open, bounded)};\footnote{Somehow irritating, the notations for an element \((a, b)\) of \(\mathbb{R}^2\) and an open interval \((a, b)\) in \(\mathbb{R}\) are the same — it’s just a fact of life that you have to determine from the context which is intended.}
(f) (a, \infty) := \{x \mid a < x\} \text{ (open, unbounded)};
(g) (-\infty, b) := \{x \mid x < b\} \text{ (open, unbounded)};
(h) [a, b) := \{x \mid a \leq x < b\} \text{ (left-half-closed or right-half-open, bounded)};
(i) (a, b] := \{x \mid a < x \leq b\} \text{ (left-half-open or right-half-closed, bounded)}.

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

\textit{Example.} The Cartesian product \(\prod_{x \in [0, \infty)} [0, x]\) comprises all functions \(f: [0, \infty) \to \mathbb{R}\) such that
\[0 \leq f(x) \leq x\text{ for all } x \in [0, \infty).\]

\textit{Exercise 3.2.} Find \(\bigcap_{x \in (0, \infty)} (0, x)\) and \(\bigcap_{x \in \mathbb{R}} (-x, x)\), and prove your answers are correct.

If \(a = b\) then the closed interval \([a, b]\) is the \textit{singleton} (which means a set containing exactly one element) \(\{a\}\), while the intervals \((a, b), [a, b), \text{ and } (a, b]\) are empty.

\textit{Exercise 3.3.} Prove that the function \(f: [0, \infty) \to \mathbb{R}\) defined by \(f(x) = x^2\) is strictly increasing.

\textit{Definition.} A \textit{sequence} is a function with domain \(\mathbb{N}\).

If \(x\) is a sequence, then for each \(n \in \mathbb{N}\) the \textit{\(n\)th term} of \(x\) is defined as
\[x_n := x(n).\]

\textit{Notation and Terminology.} A sequence \(x\) is written \((x_n)_{n=1}^{\infty}\), or just \((x_n)\). Another popular notation for the sequence is \(\{x_n\}\), but we won’t use this, because it looks too much like a \textit{set} rather than a \textit{function}. The “tuple” notation \((x_n)\) reminds us of an element of a Cartesian product \(\prod_{n=1}^{\infty} A_n\), which is actually quite appropriate, since every sequence really is such an element.

Carrying the “tuple” notation a bit further, a sequence \((x_n)\) may be written \((x_1, x_2, \ldots)\). In fact, sometimes a sequence is specified by giving the first few terms, with the understanding that the pattern continues.
This is dangerous, because you are then depending upon the reader to
guess the same pattern as you intend. For example, is \((0, 1, \ldots)\) the
sequence \((n - 1)\) or the sequence \(((n - 1)^2)\)? The rule that saves us here
is Occam’s Razor: when there are two explanations for something, the
simpler one is most likely to be correct. Strictly speaking, specifying
a sequence by only giving the first few terms is an abuse: no matter
how many terms you specify, there are infinitely many sequences which
start out that way. But in practice there is usually no confusion.

**Exercise 3.4.** For each sequence below, find \(x_{87}\):

(a) \((1, 1/2, 1/3, \ldots)\);
(b) \((k, k + 1, k + 2, \ldots)\);
(c) \((1, -1, 1, -1, \ldots)\).

**Notation and Terminology.** If all the values \(x_n\) of a sequence \((x_n)\)
are elements of 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 \(\prod_{n=1}^{\infty} S\). Also, we
call a sequence in \(\mathbb{R}\) a **real sequence**

Actually, it’s sometimes convenient to allow sequences to “start some-
where other than 1”, that is, to allow the domain to be of the form
\(\{n \in \mathbb{Z} \mid n \geq k\}\) for some \(k \in \mathbb{Z}\), and in fact \(k = 0\) is used a lot.
Many of the essential properties of sequences don’t depend upon the
“starting point”, and for convenience we develop the general theory
for sequences starting at 1. Unless otherwise specified, a sequence is
assumed to have domain \(\mathbb{N}\).

Because the domain is so special, it’s much easier to check mono-
tonicity for a sequence:

**Exercise 3.5.** Prove that a real sequence \((x_n)\) is increasing if and only
if

\[
(2) \quad x_n \leq x_{n+1} \quad \text{for all } n \in \mathbb{N},
\]

and state and prove similar assertions for decreasing, strictly increasing,
and strictly decreasing. Hint: the main thing is to prove, by induction
on \(k\), that if \((x_n)\) satisfies \(2\) then for each \(n \in \mathbb{N}\) the inequality \(x_n \leq x_k\)
is true for all \(k > n\).

In the above exercise you are to apply induction to show that some-
ting is true for all \(k = n+1, n+2, n+3, \ldots\) — as opposed to what’s in
the “official” statement of Mathematical Induction, where \(k = 1, 2, \ldots\)
— this is pretty obviously a valid use of induction, and is frequently
used without further comment.

**Exercise 3.6.** (a) Prove by induction that if \(a > 1\) then the se-
quence \((a^n)\) is strictly increasing.
(b) Use (a) to prove that if \( a > 1 \) then \( a^n > 1 \) for all \( n \in \mathbb{N} \).

(c) Use (b) to prove that if \( 0 < b < 1 \) then \( b^n < 1 \) for all \( n \in \mathbb{N} \).

**Definition.** The absolute value of \( x \in \mathbb{R} \) is

\[
|x| := \begin{cases} 
  x & \text{if } x \geq 0 \\
  -x & \text{if } x < 0.
\end{cases}
\]

Note that \(|-x| = |x|\) for all \( x \in \mathbb{R} \).

**Lemma 3.7.** For all \( a, b \in \mathbb{R} \):

(i) \( \pm a \leq |a| \);

(ii) \( |a| \leq b \) 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, by the above we have \( -a \leq |−a| \). Therefore \( -a \leq |a| \) since \( |-a| = |a| \).

(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 \( |a| \leq b \). \(\square\)

**Triangle Inequality.** For all \( x, y \in \mathbb{R} \),

\[
|x + y| \leq |x| + |y| \quad \text{(alternate version)}.
\]

**Proof.** First, \( x \leq |x| \) and \( y \leq |y| \), so

\[
x + y \leq |x| + |y|.
\]

Next, \( -x \leq |x| \) and \( -y \leq |y| \), so

\[
-(x + y) = -x - y \leq |x| + |y|.
\]

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

\[ |x| - |y| \leq |x - y|. \]

Applying this inequality with \( x \) and \( y \) reversed, we get

\[ |y| - |x| \leq |y - x|. \]

But

\[ |y - x| = |-(y - x)| = |x - y|, \]

so

\[ |y| - |x| \leq |x - y|. \]

Thus

\[ \pm(|x| - |y|) \leq |x - y|, \]

therefore

\[ | |x| - |y| | \leq |x - y|. \] \( \square \)

**Definition.** Let \( A \subset \mathbb{R} \) and \( x \in A \).

(i) \( x \) is the maximum of \( A \) if \( y \leq x \) for all \( y \in A \).

(ii) \( x \) is the minimum of \( A \) if \( x \leq y \) for all \( y \in A \).

**Notation and Terminology.** \( \max A \) denotes the maximum of \( A \), and \( \min A \) the minimum. The maximum is also called the largest, or greatest element of \( A \), and similarly for smallest, least, and minimum.

It’s easy to see that \( A \) can have at most one maximum — this is why it makes sense to use the definite article “the” in the definition and introduce the notation “max \( A \)” (and similarly for minimum). You should convince yourself of the uniqueness of the maximum, and likewise for similar situations elsewhere.

It’s easy to see that \( \min A = -\max(-A) \) (where \( -A := \{ -x \mid x \in A \} \)). 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, and vice-versa.

If \( A \) is finite\(^{19}\) and nonempty, then \( \max A \) exists\(^{20}\). Similarly for \( \min A \).

However, infinite sets can easily fail to have a maximum or minimum, for example:

**Exercise 3.8.** Prove that the left-half-open interval \((0, 1]\) has a maximum but no minimum.

**Exercise 3.9.** Let \( A \subset B \subset \mathbb{R} \) and \( x \in A \). Prove that if \( x = \min B \) then \( x = \min A \).

\(^{19}\)And you can look in the next section for the definition of this!

\(^{20}\)as you can check using induction
Definition. Let $A \subset \mathbb{R}$ and $x \in \mathbb{R}$.

(i) $x$ is an upper bound of $A$ if $y \leq x$ for all $y \in A$.
(ii) $x$ is a lower bound of $A$ if $x \leq y$ for all $y \in A$.
(iii) $A$ is bounded above if it has an upper bound.
(iv) $A$ is bounded below if it has a lower bound.

Exercise 3.10. Prove that if $A$ is bounded above, then the set $U$ of upper bounds of $A$ is an interval which is unbounded to the right. This means that you must show that for all $a, b \in \mathbb{R}$, if $a < b$ and $a \in U$, then $b \in U$.

Exercise 3.11. Prove that every real number is simultaneously an upper and a lower bound for $\emptyset$.

As with max and min, it’s 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.

Some sets of real numbers, even if they don’t have a maximum, might have something almost as good:

Definition. Let $A \subset \mathbb{R}$.

(i) The supremum of $A$ is the minimum of the set of upper bounds of $A$.
(ii) The infimum of $A$ is the maximum of the set of lower bounds of $A$.

Notation and Terminology. The supremum of $A$ is denoted sup $A$ and also (for obvious reasons) called the least upper bound of $A$. Similarly for infimum, inf $A$, and greatest lower bound.

The supremum of $A$ is unique if it exists, and similarly for infimum. It follows from our knowledge concerning max, min, upper bounds, and lower bounds that $A$ has an inf if and only if $(-A)$ has a sup, in which case inf $A = -\sup(-A)$. Consequently, any fact concerning suprema gives more-or-less automatically a corresponding fact concerning infima.

Here’s a more explicit reformulation of the definition of supremum:

Observation. (i) $x = \sup A$ if and only if both:

(1) $x$ is an upper bound of $A$, and
(2) $x \leq y$ for every upper bound $y$ of $A$.

(ii) If $\sup A$ exists, then the set of upper bounds of $A$ is the interval $[\sup A, \infty)$.

Here’s how suprema are related to maxima:
Observation. A \( \subset \mathbb{R} \) has a maximum if and only if both:

(i) A has a supremum, and
(ii) \( \sup A \in A \),
in which case \( \max A = \sup A \).

Example. We have \( \sup(0, 1) = 1 \), but \((0, 1)\) has no maximum.

It’s built into the definition that if A has a supremum then A is bounded above. However, the converse is false:

Example. The empty set is bounded above but has no supremum.

Thus, if \( \sup A \) exists then A is nonempty and bounded above. The last axiom of the real numbers, certainly the deepest[^21], guarantees that there are no other obstructions to the existence of a supremum:

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

Observation. It follows that if \( A \subset \mathbb{R} \) is nonempty and bounded below then A has an infimum. When we invoke this we just say “by the Completeness Axiom”.

The following trivial consequence of being the least upper bound is surprisingly useful — in fact, so useful that it’s given a name:

Approximation of Suprema. For all \( t < \sup A \) there exists \( x \in A \) such that \( t < x \).

**Exercise 3.12.** Prove that if \( A, B \subset \mathbb{R} \) are nonempty and bounded above, then
\[
\sup(A \cup B) = \max\{\sup A, \sup B\}.
\]

**Exercise 3.13.** Prove that if \( A, B \subset \mathbb{R} \) are nonempty and bounded above, then
\[
\sup(A + B) = \sup A + \sup B,
\]
where \( A + B := \{x + y \mid x \in A, y \in B\} \).

The following result confirms our intuition that if we add up the number 1 enough times we’ll get beyond any given real number. It’s perhaps surprising that we need the Completeness Axiom of the real numbers to prove it:

[^21]: Indeed, its crucial role in analysis was not recognized until the 19th century.
[^22]: Informally, the reason for the name is that the axiom says the real line has no gaps.
Archimedean Principle. For all $x \in \mathbb{R}$ there exists $n \in \mathbb{N}$ such that $n > x$.

Proof. We argue by contradiction. Suppose not. More precisely, assume the statement of the theorem is false. Then there exists $x \in \mathbb{R}$ such that $n \leq x$ for all $n \in \mathbb{N}$. In particular, $\mathbb{N}$ is bounded above. Of course $\mathbb{N} \neq \emptyset$, so by the Completeness Axiom $\sup \mathbb{N}$ exists. Since $\sup \mathbb{N} - 1 < \sup \mathbb{N}$, 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. We conclude that our initial assumption must be false, therefore the statement of the theorem is true.

More generally (and this follows very easily from the Archimedean Principle):

Exercise 3.14. Prove:

(a) for all $a > 0$ and $x \in \mathbb{R}$ there exists $n \in \mathbb{N}$ such that $na > x$;
(b) for all $a > 0$ there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < a$;
(c) 0 is the infimum of the set $\left\{ \frac{1}{n} \bigg| n \in \mathbb{N} \right\}$.

When the above exercise is invoked we just say “by the Archimedean Principle”.

Exercise 3.15. Prove that if $a > 1$ then for all $M \in \mathbb{R}$ there exists $n \in \mathbb{N}$ such that $a^n > M$. Hint: $a = 1 + b$ with $b > 0$.

---

$^{23}$not the one involving running down the street screaming “Eureka”!

$^{24}$Here’s a caution about indirect proofs, which means either by contrapositive or by contradiction. These are inherently harder to follow than direct proofs, so you must take extra care in writing them. It’s always helpful to announce up front that an indirect proof is coming. Actually, proofs by contradiction are the trickiest of all to write clearly, and should probably be avoided whenever possible. When you find yourself proving something by contradiction, a lot of the time you could rewrite the proof using contrapositive without too much pain, and it could make the proof clearer. As a matter of taste I eschew proof by contradiction unless it’s necessary or makes the proof significantly shorter or easier to follow. Proofs by contradiction are inherently obscure — think of it: you start by assuming something which you know to be false, which is confusing in itself, but even worse it hurls you into some fantasy land where every proposition is both true and false; it’s a dangerous place to be.
Observation. It follows easily from the Archimedean Principle that for all \( x \in \mathbb{R} \) there exists \( n \in \mathbb{Z} \) such that \( n \leq x \), and when we invoke this we just say “by the Archimedean Principle”.

For sets of integers we can improve upon the Completeness Axiom:

**Proposition 3.16.** Let \( A \subset \mathbb{Z} \) be nonempty.

(i) If \( A \) is bounded above, then \( A \) has a maximum.

(ii) If \( A \) is bounded below, then \( A \) has a minimum.

**Proof.** For (i), by the Completeness Axiom \( s := \sup A \) exists. Then \( s - 1 < s \), so by Approximation of Suprema there exists \( n \in A \) such that \( n > s - 1 \). For all \( k \in A \) we have

\[
k \leq s < n + 1,
\]

so we must have \( k \leq n \). Therefore \( n = \max A \).

For (ii), just take negatives. \( \square \)

As an immediate corollary, we get:

**Well-Ordering Principle.**

25 Every nonempty subset of \( \mathbb{N} \) has a minimum.

The following corollary shows that the integers mark off the real line as a sort of “infinite ruler”.

**Floor Theorem.**

26 For all \( x \in \mathbb{R} \) there exists a unique \( n \in \mathbb{Z} \) such that

\[
n \leq x < n + 1.
\]

**Proof.** Let \( x \in \mathbb{R} \). We must show two things: that there exists \( n \in \mathbb{Z} \) such that \( n \leq x < n + 1 \), and that this \( n \) is unique, i.e., that there is at most one such \( n \).

Put \( A = \{k \in \mathbb{Z} \mid k \leq x\} \). Then \( A \) is a set of integers, which is nonempty by the Archimedean Principle, and is bounded above by \( x \). Thus \( n := \max A \) exists. 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 \( n \leq x < n + 1 \). This proves existence.

For the uniqueness, we argue by contradiction: suppose that in addition to the \( n \) we found above, we also have \( k \in \mathbb{Z} \) such that \( k \leq x < k + 1 \)

25 We don’t really need the Completeness Axiom for this — it’s actually equivalent to the Principle of Mathematical Induction — but it’s easier to derive it this way.

26 So called because the unique integer \( n \) is popularly called the “floor” of \( x \).
and \( k \neq n \). Without loss of generality \( n < k \). Then \( n + 1 \leq k \) since \( n \) and \( k \) are integers. But then 
\[
x < n + 1 \leq k \leq x,
\]
a contradiction. Therefore such a \( k \) does not exist. \( \square \)

**Density of Rationals.** *Strictly between any two distinct real numbers there exists a rational number.*

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

The Completeness Axiom is what allows for *decimal expansions* of real numbers: Given \( x \geq 0 \), first define 
\[
n = \max \{ l \in \mathbb{Z} \mid l \leq x \}.
\]
Next, define \( d_1, d_2, \ldots \) by 
\[
d_1 = \max \left\{ l \in \mathbb{Z} \mid n + \frac{l}{10} \leq x \right\},
\]
\[
d_2 = \max \left\{ l \in \mathbb{Z} \mid n + \frac{d_1}{10} + \frac{l}{10^2} \leq x \right\},
\]
and continuing inductively.\(^{27}\) Each \( d_i \) is an integer between 0 and 9, inclusive. Put 
\[
A = \left\{ n + \sum_{i=1}^{j} \frac{d_i}{10^i} \mid j \in \mathbb{N} \right\}.
\]
Then \( x = \sup A \), and we say 
\[
x = n.d_1d_2 \ldots
\]
is a decimal expansion of \( x \).

\(^{27}\)Here the choice of each successive \( d_n \) depends 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’ll allow ourselves to do this kind of thing willy-nilly.
On the other hand, given a nonnegative integer $n$ and integers $d_1, d_2, \ldots$ between 0 and 9, again put

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

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 $n/10^k$ 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.

**Density of Irrationals.** *Strictly between any two distinct real numbers there exists an irrational number.*

*Proof.* Let $a < b$, and pick any irrational number $x$. Then $a + x < b + x$, so by Density of Rationals we can choose $y \in \mathbb{Q}$ such that $a + x < y < b + x$. Then

$$a < y - x < b,$$

and $y - x$ is irrational since $y$ is rational and $x$ is irrational. \square

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 $y^n = x$.

In both of the above cases, $y$ is called the $n$th root of $x$, denoted $x^{1/n}$ or $\sqrt[n]{x}$. For example, let’s verify the existence of square roots:
Proposition 3.17. For every \( t \geq 0 \) there exists a unique \( s \geq 0 \) such that \( s^2 = t \).

Proof. Note straightaway that if \( 0 \leq x < y \) then \( x^2 < y^2 \), so there is at most one such \( s \).

The case \( t = 0 \) is trivial, so assume \( t > 0 \). Define

\[
A = \{ x > 0 \mid x^2 < t \}.
\]

Then \( A \) is nonempty since if \( 0 < x < \min\{1, t\} \) then \( x^2 \leq x < t \), and \( A \) is bounded above because if \( y > \max\{1, t\} \) then \( y^2 \geq y \geq t \), making \( \max\{1, t\} \) an upper bound. Thus by the Completeness Axiom we can put \( s = \sup A \). Then \( s > 0 \). We show \( s^2 = t \) by contradiction: first suppose \( s^2 < t \). Then for \( 0 < r < 1 \) we have

\[
(s + r)^2 = s^2 + 2sr + r^2 \leq s^2 + 2sr + r = s^2 + r(2s + 1),
\]

which is less than \( t \) if \( r < (t - s^2)/(2s + 1) \). But then for any such \( r \) we have \( s + r \in A \) and \( s + r > \sup A \), a contradiction. On the other hand, suppose \( s^2 > t \). Then for \( r > 0 \) we have

\[
(s - r)^2 = s^2 - 2sr + r^2 \geq s^2 - 2sr,
\]

which is greater than \( t \) if \( r < (s^2 - t)/(2s) \). For any such \( r \), by Approximation of Suprema there exists \( x \in A \) such that \( x > s - r \). But then

\[
t < (s - r)^2 < x^2,
\]

and contradiction. Therefore we must have \( s^2 = t \) \( \Box \)

We’ll defer the justification of the existence of more general roots until after the Intermediate Value Theorem (and we won’t run the risk of circular reasoning since we’ll only use more general roots in examples until then). However, we’ll actually need square roots soon, to define the norm in \( \mathbb{R}^n \).

Exercise 3.18. Prove that if \( 0 \leq a < b \) then \( \sqrt{a} < \sqrt{b} \).

Exercise 3.19. Let \( x_0 = 1 \), and for \( n \in \mathbb{N} \) inductively define

\[
x_n = \sqrt{1 + x_{n-1}}.
\]

Prove that

\[
1 < x_n < x_{n+1} \quad \text{for all } n \in \mathbb{N}.
\]

Hint: induction.

Rational powers. For each \( n \in \mathbb{N} \) and \( k \in \mathbb{Z} \),

\[
x^{k/n} := (x^{1/n})^k \quad \text{(where } x \geq 0 \text{ if } n \text{ is even}).
\]
It’s 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 \( x, y \in \mathbb{R} \):

(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 won’t need these in the formal development of the theory, we don’t prove them here. Much later (after the material on integration) we’ll 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’s 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.

**Proposition 3.20.** \( \sqrt{2} \) is irrational.

**Proof.** We argue by contradiction.\(^{28}\) Suppose \( \sqrt{2} = \frac{p}{q} \) for integers \( p \) and \( q \); 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. \( \Box \)

\(^{28}\)If you were going to be stranded on a desert island and could only take one proof by contradiction with you, you should choose this one. Half a millennium BC, there was a group of philosopher-mathematicians called the Pythagoreans — led by the man himself — who believed (in modern terminology) that all numbers were rational. The (apocryphal) story is that, while the group was on a sea-voyage, one of their members proved that \( \sqrt{2} \) is irrational, and the group threw him overboard to cover up the discovery.
4. Countability

It’s obvious that there are lots of natural numbers — in fact, infinitely many. And it’s obvious that there are more real numbers than natural numbers — after all, every natural number is a real number, and there are real numbers that aren’t natural. But it’s not so obvious that there is a much more subtle difference between the abundance of natural numbers and of real numbers. The same reasoning would lead us to conclude that there must be more rational numbers than natural numbers. However, it turns out that, in a technical sense we’ll discuss in this section, there are more real numbers than natural numbers, but exactly as many rational numbers as natural numbers. And we’ll need to deal with this difference, so in this section we’ll find out what it’s all about.

Main definitions

(i) \( B \) is an image of \( A \) if there exists an onto function \( f : A \to B \).
(ii) \( B \) is a 1-1 image of \( A \) if there exists a 1-1 onto function \( f : A \to B \).

A set is:

(i) finite if it’s either empty or an image of \( \{1, \ldots, n\} \) for some \( n \in \mathbb{N} \);
(ii) infinite if it’s not finite;
(iii) countable if it’s either empty or an image of \( \mathbb{N} \);
(iv) uncountable if it’s not countable.

Discussion

Thus a countable set may be finite or infinite, but every uncountable set is infinite.

A set is countable if and only if it’s the range of a sequence.

Exercise 4.1. Prove that every infinite subset of \( \mathbb{N} \) is unbounded above.

Let’s agree that we understand finite sets pretty well.\(^2\) Infinite sets, on the other hand, have some amazing properties (for example, a set is infinite if and only if it is an image of one of its proper subsets), and have mystified a lot of smart people for a long time. The modern

---

\(^2\)Strictly speaking, the familiar properties of finite sets should be proven, which would involve tedious but routine induction arguments.
Main results

We’ll state the main results on countability, postponing the proofs until we develop some tools.

**Theorem 4.2.** 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) $\mathbb{Q}$.

**Theorem 4.3.** The following sets are uncountable:

(i) every nondegenerate interval;
(ii) $\mathbb{R}$;
(iii) the irrational numbers.

Theory

Let’s collect some handy tools for detecting countability:

**Observation.** If $B$ is a 1-1 image of $A$, then:

(i) $A$ is infinite if and only if $B$ is;
(ii) $A$ is countable if and only if $B$ is.

**Lemma 4.4.** Every image of a countable set is countable.

*Proof.* Assume $B$ is countable and $f : B \to A$ is onto. Without loss of generality $B$, hence $A$, is nonempty. Since $B$ is countable, it’s the range of a sequence $(x_n)$. Then $(f(x_n))$ is a sequence with range $A$. Therefore $A$ is countable. \(\square\)

**Lemma 4.5.** Every subset of a countable set is countable.

---

\[30\text{When the theory of infinite sets was first worked out in the late 19th century, by Cantor, it was so controversial that other mathematicians ridiculed Cantor for it. Unfortunately Cantor was delicate — he broke down under this ridicule, spending the last years of his life in an insane asylum. But now we accept his theory as quite natural.}\]
Proof. Let $A \subset B$ with $B$ countable.

Case 1. $A = \emptyset$. Then $A$ is trivially countable.

Case 2. $A \neq \emptyset$. Then we can choose $a \in A$. Define $f : B \to A$ by

$$f(x) = \begin{cases} x & \text{if } x \in A \\ a & \text{if } x \notin A. \end{cases}$$

Then $f$ is onto, so $A$ is countable since $B$ is. \hfill \Box

**Corollary 4.6.** $A$ is countable if there exists a countable set $B$ and a 1-1 function from $A$ to $B$.

Proof. Assume $B$ is countable and $f : A \to B$ is 1-1. Then $\text{ran } f$ is countable since $\text{ran } f \subset B$. Therefore $A$ is countable since $f : A \to \text{ran } f$ is 1-1 onto. \hfill \Box

**Proofs of main results**

**Proof of Theorem 4.2.** (i). Define $f : \mathbb{N}^2 \to \mathbb{N}$ 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 $\bigcup_{i \in I} A_i$ where $I$ and every $A_i$ are countable. Without loss of generality $I$ and every $A_i$ are nonempty, because $I = \emptyset$ makes the whole union empty, while an empty $A_i$ has no effect on the union. Let $I = \{i_n \mid n \in \mathbb{N}\}$, and for each $n \in \mathbb{N}$ put $B_n = A_{i_n}$. Then

$$\bigcup_{i \in I} A_i = \bigcup_{n=1}^{\infty} B_n.$$

Next, for each $n \in \mathbb{N}$ let $B_n = \{x_{n,k} \mid k \in \mathbb{N}\}$. Then

$$\bigcup_{n=1}^{\infty} B_n = \{x_{n,k} \mid (n,k) \in \mathbb{N}^2\},$$

which 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).

Thus it suffices to prove the following: for every $n \in \mathbb{N}$, every product of $n$ countable sets is countable. We argue by induction. For $n = 1$
the statement is trivially true, because there is an obvious 1-1 onto
function from $\prod_{i=1}^{1} A_i$ to $A_1$.

Let $n > 1$, and assume that every product of $n - 1$ countable sets
is countable. We deduce that every product of $n$ countable sets is
countable, which will complete the proof. Let $A_1, \ldots, A_n$ be countable.
We use a trick: for each $y \in A_1$ define

$$B_y = \{(x_1, \ldots, x_n) \in \prod_{i=1}^{n} A_i \mid x_1 = y\}.$$

Then

$$\prod_{i=1}^{n} A_i = \bigcup_{y \in A_1} B_y,$$

a countable union of sets. Thus it suffices to show that every $B_y$ is
countable. Fix $y \in A_1$, and define

$$f: \prod_{i=2}^{n} A_i \to B_y$$

by

$$f(x_2, \ldots, x_n) = (y, x_2, \ldots, x_n).$$

It’s obvious that $f$ is onto. By the induction hypothesis, $\prod_{i=2}^{n} A_i$ is
countable. Therefore $B_y$ is countable, being an image of a countable
set.

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

(v). $\mathbb{Z}^2$ is a finite product of countable sets, hence is countable.

(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.

In the above proof we defined several functions with domain a Carte-
sian product, and there is a minor notational issue to clear up: for
example we defined $f: \mathbb{N}^2 \to \mathbb{N}$ by $f(n, k) = 2^n3^k$. Strictly speaking,
the rules of function notation would force us to write “$f((n, k))$”, since
it should be of the form $f(x)$ with $x \in \mathbb{N}^2$, and a typical element of $\mathbb{N}^2$
is an ordered pair. However, the double parentheses in “$((n, k))$” are
cumbersome and confusing, and in practice is not used.

\[31\] It’s also 1-1, but we don’t need this fact.
Also in the above proof, we said “let \( B_n = \{ x_{n,k} \mid k \in \mathbb{N} \} \)”. The subscript “\( n, k \)” looks a bit odd at first, but we’re using it in same sense that the usual subscript notation for sequences is used. We know \( B_n \) is the range of some sequence, and we have a sequence for every \( n \), so it behooves us to give some thought to the choice of notation. Our choice has the advantage of reminding us that we’re dealing with sequences; we’re saying \( B_n \) is the range of the sequence \( k \mapsto x_{n,k} \), and then the union is the range of the function \( (n, k) \mapsto x_{n,k} \) with domain \( \mathbb{N}^2 \).

**Exercise 4.7.** A real number is **algebraic** if it’s a root of a polynomial with integer coefficients. Prove that the set of all algebraic numbers is countable.

**Exercise 4.8.** Prove that the family of all finite subsets of \( \mathbb{N} \) is countable.

**Proof of Theorem 4.3.** (i). To show every nondegenerate interval is uncountable, it suffices to show \( [0, 1] \) is uncountable, because every nondegenerate interval contains a 1-1 image of \( [0, 1] \), and a set containing an uncountable subset must itself be uncountable. Moreover it suffices to show the subset\(^{32}\)

\[
A := \{ .a_1a_2 \cdots \mid a_i = 1 \text{ or } 2 \text{ for all } i \in \mathbb{N} \}
\]

of \([0, 1]\) is uncountable. We use what’s become known as the **Cantor Diagonalization Argument**: let \( (x_n) \) be a sequence in \( A \). We show \( A \neq \{ x_n \mid n \in \mathbb{N} \} \). For each \( n \in \mathbb{N} \) let

\[
x_n = .a_{n1}a_{n2}a_{n3} \cdots.
\]

Define \( b = .b_1b_2 \cdots \in A \) by

\[
b_i = \begin{cases} 
  1 & \text{if } a_{ii} = 2 \\
  2 & \text{if } a_{ii} = 1.
\end{cases}
\]

Then \( b \) differs from each \( x_n \) in the \( n \)th decimal place, so \( b \notin \{ x_n \mid n \in \mathbb{N} \} \).

(ii). This is a special case of (i), since \( \mathbb{R} \) is the nondegenerate interval \( (-\infty, \infty) \).

(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. \( \square \)

**Exercise 4.9.** Prove that the family of all subsets of \( \mathbb{N} \) is uncountable.

\(^{32}\)chosen to avoid the nuisance of ambiguous decimal expansions
Auxiliary results

There are four more countability tools that are handy, although we won’t use them in this section:

**Lemma 4.10.** If $A$ is countable, then there exists a 1-1 function from $A$ to $\mathbb{N}$.

**Proof.** Assume $A$ is countable.

Case 1. $A = \emptyset$. Then the empty set is a 1-1 function from $A$ to $\mathbb{N}$.

Case 2. $A \neq \emptyset$. Since $A$ is countable, there exists an onto function $f : \mathbb{N} \to A$. Our strategy is to get a 1-1 function $A \to \mathbb{N}$ by picking a single inverse image for each element of $A$. But we want a definite way to do this. Fortunately, the properties of the natural numbers give us a way to do it: Define $g : A \to \mathbb{N}$ by

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

Note that $g$ is well-defined by the Well-Ordering Principle, since the pre-image $f^{-1}(\{x\})$ is nonempty because $f$ is onto. We must show that $g$ is 1-1.

Let $x, y \in A$ with $x \neq y$. Then the sets $\{x\}$ and $\{y\}$ are disjoint, hence so are $f^{-1}(\{x\})$ and $f^{-1}(\{y\})$. Thus we must have

$$g(x) = \min f^{-1}(\{x\}) \neq \min f^{-1}(\{y\}) = g(y). \quad \square$$

Our next result shows that $\mathbb{N}$ is the smallest infinite set:

**Lemma 4.11.** $A$ is infinite if and only if there exists a 1-1 sequence in $A$.

**Proof.** First assume $A$ is infinite. We will define a sequence $(x_n)$ in $A$ such that $x_{n+1} \notin \{x_1, \ldots, x_n\}$ for all $n \in \mathbb{N}$, which will imply $(x_n)$ is 1-1. First, since $A$ is infinite, it’s nonempty, so we can pick $x_1 \in A$.

Let $n \in \mathbb{N}$, and assume that we have distinct $x_1, \ldots, x_n \in A$. Since $A$ is infinite and $\{x_1, \ldots, x_n\}$ is finite, we can pick $x_{n+1} \in A \setminus \{x_1, \ldots, x_n\}$. This shows that we can inductively define a sequence with the desired properties.

Conversely, if $(x_n)$ is a 1-1 sequence in $A$, then $A$ is infinite since it contains the infinite set $\{x_n \mid n \in \mathbb{N}\}$. \quad \square

If $A \subset \mathbb{N}$ we can do better:

**Lemma 4.12.** Every infinite subset $A$ of $\mathbb{N}$ is the range of a strictly increasing sequence $(n_k)$, so that in particular $A$ is a 1-1 image of $\mathbb{N}$.

**Proof.** Define $n_k$ inductively by

$$n_k = \begin{cases} 
\min A & \text{if } k = 1 \\
\min A \setminus \{n_1, \ldots, n_{k-1}\} & \text{if } k > 1.
\end{cases}$$
First let’s verify that the sequence \((n_k)\) is well-defined; this means we have to show that the inductive definition can be applied for every \(k \in \mathbb{N}\). We can accomplish this by showing that for every \(k > 1\), if \(n_1, \ldots, n_{k-1}\) have been defined then \(n_k\) can be defined as above. Since \(\{n_1, \ldots, n_{k-1}\}\) is a finite subset of \(A\) and \(A\) is infinite, the difference \(A \setminus \{n_1, \ldots, n_{k-1}\}\) is nonempty, so has a minimum by the Well-Ordering Principle. Thus \(n_k\) can be defined.

Next, for all \(k \in \mathbb{N}\) we have \(n_{k+1} > n_k\) since we’re taking a minimum in the integers, hence the sequence \((n_k)\) is strictly increasing, in particular 1-1.

Finally, let’s show \(A = \{n_k \mid k \in \mathbb{N}\}\). Let \(n \in A\). Since \((n_k)\) is 1-1 and \(\mathbb{N}\) is infinite, \(\{n_k \mid k \in \mathbb{N}\}\) is an infinite subset of \(A\). Thus there exists \(k \in \mathbb{N}\) such that \(n < n_k\). Since \(n_{k+1} > n_k\), without loss of generality \(k > 1\). Then
\[
 n < n_k = \min A \setminus \{n_1, \ldots, n_{k-1}\}.
\]
Since \(n \in A\),
\[
 n \notin A \setminus \{n_1, \ldots, n_{k-1}\},
\]
so we must have
\[
 n \in \{n_1, \ldots, n_{k-1}\},
\]
and we’re done. \(\square\)

The above proof was very fussy, but the basic idea was simple: pick the smallest element of \(A\), then the next smallest, and so on. The result itself says that \(A\) can be listed in increasing order — this might sound obvious, but a proof was necessary. It’s certainly not true for all countable subsets of \(\mathbb{R}\):

**Exercise 4.13.** Prove that there is no strictly increasing sequence whose range is \(\mathbb{Q}\).

To complete our “countability toolkit”, we show that, among infinite sets, all the countable ones look like \(\mathbb{N}\):

**Lemma 4.14.** \(A\) is countable and infinite if and only if it’s a 1-1 image of \(\mathbb{N}\).

**Proof.** The converse direction is trivial, so assume \(A\) is countable and infinite. Then there exists a 1-1 function \(g: A \to \mathbb{N}\), and since \(A\) is infinite so is \(g(A)\). Since \(g: A \to g(A)\) is 1-1 onto, we can replace \(A\) by \(g(A)\), so without loss of generality \(A \subseteq \mathbb{N}\). Then we are done by the preceding lemma. \(\square\)
5. Metric spaces

The mathematical theory of analysis, which is our main interest in this course, centers around being able to control distances. In calculus you learned a little about controlling distances in the real line \( \mathbb{R} \), and to a lesser extent in \( \mathbb{R}^2 \) and \( \mathbb{R}^3 \). The general sort of structure necessary for this sort of analysis is a metric space, which includes all the Euclidean spaces \( \mathbb{R}^n \) as well as many others:

**Definition.** Let \( X \) be a set. A **metric** on \( X \) is a function \( d: X^2 \to \mathbb{R} \) such that for all \( x, y, z \in X \):

1. \( d(x, y) \geq 0 \), and \( d(x, y) = 0 \) if and only if \( x = y \) (**positivity**);
2. \( d(x, y) = d(y, x) \) (**symmetry**);
3. \( d(x, y) \leq d(x, z) + d(z, y) \) (**triangle inequality**).

The pair \( (X, d) \) is a **metric space**, although we usually abuse the terminology by referring to \( X \) itself as the metric space if \( d \) is understood.

**Example.** The usual distance on the real line, defined by

\[
d(x, y) = |x - y| \quad \text{for } x, y \in \mathbb{R},
\]

is the archetypal example of a metric. Thus \( \mathbb{R} \) becomes a metric space — indeed, the axioms for a metric space were chosen to mimic the properties of distance in \( \mathbb{R} \). However, it’s important to keep in mind that a metric space is not just the set, but also the specific metric on that set. There are other metrics on \( \mathbb{R} \); the following exercise shows how to get a metric on \( \mathbb{R} \) (in fact, any set) which is very different from the familiar Euclidean distance:

**Exercise 5.1.** Let \( X \) be a set, and define \( d: X^2 \to \mathbb{R} \) by

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

Prove that \( d \) is a metric on \( X \) (the **discrete metric**).

**Notation and Terminology.** In the following, letters such as \( X \) and \( Y \) will refer to metric spaces, unless otherwise specified, and \( d \) will denote a metric on any of these sets.

An element of a metric space \( X \) is also called a **point** of \( X \).

**Alternate Triangle Inequality.** For all \( x, y, z \in X \),

\[
|d(x, z) - d(y, z)| \leq d(x, y).
\]

---

[^33]: Please keep in mind that the “\( d \)” on \( X \) is probably unrelated to the “\( d \)” on \( Y \)!
Proof. We have
\[ d(x, z) \leq d(x, y) + d(y, z) \]
and
\[ d(y, z) \leq d(y, x) + d(x, z) = d(x, y) + d(x, z), \]
so
\[ d(x, z) - d(y, z) \leq d(x, y) \quad \text{and} \quad d(y, z) - d(x, z) \leq d(x, y), \]

hence
\[ |d(x, z) - d(y, z)| \leq d(x, y). \]

\[ \square \]

Proposition 5.2. If \((X, d)\) is a metric space and \(Y \subset X\), then the restriction \(d|Y^2\) is a metric on \(Y\).

Proof. Obviously, the restriction \(d|Y^2\) still satisfies the axioms of a metric. \(\square\)

Thus, from one metric space we can get lots more:

Definition. In the situation of the above proposition, \(Y\) with the restricted metric is a subspace of \(X\).

Example. The closed unit interval \([0, 1]\), a subspace of \(\mathbb{R}\), is a very important metric space.

The Euclidean spaces \(\mathbb{R}^n\) are very special sorts of metric spaces, where the metric comes from a norm:

Definition. Let \(X\) be a (real) vector space. A norm on \(X\) is a function \(\| \cdot \| : X \to \mathbb{R}\) such that for all \(x, y \in X\):

(i) \(\|x\| \geq 0\), and if \(\|x\| = 0\) then \(x = 0\) (positivity);
(ii) \(\|cx\| = |c|\|x\|\) for all \(c \in \mathbb{R}\) (homogeneity);
(iii) \(\|x + y\| \leq \|x\| + \|y\|\) (the triangle inequality).

The pair \((X, \| \cdot \|)\) is normed space, although we usually abuse the terminology by referring to \(X\) itself as the normed space if \(\| \cdot \|\) is understood.

Exercise 5.3. Prove that \(\|0\| = 0\) in any normed space.

Proposition 5.4. Every normed space becomes a metric space if we define
\[ d(x, y) = \|x - y\|. \]

Proof. Let \(X\) be a normed space, and let \(x, y, z \in X\).

First, by definition of norm we have \(d(x, y) \geq 0\), and if \(d(x, y) = 0\) then \(x - y = 0\), hence \(x = y\).
Next,
\[ d(x, y) = \| x - y \| = | -1 | \| x - y \| = \|(-1)(x - y)\| = \|y - x\| = d(y, x). \]

Finally,
\[ d(x, y) = \| x - y \| = \| x - z + z - y \| \leq \| x - z \| + \| z - y \| = d(x, z) + d(z, y). \quad \square \]

**Proposition 5.5.** The formula
\[ \| x \| : = \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2} \]
defines a norm on \( \mathbb{R}^n \), called the Euclidean norm.

Before proving this proposition, we need an auxiliary result, which is actually one of the most important inequalities in analysis:

**Cauchy-Schwartz Inequality.** For all \( x, y \in \mathbb{R}^n \), \( |x \cdot y| \leq \|x\| \|y\| \), where \( x \cdot y = \sum_{i=1}^{n} x_i y_i \).

**Proof.** Consult your linear algebra book. \( \square \)

**Proof of Proposition 5.5.** The only nonobvious property is the triangle inequality: for \( x, y \in \mathbb{R}^n \) we have
\[ \|x + y\|^2 = (x + y) \cdot (x + y) = x \cdot x + y \cdot y + x \cdot y + y \cdot x \]
\[ = \|x\|^2 + \|y\|^2 + 2x \cdot y \]
\[ \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \]
\[ = \left( \|x\| + \|y\| \right)^2, \]
hence \( \|x + y\| \leq \|x\| + \|y\| \). \( \square \)

**Exercise 5.6.** Prove that
\[ \| x \|_{\infty} := \max_{i} |x_i| \]
defines a norm on \( \mathbb{R}^n \).

**Definition.** Let \( X \) be a metric space, \( x \in X \), and \( r > 0 \).

(i) The **open ball of radius** \( r \) and **center** \( x \) is
\[ B_r(x) := \{ y \in X \mid d(x, y) < r \}. \]

---

In your calculus course you probably saw a geometric proof of the Cauchy-Schwartz Inequality for \( \mathbb{R}^2 \) and \( \mathbb{R}^3 \), involving the Law of Cosines; the general proof requires a completely different argument, depending not upon trigonometry but rather upon the properties of the dot product \( x \cdot y \).
(ii) The closed ball of radius $r$ and center $x$ is
$$\overline{B}_r(x) := \{y \in X \mid d(x, y) \leq r\}.$$ 

Example. In $\mathbb{R}^3$, a ball is exactly what the word says — a closed ball containing all of the boundary sphere, and an open ball none of it. In $\mathbb{R}^2$ balls are disks.

The following definition concerns the most important sets for analysis: those which contain everything sufficiently close to any of their elements:

**Definition.** A subset $A$ of $X$ is open in $X$ if for all $x \in A$ there exists $r > 0$ such that $B_r(x) \subset A$.

Open sets include all the open balls, but also much more exotic sets, as indicated in the following proposition; indeed, it’s impossible to visualize all open subsets of $\mathbb{R}^2$, for example — although we’ll see later that the open subsets of $\mathbb{R}$ are just the countable unions of open intervals.

**Proposition 5.7.** The following are open sets in $X$:

(i) $\emptyset$ and $X$;

(ii) any union of open sets in $X$;

(iii) any finite intersection of open sets in $X$;

(iv) any open ball in $X$.

**Proof.**

(i). Suppose $\emptyset$ is not open. Then there exists $x \in \emptyset$ such that $B_r(x) \not\subset \emptyset$ for all $r > 0$. But this contradicts $\emptyset$ being empty. Thus $\emptyset$ must be open.

For the other part, let $x \in X$. Then $B_1(x) \subset X$ by definition. Thus $X$ is open.

(ii). Let $\mathcal{F}$ be a family of open sets in $X$, and let $x \in \bigcup \mathcal{F}$. Then there exits $A \in \mathcal{F}$ such that $x \in A$. Since $A$ is open, there exists $r > 0$ such that $B_r(x) \subset A$. Since $A \subset \bigcup \mathcal{F}$, we get $B_r(x) \subset \bigcup \mathcal{F}$. Thus $\bigcup \mathcal{F}$ is open.

(iii). Let $A_1, \ldots, A_n$ be open in $X$, and let $x \in \bigcap_{i=1}^n A_i$. For each $i = 1, \ldots, n$, pick $r_i > 0$ such that $B_{r_i}(x) \subset A_i$. Then $\{r_1, \ldots, r_n\}$ is a finite subset of $\mathbb{R}$, so it has a minimum; denote it by $r$. Then $r > 0$, and for each $i = 1, \ldots, n$,
$$B_r(x) \subset B_{r_i}(x) \subset A_i,$$

hence $B_r(x) \subset \bigcap_{i=1}^n A_i$. Thus $\bigcap_{i=1}^n A_i$ is open.

(iv). Let $x \in X$, $r > 0$, and $y \in B_r(x)$. Then $s := r - d(x, y) > 0$. Let $z \in B_s(y)$. Then
$$d(z, x) \leq d(z, y) + d(y, x) < s + d(y, x) = r,$$
so \( z \in B_r(x) \). Thus \( B_s(y) \subset B_r(x) \), and we’ve shown \( B_r(x) \) is open. \( \square \)

There’s a subtlety concerning subspaces: when dealing with subspaces, it can be confusing to talk about open sets, because a set which is open in the subspace need not be open in the “ambient” metric space.

**Exercise 5.8.** Let \( A \) be a subspace of a metric space \( X \), and let \( B \subset A \). Prove that \( B \) is open as a subset of the metric space \( A \) if and only if there exists a set \( U \) which is open in \( X \) such that \( B = A \cap U \).

To enhance clarity, it’s occasionally helpful to use the following terminology:

**Definition.** Let \( X \) be a metric space, and let \( B \subset A \subset X \). We say \( B \) is relatively open in \( A \) if it’s open in the subspace \( A \).

Thus relatively open subsets of \( A \) are precisely the intersections with \( A \) of open subsets of \( X \).

**Example.** \([0, 1)\) is relatively open in \([0, 2] \).\(^{35}\)

But of course \([0, 1)\) is not open in \( \mathbb{R} \).

Almost as important as the open sets are the closed ones:

**Definition.** A subset \( A \) of \( X \) is closed in \( X \) if \( A^c \) is open in \( X \).

At first glance it might seem that the above definition is not worthwhile, since closed sets are so trivially related to open ones; however, it turns out that the concept of closedness is extremely handy in its own right.

**Exercise 5.9.** Prove that the set \( \mathbb{Q} \) is neither open nor closed in \( \mathbb{R} \).

**Exercise 5.10.** Prove that the set \( A := \{1/n \mid n \in \mathbb{N}\} \) is not closed in \( \mathbb{R} \).

**Exercise 5.11.** Prove that the set \( B := \{0\} \cup \{1/n \mid n \in \mathbb{N}\} \) is closed in \( \mathbb{R} \). You may not use any fact not yet introduced.

Just as for openness and subspaces, we make the

**Definition.** Let \( X \) be a metric space, and let \( B \subset A \subset X \). We say \( B \) is relatively closed in \( A \) if it’s closed in the subspace \( A \).

And we have:

**Exercise 5.12.** Prove that if \( B \) is a subset of a subspace \( A \) of a metric space \( X \), then \( B \) is relatively closed in \( A \) if and only if there exists a closed set \( C \) in \( X \) such that \( B = A \cap C \).

\(^{35}\)where we are regarding the latter as a subspace of \( \mathbb{R} \) with the usual metric, which is the default assumption unless we say otherwise.
Proposition 5.13. The following are closed sets in $X$:

(i) $\emptyset$ and $X$;
(ii) any intersection of closed sets in $X$;
(iii) any finite union of closed sets in $X$;
(iv) any closed ball in $X$;
(v) any singleton in $X$.

Proof. (i). This follows immediately from $\emptyset^c = X$ and $X^c = \emptyset$.
(ii). Let $F$ be a family of closed sets. Then 
$$\left( \bigcap F \right)^c = \bigcup_{A \in F} A^c$$
is open since each $A^c$ is, so $\bigcap F$ is closed.
(iii). Let $A_1, \ldots, A_n$ be closed. Then 
$$\left( \bigcup_{i=1}^n A_i \right)^c = \bigcap_{i=1}^n A_i^c$$
is open since each $A_i^c$ is, so $\bigcup_{i=1}^n A_i$ is closed.
(iv). Let $x \in X$, $r > 0$, and $y \in \overline{B}_r(x)^c$. Then $s := d(x, y) - r > 0$ since $d(x, y) \leq r$. Let $z \in B_s(y)$. Then 
$$d(z, x) \geq d(x, y) - d(z, y) > d(x, y) - s = r,$$
so $z \in \overline{B}_r(x)^c$. Thus $B_s(y) \subset \overline{B}_r(x)^c$. We’ve shown $\overline{B}_r(x)^c$ is open, hence $\overline{B}_r(x)$ is closed.
(v). Let $x \in X$ and $y \in \{x\}^c$. Then $r := d(x, y) > 0$. Let $z \in B_r(y)$. Then 
$$d(z, x) \geq d(x, y) - d(z, y) > d(x, y) - r > 0,$$
so $z \in \{x\}^c$. Then $B_r(y) \subset \{x\}^c$, so $\{x\}^c$ is open, hence $\{x\}$ is closed.
\qed

Definition. Let $A \subset X$. The **closure** of $A$ is the intersection of all closed sets containing $A$.

Notation and Terminology. $\overline{A}$ denotes the closure of $A$.

Proposition 5.14. Let $A \subset X$. Then:

(i) $\overline{A}$ is a closed set containing $A$, and
(ii) if $B$ is any closed set containing $A$, then $B \supset \overline{A}$.

Proof. (i). Let $\mathcal{C} = \{ C \subset X \mid C$ is closed and $C \supset A \}$. Since each $C \in \mathcal{C}$ is closed and contains $A$, $\overline{A} = \bigcap \mathcal{C}$ is closed and contains $A$.
(ii). If $B$ is closed and contains $A$, then $B \in \mathcal{C}$, hence 
$$B \supset \bigcap \mathcal{C} = \overline{A}. \qed$$
Exercise 5.15. Let $A$ be a subspace of a metric space $X$, and let $B \subseteq A$. Let $\overline{B}$ denote the closure of $B$ in the ambient metric space $X$, and $\overline{B}^A$ the closure of $B$ in the subspace $A$. Prove that $\overline{B}^A = A \cap \overline{B}$.

Definition. With the above notation, we call $\overline{B}^A$ the relative closure of $B$ in $A$.

Exercise 5.16. Find the relative closure of $(1, 2)$ in $(0, 2)$.

Definition. Let $A \subseteq X$. A cluster point of $A$ is an element $t \in X$ such that for all $\epsilon > 0$ there exists $x \in A \setminus \{t\}$ such that $d(x, t) < \epsilon$.

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

Exercise 5.17. Prove that every real number is a cluster point of both $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$.

Exercise 5.18. Write a useful negation of “$t$ is a cluster point of $A$”.

In the definition of cluster point, once we specify $x \in A$ we can satisfy the requirement that $x \neq t$ by requiring $0 < d(x, t)$. Thus we can rephrase the condition as “for all $\epsilon > 0$ there exists $x \in A \setminus \{t\}$ such that $0 < d(x, t) < \epsilon$”.

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 5.19. If $A \subseteq X$ and $t \in A'$, then for all $\epsilon > 0$ the set $B_\epsilon(t) \cap A$ is infinite.

Proof. Arguing the contrapositive, suppose there exists $\epsilon > 0$ such that $B_\epsilon(t) \cap A$ is finite. Put $r = \min \{d(x, t) \mid x \in B_\epsilon(t) \cap A \setminus \{t\}\}$. Then $r > 0$ and $d(x, t) \geq r$ for all $x \in A \setminus \{t\}$, so $t \notin A'$.

Exercise 5.20. Prove that a finite subset of a metric space has no cluster points.

The following result shows that the cluster points are exactly what must be added to a set to get the closure, and also gives a useful test for membership in the closure, resembling the definition of cluster point:
Proposition 5.21. Let $A \subset X$. Then:

(i) $\overline{A} = A \cup A'$;
(ii) $A$ is closed if and only if $A \supset A'$;
(iii) if $x \in X$, then $x \in \overline{A}$ if and only if $B_r(x) \cap A \neq \emptyset$ for all $r > 0$.

Proof. (i). First, to show $\overline{A} \subset A \cup A'$, since $A \subset A \cup A'$ it suffices to show $A \cup A'$ is closed, equivalently its complemen is open. Let $x \in (A \cup A')^c$. Since $x \notin A'$, there exists $r > 0$ such that for all $y \in A$ either $d(y, x) = 0$ or $d(y, x) \geq r$. But $x \notin A$, so for all $y \in A$ we have $d(y, x) \neq 0$. Thus $B_r(x) \cap A = \emptyset$. Claim: $B_r(x) \cap A' = \emptyset$ as well. Let $y \in B_r(x)$. Then there exists $s > 0$ such that $B_s(y) \subset B_r(x)$. Hence $B_s(y) \cap A = \emptyset$, so $y \notin A'$, proving the claim. We’ve shown $(A \cup A')^c$ is open, as desired, so $\overline{A} \subset A \cup A'$.

For the opposite containment $A \cup A' \subset \overline{A}$, since $A \subset \overline{A}$ it suffices to show $A' \subset \overline{A}$. Let $x \notin \overline{A}$. Since $\overline{A}$ is closed, there exists $r > 0$ such that $B_r(x) \cap \overline{A} = \emptyset$. Thus $B_r(x) \cap A = \emptyset$, so $x \notin A'$. Therefore $(\overline{A})^c \subset (A')^c$, so $A' \subset \overline{A}$.

(ii). It’s easy to see that $A$ is closed if and only if $A = \overline{A}$. Since $\overline{A} = A \cup A'$, we have $A = \overline{A}$ if and only if $A' \subset A$.

(iii). First let $x \in \overline{A}$ and $r > 0$. Then $x \in A$ or $x \in A'$. In the first case, $\{x\} \subset B_r(x) \cap A$, so $B_r(x) \cap A \neq \emptyset$. In the second case, $B_r(x) \cap A$ is nonempty since it’s in fact infinite.

For the converse direction, we prove the contrapositive of the desired conditional: if $x \notin \overline{A}$, then the argument from (i) shows $B_r(x) \cap A = \emptyset$ for some $r > 0$. $\square$

Example. By (i) above, we always have $A' \subset \overline{A}$. This inclusion is sometimes proper, and sometimes not: in the metric space $\mathbb{R}$, the finite set $A = \{0\}$ is closed and has no cluster points, so $\overline{A} = A$ and $A' = \emptyset$, while $\mathbb{R}$ is closed, so that $\overline{\mathbb{R}} = \mathbb{R}$, while every real number is a cluster point of $\mathbb{R}$, so $\overline{\mathbb{R}} = \mathbb{R}$.

Of course, we can rephrase the test in (iii) above as: “$x \in \overline{A}$ if and only if for all $r > 0$ there exists $y \in A$ such that $d(x, y) < r$”.

Exercise 5.22. Show that 2 is in the closure of the set $\{(2n + 1)/n \mid n \in \mathbb{N}\}$.

Exercise 5.23. Find $\overline{A}$ and $A'$ if $A = \mathbb{Q} \cap [0, 1]$.

The supremum of a set might not be a cluster point, but it’s always in the closure:

Proposition 5.24. Let $A \subset \mathbb{R}$. If $\sup A$ exists then $\sup A \in \overline{A}$. 
Proof. Our strategy is to show \(\sup A\) satisfies the test in (iii) of the preceding proposition. Let \(r > 0\). Since \(\sup A - r < \sup A\), by Approximation of Suprema there exists \(x \in A\) such that \(\sup A - r < x\). We also have \(x \leq \sup A < \sup A + r\). Thus \(x \in B_r(\sup A) \cap A\), hence \(B_r(\sup A) \cap A \neq \emptyset\).

We know that finite unions of closed sets are closed; do finite unions respect closures and cluster points? The following exercise, together with simple induction argument, shows that the operations \(A \mapsto A\) and \(A \mapsto A'\) both commute with finite unions:

**Exercise 5.25.** Let \(A, B \subset X\). Prove:

(a) \[A \cup B = \overline{A} \cup \overline{B}.\]

(b) \[(A \cup B)' = A' \cup B'.\]

Now for a few more definitions which are frequently handy:

**Definition.** Let \(A \subset X\).

(i) The **interior** of \(A\) is the union of the family of all open subsets of \(X\) which are contained in \(A\).

(ii) The **boundary** of \(A\) is \(A \setminus A^\circ\).

**Notation and Terminology.** \(A^\circ\) denotes the interior of \(A\) and \(\partial A\) the boundary.

An **interior point** of \(A\) is an element of \(A^\circ\), and a **boundary point** an element of \(\partial A\).

**Exercise 5.26.** Prove that \(x \in A^\circ\) if and only if there exists \(r > 0\) such that \(B_r(x) \subset A\).

**Exercise 5.27.** Prove that the interior is the complement of the closure of the complement, i.e.:

\[A^\circ = (\overline{A})^c.\]

**Exercise 5.28.** Prove that \(x \in \partial A\) if and only if for all \(r > 0\) we have both \(B_r(x) \cap A \neq \emptyset\) and \(B_r(x) \setminus A \neq \emptyset\).

**Exercise 5.29.** Prove that

\[\overline{A} = A \cup \partial A.\]

**Definition.** Let \(A \subset X\). Then \(A\) is **dense in** \(X\) if \(\overline{A} = X\).

**Lemma 5.30.** Let \(A \subset X\). Then \(A\) is dense in \(X\) if and only if every nonempty open subset of \(X\) intersects \(A\).
Proof. First assume $A$ is dense in $X$, and let $B$ be a nonempty open subset of $X$. Choose $x \in B$, and then choose $r > 0$ such that $B_r(x) \subset B$. Since $x \in X = \overline{A}$, $B_r(x) \cap A \neq \emptyset$. Thus $B \cap A \neq \emptyset$.

Conversely, assume every nonempty open set intersects $A$. To show $A$ is dense in $X$, it suffices to show that $X \subset \overline{A}$. Let $x \in X$. Then by hypothesis $B_r(x) \cap A \neq \emptyset$ for all $r > 0$, so $x \in \overline{A}$. Thus $X \subset \overline{A}$ as desired, therefore $A$ is dense.

Exercise 5.31. Prove that $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$ are both dense in $\mathbb{R}$. 

\hfill \Box
6. Convergence

A lot of the analysis we eventually want to do involves “continuous” phenomena: limits of functions, derivatives, and integrals. However, it turns out that the foundation of this analysis can be entirely based upon “discrete” phenomena: convergent sequences. In this section we build the general framework of techniques involving sequences, and in the next section we give some tools specifically aimed at sequences of real numbers.

Notation and Terminology. As before, the letter \( X \) will refer to a metric space, unless otherwise specified.

Definition. Let \((x_n)\) be a sequence in \(X\) and \(x \in X\). Then \((x_n)\) converges to \(x\) if for all \(\epsilon > 0\) there exists \(k \in \mathbb{N}\) such that\[
d(x_n, x) < \epsilon \quad \text{for all } n \geq k.
\]

Notation and Terminology. Any of the following mean \((x_n)\) converges to \(x\):

(i) \(x_n \to x\) as \(n \to \infty\)
(ii) \(x_n \xrightarrow{n \to \infty} x\)
(iii) \(x_n \to x\) (if it’s understood that we mean as \(n \to \infty\)),

in which case \(x\) is the limit of \((x_n)\) denoted

\[
x = \lim_{n \to \infty} x_n = \lim x_n.
\]

A sequence is convergent if it converges, otherwise it diverges or is divergent.

There can be at most one \(x\) for which \(x_n \to x\). This result is completely routine, but let’s formalize it to get practice with a typical “\(\epsilon\)-argument”:

Lemma 6.1. Let \((x_n)\) be a sequence in \(X\) and \(x, y \in X\). If \(x_n \to x\) and \(x_n \to y\), then \(x = y\).

Proof. Let \(\epsilon > 0\). Choose \(k \in \mathbb{N}\) such that

\[
d(x_n, x), d(x_n, y) < \epsilon/2 \quad \text{for all } n \geq k.
\]

Then, applying these inequalities with \(n = k\), we have

\[
d(x, y) \leq d(x, x_k) + d(x_k, y) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\]

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

\(^{36}\)and this makes sense since if \(x_n \to x\) then \(x\) is uniquely determined by \((x_n)\), as we’ll show in an elementary lemma
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 \( d(x_n, x), d(x_n, y) < \epsilon/2 \) for all \( n \geq k \)”: Since \( x_n \to x \) we can choose \( k_1 \in \mathbb{N} \) such that \( d(x_n, x) < \epsilon/2 \) for all \( n \geq k_1 \), and similarly we can choose \( k_2 \in \mathbb{N} \) such that \( d(x_n, y) < \epsilon/2 \) for all \( n \geq k_2 \), 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’s normally done without comment, as in the above proof.

Also, the above proof is actually an example of an \( \epsilon/2 \)-argument” — in order to bound a certain quantity by \( \epsilon \), we needed to bound two intermediate quantities by \( \epsilon/2 \). This sort of manipulation will occur a lot in our analysis, and we’ll also see \( \epsilon/3 \)-arguments”, as well as more complicated but related situations.

In the definition of convergence, it’s 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’ll probably have to increase the \( k \).

Also in the definition, we quantified \( n \) by “for all \( n \geq k \); from the context it’s clear that we also restrict \( n \) to be in the domain of the sequence \((x_n)\), namely \( \mathbb{N} \).

Example. Every constant sequence converges to its constant value.

Exercise 6.2. Prove that \( 1/n \to 0 \).

Lemma 6.3. Let \((x_n)\) be a sequence in \( X \) and \( x \in X \). Then \( x_n \to x \) if and only if \( d(x_n, x) \to 0 \).

Proof. Assume \( x_n \to x \). Let \( \epsilon > 0 \). Choose \( k \in \mathbb{N} \) such that \( d(x_n, x) < \epsilon \) for all \( n \geq k \). Since \( |t| = |t| \) for all \( t \in \mathbb{R} \), we have shown that \( d(x_n, x) \to 0 \). The steps are reversible, so the desired equivalence holds.

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

Proof. Let \( \epsilon > 0 \). Choose \( k \in \mathbb{N} \) such that

\[ |x_n - x|, |z_n - x| < \epsilon \]

for all \( n \geq k \).

Then for all \( n \geq k \),

\[ x - \epsilon < x_n \leq y_n \leq z_n < x + \epsilon, \]
so \(|y_n - x| < \epsilon\).

\[\square\]

**Observation.** If \(l \in \mathbb{N}\), it’s pretty obvious that a sequence \((x_n)\) converges if and only if the tail \((x_l, x_{l+1}, x_{l+2}, \ldots)\) does. Thus, for example, in the Squeeze Theorem it’s enough for the inequalities \(x_n \leq y_n \leq z_n\) to hold for large enough \(n\), that is, for there to exist \(l \in \mathbb{N}\) such that the inequalities hold for all \(n \geq l\).

The following exercise concerns sequences in \(\mathbb{R}^n\). We have to be a little careful with the notation here: not only have we frittered away the letter “\(n\)” for the dimension of the Euclidean space, so that it can’t be used as a subscript in sequences, but to make matters worse the notation for the coordinates of an element of \(\mathbb{R}^n\) typically involves subscripts: \(x = (x_1, \ldots, x_n)\). Because of this latter problem, a lot of the time when dealing with sequences in \(\mathbb{R}^n\) it’s handy to switch to *superscript* notation for sequences: \((x^{(k)})\). We put the superscript \(k\) in parentheses as a reminder that something out of the ordinary is afoot — a positive integer superscript could be mistaken for a power if we’re not careful.

**Exercise 6.4.** Let \((x^{(k)}) = ((x_1^{(k)}, \ldots, x_n^{(k)}))\) be a sequence in \(\mathbb{R}^n\) and \(x = (x_1, \ldots, x_n) \in \mathbb{R}^n\). Prove that \(x^{(k)} \rightarrow x\) if and only if \(x_i^{(k)} \xrightarrow{k \to \infty} x_i\) for each \(i = 1, \ldots, n\).

Hint: you might find it convenient to prove that for all \(y \in \mathbb{R}^n\) we have both

\[|y_i| \leq \|y\| \quad \text{for all } i = 1, \ldots, n\]

and

\[\|y\| \leq \sqrt{n}\max_i |y_i|\].

**Definition.**

(i) A subset of \(X\) is *bounded* if it is contained in some ball.

(ii) A function with values in \(X\) is *bounded* if its range is a bounded subset of \(X\).

(iii) More generally, if \(B \subset \text{dom} f\), we say \(f\) is *bounded on \(B\)* if the restriction \(f|B\) is bounded.

In particular, a sequence \((x_n)\) in \(X\) is *bounded* if its range \(\{x_n \mid n \in \mathbb{N}\}\) is bounded in \(X\).

In (i) above it doesn’t matter whether we use open or closed balls, since

\[B_r(x) \subset \overline{B}_r(x) \subset B_{2r}(x)\].

Do we have to pay attention to where the balls are centered? No:
Exercise 6.5. Prove that if \( A \subset X \) and \( X \neq \emptyset \), then the following are equivalent:

(i) \( A \) is bounded;
(ii) there exist \( t \in X \) and \( M \in \mathbb{R} \) such that \( d(x, t) \leq M \) for all \( x \in A \);
(iii) for all \( t \in X \) there exists \( M \in \mathbb{R} \) such that \( d(x, t) \leq M \) for all \( x \in A \);
(iv) there exists \( M \in \mathbb{R} \) such that \( d(x, y) \leq M \) for all \( x, y \in A \).

Observation. A subset of \( \mathbb{R} \) is bounded if and only if it’s both bounded above and bounded below.

In view of the above observation, it makes sense to define corresponding notions for functions, but first another term:

Definition. A function is real-valued if its range is a subset of \( \mathbb{R} \).

Definition. A real-valued function is bounded above if its range is bounded above, and similarly for bounded below.

Thus a real-valued function is bounded if and only if it’s bounded both above and below.

Proposition 6.6. Every convergent sequence is bounded.

Proof. Our strategy is to use the definition of convergence to get all but finitely many terms of the sequence in some ball, then enlarge this ball finitely many times to capture those finitely many terms initially left out.

Let \( x_n \to x \). Since \( 1 > 0 \), by definition of convergence we can choose \( k \in \mathbb{N} \) such that \( d(x, x_n) < 1 \) for all \( n \geq k \). Without loss of generality \( k > 1 \). Put

\[
M = \max\{1, d(x_1, x), \ldots, d(x_{k-1}, x)\}
\]

Then \( d(x_n, x) \leq M \) for all \( n \in \mathbb{N} \). \( \square \)

The following result begins our campaign to show how sequences can form the foundation of analysis:

Proposition 6.7. Let \( A \subset X \) and \( t \in X \). Then:

(i) (Sequential Characterization of Cluster Points) \( t \in A' \) if and only if there exists a sequence in \( A \setminus \{t\} \) converging to \( t \);
(ii) (Sequential Characterization of Closure) $t \in \overline{A}$ if and only if there exists a sequence in $A$ converging to $t$;

(iii) (Sequential Characterization of Closedness) $A$ is closed in $X$ if and only if for every sequence $(x_n)$ in $A$, if $(x_n)$ converges in $X$ then $\lim x_n \in A$.

Proof. (i). First assume $t \in A'$. For each $n \in \mathbb{N}$ choose $x_n \in A$ such that $0 < d(x_n, t) < 1/n$. Then $(x_n)$ is a sequence in $A \setminus \{t\}$. Since $1/n \to 0$, we have $d(x_n, t) \to 0$ by the Squeeze Theorem, hence $x_n \to t$.

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

(ii). First assume $t \in \overline{A}$. If $t \in A$, then the constant sequence $(t, t, \ldots)$ in $A$ converges to $t$. On the other hand, if $t \in A'$, then the preceding part gives a sequence in $A$ converging to $t$.

Conversely, assume $(x_n)$ is a sequence in $A$ converging to $t$, and let $\epsilon > 0$. By convergence, we can choose $x_n \in A$ such that $d(x_n, t) < \epsilon$, and we’ve shown $t \in \overline{A}$.

(iii). First assume $A$ is closed, and let $(x_n)$ be a sequence in $A$ which converges in $X$. Then $\lim x_n \in \overline{A}$ by the preceding part. But $\overline{A} = A$ since $A$ is closed. Thus $\lim x_n \in A$.

Conversely, assume the condition regarding sequences. Let $x \in \overline{A}$. Then there exists a sequence $(x_n)$ in $A$ converging to $x$. By hypothesis, $x \in A$. We’ve shown $\overline{A} \subseteq A$, hence $A = \overline{A}$, so $A$ is closed. $\square$

Given one sequence, the following definition shows how to get lots more:

**Definition.** For sequences $(x_n)$ and $(y_k)$, we say $(y_k)$ is a subsequence of $(x_n)$ if there exists a strictly increasing sequence $(n_k)$ in $\mathbb{N}$ such that

$$y_k = x_{n_k} \quad \text{for all } k \in \mathbb{N}.$$ 

Thus, a subsequence of $(x_n)$ is a composition of the function $n \mapsto x_n : \mathbb{N} \to X$ with a strictly increasing function $k \mapsto n_k : \mathbb{N} \to \mathbb{N}$.

**Observation.** In the above definition, the requirement that $(n_k)$ be strictly increasing implies that $n_k \geq k$ for all $k \in \mathbb{N}$.

How do we get a subsequence of $(x_n)$? The definition tells us to pick a strictly increasing sequence $(n_k)$ in $\mathbb{N}$. There are a couple of subtle points here: First of all, of course the range $\{n_k \mid k \in \mathbb{N}\}$ is an infinite subset of $\mathbb{N}$. Conversely, Lemma 4.12 tells us that every infinite subset of $\mathbb{N}$ is the range of a strictly increasing sequence, in fact it’s easy to see that this strictly increasing sequence is unique. Thus we get every
subsequence of \((x_n)\) by restricting to an infinite subset of \(\mathbb{N}\). How many different subsequences do we get?

Here comes the second subtle point: although there are uncountably many infinite subsets of \(\mathbb{N}\), different infinite subsets might give the same subsequence, depending upon the values of \((x_n)\). Here’s an extreme case:

**Example.** If \((x_n)\) is a constant sequence, then for any strictly increasing sequence \((n_k)\) in \(\mathbb{N}\) we have

\[ x_k = x_{n_k} \quad \text{for all } k \in \mathbb{N}. \]

Conclusion: a constant sequence has only one subsequence.

To summarize part of the above discussion: specifying a subsequence of \((x_n)\) is equivalent to restricting to an infinite subset of \(\mathbb{N}\), and moreover, different subsets might give the same subsequence.

This being said, a moment’s reflection reveals that in general a sequence can conceivably have a bewildering profusion of subsequences; certainly there is in general no simple way to write them all down. Nevertheless, there is some commonality among them:

**Lemma 6.8.** Let \((x_n)\) be a sequence in \(X\) and \(x \in X\). If \((x_n)\) converges to \(x\) then every subsequence of \((x_n)\) also converges to \(x\).

Proof. Let \(x_n \to x\), and let \((n_k)\) be a strictly increasing sequence in \(\mathbb{N}\). Let \(\epsilon > 0\). Choose \(j \in \mathbb{N}\) such that \(d(x_n, x) < \epsilon\) for all \(n \geq j\). Then for all \(k \geq j\) we have \(n_k \geq k \geq j\), hence \(d(x_{n_k}, x) < \epsilon\). Thus \(x_{n_k} \to x\). \(\square\)

What if \((x_n)\) does not converge to \(x\)? This also has an interesting subsequential interpretation:

**Lemma 6.9.** Let \((x_n)\) be a sequence in \(X\) and \(x \in X\). If \((x_n)\) does not converge to \(x\) then there exists \(\epsilon > 0\) and a subsequence of \((x_n)\) in \(X \setminus B_{\epsilon}(x)\).

Proof. Since \(x_n \not\to x\), there exists \(\epsilon > 0\) such that for all \(k \in \mathbb{N}\) there exists \(n \geq k\) such that \(d(x_n, x) \geq \epsilon\). Thus \(\{n \in \mathbb{N} \mid d(x_n, x) \geq \epsilon\}\) is infinite, giving a subsequence in \(X \setminus B_{\epsilon}(x)\). \(\square\)

Another way to think of the above conclusion is that \((x_n)\) has a subsequence no subsequence of which converges to \(x\).

\((x_n)\) may not converge, but it may have many convergent subsequences. It’s useful to consider their limits:

\(38\)Remember that a subsequence is a sequence, which is a function, so all it knows about is what it’s values are. The subsequence doesn’t know anything about how it came to be — subsequences don’t study philosophy!
**Definition.** A *subsequential limit* of \((x_n)\) is a limit of a convergent subsequence.

A sequence might have many subsequential limits (in fact, uncountably many) — the following lemma tells us exactly what they are:

**Lemma 6.10.** Let \((x_n)\) be a sequence in \(X\) and \(x \in X\). Then the following are equivalent:

(i) \(x\) is a subsequential limit of \((x_n)\);
(ii) either \(x\) is a cluster point of the range \(\{x_n \mid n \in \mathbb{N}\}\) or the set \(\{n \in \mathbb{N} \mid x_n = x\}\) is infinite;
(iii) for all \(\epsilon > 0\) and \(k \in \mathbb{N}\) there exists \(n > k\) such that \(d(x_n, x) < \epsilon\).

**Proof.** Throughout this proof let \(A = \text{ran}(x_n)\).

(i) implies (ii). Let \((x_{n_k})\) be a subsequence of \((x_n)\) converging to \(x\). We must show that either \(x\) is a cluster point of \(A\) or the set \(\{n \in \mathbb{N} \mid x_n = x\}\) is infinite. Equivalently, if we assume \(x \notin A'\), we must show that \(x\) is a cluster point of \(A\).

(ii) implies (iii). Let \(\epsilon > 0\) and \(k \in \mathbb{N}\). There are two cases to consider:

Case 1. \(x \in A'\). Then \(B_\epsilon(x) \cap A\) is infinite, so the set \(\{n \in \mathbb{N} \mid d(x_n, x) < \epsilon\}\) is infinite, hence contains some an element greater than \(k\).

Case 2. The set \(\{n \in \mathbb{N} \mid x_n = x\}\) is infinite. Then there exists \(n > k\) such that \(x_n = x\), hence \(d(x_n, x) < \epsilon\).

(iii) implies (i). First, applying the hypothesis with \(\epsilon = k = 1\), we can choose \(n_1 > 1\) such that \(d(x_{n_1}, x) < 1\). Then for each \(j = 2, 3, \ldots\) we can inductively apply the hypothesis with \(\epsilon = 1/j\) and \(k = n_{j-1}\) to choose \(n_j > n_{j-1}\) such that \(d(x_{n_j}, x) < 1/j\). We have defined a strictly increasing sequence \((n_j)\) in \(\mathbb{N}\) such that \(d(x_{n_j}, x) < 1/j\) for all \(j \in \mathbb{N}\). Since \(1/j \to 0\), by the Squeeze Theorem we have \(d(x_{n_j}, x) \to 0\), hence \(x_{n_j} \to x\). \(\square\)

**Exercise 6.11.** Find two different subsequential limits of \((-1)^n\), and prove that you are right.

---

39If \(x\) is a cluster point of the set \(A := \{x_n \mid n \in \mathbb{N}\}\), it might seem that Lemma 6.7 already told us that there is a subsequence of \((x_n)\) converging to \(x\), but this is wrong: not only does it not tell us this, but in fact Lemma 6.7 gives us no information about subsequences: 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 \((x_n)\).
Exercise 6.12. Prove that there exists a real sequence having every real number as a subsequential limit.

Exercise 6.13. Let \((x_n)\) be a sequence in a metric space \(X\), and let \(x \in X\). Assume \(x_n \to x\). Use the preceding lemma to prove that the subspace
\[ A := \{x\} \cup \{x_n \mid n \in \mathbb{N}\} \]
is closed in \(X\). Hint: prove that \(x\) is the only possible cluster point of \(A\).
7. Sequences of real numbers

In this section we discuss phenomena specifically germane to real sequences.

Of course, just as in any metric space, a convergent real sequence must be bounded. Although it can be difficult to determine whether a bounded real sequence converges, sometimes it’s easy:

**Exercise 7.1.** Prove that a bounded increasing sequence \((x_n)\) converges to the supremum of its range.

Of course, there is a corresponding result for decreasing sequences, and it’s handy to record for easy reference the following consequence:

**Corollary 7.2.** Every bounded monotone sequence converges.

**Exercise 7.3.** Find the sup and inf of the set \(A = \left\{ (-1)^n n/(n + 1) \mid n \in \mathbb{N} \right\} \).

Monotone sequences are very special, however all real sequences contain them:

**Proposition 7.4.** Every real sequence has a monotone subsequence.

**Proof.** Let \((x_n)\) be a real sequence. We introduce an auxiliary object: put

\[ A = \left\{ k \in \mathbb{N} \mid x_k = \max \{ x_n \mid n \geq k \} \right\}. \]

There are two possibilities for \(A\):

**Case 1.** \(A\) is finite. Then we can choose \(l \in \mathbb{N}\) such that \(n < l\) for all \(n \in A\). By construction, for all \(n \geq l\) there exists \(j > n\) such that \(x_n < x_j\). Define \(n_1 = l\), and for \(k > 1\) inductively choose \(n_k > n_{k-1}\) such that \(x_{n_{k-1}} < x_{n_k}\). Then the subsequence \((x_{n_k})\) is increasing.

**Case 2.** \(A\) is infinite. We can restrict the sequence \((x_n)\) to \(A\) to get a subsequence, and by construction of \(A\) this subsequence is decreasing, since if \(k < l\) in \(A\) then

\[ x_k = \max \{ x_n \mid n \geq k \} \geq \max \{ x_n \mid n \geq l \} = x_l. \]

\[ \square \]

It’s no exaggeration to say that the following easy corollary is one of the foundations of analysis:

**Bolzano-Weierstrass Theorem.** Every bounded real sequence has a convergent subsequence.

**Proof.** Let \((x_n)\) be a bounded real sequence. By the preceding result, \((x_n)\) has a monotone subsequence. Since every bounded monotone sequence converges, we are done. \[ \square \]
We can rephrase the above theorem as: “every bounded real sequence has a subsequential limit”.

**Exercise 7.5.** Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded function. Prove that there exists a strictly increasing sequence $(x_n)$ in $\mathbb{R}$ such that the sequence $(f(x_n))$ converges.

The following lemma is a technical tool which makes some proofs easier:

**Lemma 7.6.** For all real sequences $(x_n)$ and $(y_n)$, if $x_n \to 0$ and $(y_n)$ is bounded, then $x_n y_n \to 0$.

**Proof.** Since $(y_n)$ is bounded, we can choose $M \in \mathbb{R}$ such that $|y_n| \leq M$ for all $n \in \mathbb{N}$. Without loss of generality $M > 0$. Let $\epsilon > 0$. Choose $k \in \mathbb{N}$ such that $|x_n| < \epsilon/M$ for all $n \geq k$. Then

$$|x_n y_n| = |x_n||y_n| \leq M|x_n| < \epsilon$$

for all $n \geq k$. □

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’s an upper bound for $\{|y_n| \mid n \in \mathbb{N}\}$; 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

$$|x_n y_n| \leq M|x_n| < \frac{M\epsilon}{M + 1} < \epsilon.$$

The following result gives the familiar arithmetic of limits for sequences:

**Arithmetic of Convergence.** For all convergent real 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). Let $\epsilon > 0$. Choose $k \in \mathbb{N}$ such that

$$|x_n - x|, |y_n - y| < \frac{\epsilon}{2} \quad \text{for all } n \geq k.$$
Then

\[ |(x_n + y_n) - (x + y)| \leq |x_n - x| + |y_n - y| < \epsilon \quad \text{for all } n \geq k. \]

(ii). We have

\[ x_ny_n - xy = x_ny_n - xy_n + xy_n - xy = (x_n - x)y_n + x(y_n - y) \to 0, \]

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

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

(iv). We have

\[ \frac{x_n}{y_n} - \frac{x}{y} = \frac{1}{y_n} \left( x_n - \frac{x y_n}{y} \right), \]

and the stuff inside the parentheses goes to 0 by Parts (i) and (iii). Thus it suffices to show the sequence \( \frac{1}{y_n} \) has a bounded tail: there exists \( k \in \mathbb{N} \) such that for all \( n \geq k \) we have \( |y_n - y| < |y|/2 \), so \( |y_n| \geq |y|/2 \), hence

\[ \left| \frac{1}{y_n} \right| \leq \frac{2}{|y|}. \]

A couple of times in the above proof we used the shortcut of choosing something to perform several consistent jobs at once. In particular, for (iv) we chose \( k \) such that both \( k \geq l \) and \( |y_n - y| < |y|/2 \) for all \( n \geq l \).

The following result shows that limits preserve weak inequalities:

**Proposition 7.7.** For all convergent real 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 \( z \geq 0 \). But \( [0, \infty) \) is closed, so since \( z_n \in [0, \infty) \) for all \( n \) we must have \( \lim z_n \in [0, \infty) \).

The following result gives a kind of “continuity” of the metric function (see Lecture 10 for the definition of continuity):

**Lemma 7.8.** For all sequences \( (x_n) \) and \( (y_n) \) in a metric space \( X \), if \( x_n \to x \) and \( y_n \to y \), then \( d(x_n, y_n) \to d(x, y) \).

**Proof.** We have

\[ |d(x_n, y_n) - d(x, y)| = |d(x_n, y_n) - d(x, y_n) + d(x, y_n) - d(x, y)| \]
\[ \leq |d(x_n, y_n) - d(x, y_n)| + |d(x, y_n) - d(x, y)| \]
\[ \leq d(x_n, x) + d(y_n, y). \]

Since \( x_n \to x \), we have \( d(x_n, x) \to 0 \), and similarly \( d(y_n, y) \to 0 \), hence \( d(x_n, x) + d(y_n, y) \to 0 \). Thus \( |d(x_n, y_n) - d(x, y)| \to 0 \) by the Squeeze Theorem, therefore \( d(x_n, y_n) \to d(x, y) \).
In the above proof we got to use the Alternate Triangle Inequality twice.

**Infinite limits**

Real sequences can diverge in several ways. The following definition describes two of these ways which turn out to be particularly common:

**Definition.** Let \((x_n)\) be a real sequence.

(i) \((x_n)\) goes to \(\infty\) if for all \(z \in \mathbb{R}\) there exists \(k \in \mathbb{N}\) such that
\[
x_n > z \quad \text{for all } n \geq k.
\]

(ii) \((x_n)\) goes to \(-\infty\) if for all \(z \in \mathbb{R}\) there exists \(k \in \mathbb{N}\) such that
\[
x_n < z \quad \text{for all } n \geq k.
\]

**Notation and Terminology.** Our notation for “infinite limits” is similar to that for finite limits; for example, we write \(x_n \to \infty\) or \(\lim x_n = \infty\).

Occasionally it’s convenient to write \(+\infty\) for \(\infty\). For example, both \(\infty\) and \(-\infty\) can be indicated at the same time by \(\pm\infty\).

However, we have to be extremely careful not to confuse “infinite limits” with “ordinary limits” — if \(\lim x_n = \infty\) then the sequence \((x_n)\) diverges! This is one of those circumstances where the notation and terminology could perhaps be made more logical and less prone to misinterpretation,\(^{41}\) but it’s so ingrained by now that we can’t do anything about it. Anyway, it would be unwise to say something like “\((x_n)\) has a limit” if in fact the situation is \(x_n \to \pm\infty\).

**Examples.**

(i) \(n \to \infty\) by the Archimedean Principle.

(ii) If \(a > 1\), Exercise 3.15 tells us that \(a^n \to \infty\).

**Exercise 7.9.** Give examples of divergent real sequences \((x_n)\) for which \(x_n \not\to \infty\) and \(x_n \not\to -\infty\) and which are:

(a) bounded;

(b) unbounded.

**Exercise 7.10.** Give an example of a real sequence \((x_n)\) such that both:

\(^{40}\)Note that in this second part we do a familiar thing: reverse all inequalities to get an analogous concept.

\(^{41}\)In fact, this is a particularly galling instance — I’m sorely tempted to eschew writing \(\lim x_n = \infty\) altogether, but it would cause some inconvenience down the road!
(i) \( x_n \to \infty \), and

(ii) for every \( k \in \mathbb{N} \) the sequence \( (x_k, x_{k+1}, \ldots) \) is not monotone increasing.

The following result gives a way to turn infinite limits into ordinary ones:

**Proposition 7.11.** For any real sequence \( (x_n) \), we have \( x_n \to \infty \) if and only if there exists \( k \in \mathbb{N} \) such that both

(i) \( x_n > 0 \) for all \( n \geq k \), and

(ii) the tail \( (1/x_n)_{n \geq k} \) converges to 0.

Similarly for \( -\infty \).

**Proof.** We only give the argument for the case \( \infty \); the case \( -\infty \) is similar. Since the condition (i) explicitly requires all but finitely many terms to be positive, and since the condition \( x_n \to \infty \) implies the same property, and furthermore since the important properties of the sequence \( (x_n) \) don’t change if we delete finitely many terms, without loss of generality \( x_n > 0 \) for all \( n \in \mathbb{N} \).

First assume \( x_n \to \infty \). Let \( \epsilon > 0 \). Then we can choose \( l \in \mathbb{N} \) such that for all \( n \geq l \) we have \( x_n > 1/\epsilon \), hence \( 1/x_n < \epsilon \).

Conversely, assume \( 1/x_n \to 0 \). Let \( z \in \mathbb{R} \). We want to show that the terms are eventually bigger than \( z \), so without loss of generality \( z > 0 \). Then we can choose \( l \in \mathbb{N} \) such that for all \( n \geq l \) we have \( 1/x_n < 1/z \), hence \( x_n > z \). \( \Box \)

The following result gives a sort of limited “arithmetic” for infinite limits:

**Arithmetic of Infinite Limits.** Let \( (x_n) \) and \( (y_n) \) be real sequences, and assume \( x_n \to \infty \). Then:

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

(ii) \( x_n y_n \to \infty \) if \( \inf y_n > 0 \);

(iii) \( y_n/x_n \to 0 \) if \( (y_n) \) is bounded and \( x_n \neq 0 \) for all \( n \in \mathbb{N} \).

**Proof.** (i). Since \( (y_n) \) is bounded below, we can choose \( t \in \mathbb{R} \) such that \( y_n \geq t \) for all \( n \in \mathbb{N} \). Let \( z \in \mathbb{R} \). Then \( z - t \in \mathbb{R} \), so we can choose \( k \in \mathbb{N} \) such that for all \( n \geq k \) we have \( x_n > z - t \), hence

\[
x_n + y_n \geq x_n + t > z.
\]

(ii). Since \( \inf y_n > 0 \), we can choose \( t > 0 \) such that \( y_n \geq t \) for all \( n \in \mathbb{N} \). Let \( z \in \mathbb{R} \). Then \( z/t \in \mathbb{R} \), so we can choose \( k \in \mathbb{N} \) such that for all \( n \geq k \) we have \( x_n > z/t \), hence

\[
x_n y_n \geq x_n t > z.
\]
(iii). We don’t have to do this one “bare-hands”: by the preceding proposition, $1/x_n \to 0$. Since $(y_n)$ is bounded we have
\[ \frac{y_n}{x_n} = y_n \left( \frac{1}{x_n} \right) \to 0. \]
\[ \square \]

Later we could slightly improve the hypothesis of (ii) to “$\lim y_n > 0$”.

**Example.** If $|x| < 1$ then $x^n \to 0$. Of course this is trivial if $x = 0$, and if $x \neq 0$ then $1/|x| > 1$, so
\[ \left( \frac{1}{|x|} \right)^n \to \infty. \]

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

Many results for finite limits have true analogues for infinite limits. For example:

**Exercise 7.12.** Prove the Squeeze Theorem for Infinite Limits: If $(x_n)$ and $(y_n)$ are real sequences such that $x_n \to \infty$ and $y_n \geq x_n$ for all $n \in \mathbb{N}$, then $y_n \to \infty$.

Actually, as usual it suffices to know that there exists $k \in \mathbb{N}$ such that $y_n \geq x_n$ for all $n \geq k$, because divergence to $\infty$ is not destroyed by modifying finitely many terms.

**Exercise 7.13.** Let $A \subset \mathbb{R}$. Prove that $A$ is unbounded above if and only if there exists a sequence $(x_n)$ in $A$ such that $x_n \to \infty$.

**Observation.** Related to the above exercise, if $A$ is bounded above we already know that $A$ contains a sequence converging to $\sup A$, since $\sup A \in A$. 
8. Lim sup and lim inf

Bounded sequences

First we’ll discuss lim sup and lim inf for bounded real sequences \((x_n)\). Recall that \((x_n)\) can have various subsequential limits. It turns out that the set of subsequential limits of \((x_n)\) can be quite complicated, but the lim sup and lim inf will pick out the largest and smallest.

Lim sups and lim infs possess a limited (!) subset of the properties of limits, but have the advantage that they always exist, whereas of course a given sequence might not have a limit.

**Definition.** Let \((x_n)\) be a bounded real sequence, and let \(S\) denote the set of subsequential limits of \((x_n)\). The **lim sup** and **lim inf** of \((x_n)\) are defined as

\[
\limsup x_n := \sup S \\
\liminf x_n := \inf S,
\]

respectively.

Of course, we know the above set \(S\) is nonempty by the Bolzano-Weierstrass Theorem, and it’s bounded because if \(|x_n| \leq M\) for all \(n \in \mathbb{N}\) then every subsequential limit will satisfy the same inequality. Thus the lim sup and lim inf both exist.

In fact, the above set \(S\) has both a max and a min, so the lim sup and lim inf are actually the largest and smallest subsequential limits:

**Proposition 8.1.** Let \((x_n)\) be a bounded real sequence. Then \(\limsup x_n\) and \(\liminf x_n\) are subsequential limits of \((x_n)\).

**Proof.** By symmetry (see the following exercise), it suffices to show it for the lim sup. For convenience in this proof, let \(S\) denote the set of subsequential limits of \((x_n)\), and put

\[s = \limsup x_n.\]

We argue by contradiction: suppose \(s \notin S\). Then there exists \(\epsilon > 0\) and \(k \in \mathbb{N}\) such that

\[|x_n - s| \geq \epsilon \quad \text{for all } n > k.\]

Since \(s - \epsilon < s\), by Approximation of Suprema there exists \(t \in S\) such that \(t > s - \epsilon\). Put \(\delta = t - s + \epsilon\). Then \(\delta > 0\) and

\[(t - \delta, t + \delta) \subset (s - \epsilon, s + \epsilon).\]
Since $t \in S$, there exists $n > k$ such that 

$$|x_n - t| < \delta.$$ 

But then 

$$|x_n - s| < \epsilon,$$ 

a contradiction. \hfill \Box

The result contained in the following exercise shows that from every fact concerning lim sup we can more-or-less automatically deduce a corresponding fact for lim inf:

**Exercise 8.2.** Prove that for any bounded real sequence $(x_n)$,

$$\lim x_n = -\lim (-x_n).$$

The following result shows how lim sup and lim inf pick out the real sequences which converge:

**Corollary 8.3.** A bounded real sequence $(x_n)$ converges if and only if $\lim x_n = \lim x_n$.

**Proof.** First assume $(x_n)$ converges, with limit $x$. Then every subsequence of $(x_n)$ also converges to $x$, so $\lim x_n = \lim x_n = x$ by the preceding proposition.

Conversely, assume $(x_n)$ diverges. Then in particular $x_n \not\to \lim x_n$. Thus there exists $\epsilon > 0$ and a subsequence $(y_k)$ of $(x_n)$ such that

$$|y_k - \lim x_n| \geq \epsilon \quad \text{for all } k \in \mathbb{N}.$$ 

Then $(y_k)$ is a bounded real sequence, hence has a subsequential limit $y$ by the Bolzano-Weierstrass Theorem. The above inequalities imply $|y - \lim x_n| \geq \epsilon$. Thus $y \neq \lim x_n$. Since $y$ and $\lim x_n$ are both subsequential limits of $(x_n)$, we conclude $\lim x_n \neq \lim x_n$. \hfill \Box

No matter how complicated the set of subsequential limits of a bounded real sequence gets, the lim sup and lim inf tell us what the largest and smallest ones are. When there are only finitely many subsequential limits, the situation can be handled in a very elementary way; a special case is illustrated by the following exercise:

**Exercise 8.4.** Let $(x_n)$ be a bounded real sequence, and let $A$ and $B$ be disjoint infinite subsets of $\mathbb{N}$ with union $\mathbb{N}$. Assume that the restrictions of the sequence $(x_n)$ to $A$ and $B$ converge to $x$ and $y$, respectively. Prove that $x$ and $y$ are the only subsequential limits of $(x_n)$. 
The above exercise can be generalized in an obvious way to finitely many subsequences which “exhaust” \((x_n)\).

Here’s an opportunity to apply this useful device:

**Exercise 8.5.** Define a sequence \((x_n)\) by
\[
x_n = (-1)^n + \frac{1}{n}.
\]
Find \(\lim x_n\) and \(\lim x_n\).

The following exercise gives a limited sort of “arithmetic” with \(\lim\) sup (and similar results could be proven for \(\lim\) inf):

**Exercise 8.6.** Let \((x_n)\) and \((y_n)\) be real sequences. Assume \(x_n \to x\) and \((y_n)\) is bounded. Prove:
\[
\begin{align*}
(a) \quad \lim(x_n + y_n) &= x + \lim y_n; \\
(b) \quad \lim(x_n y_n) &= x \lim y_n \text{ if } x > 0.
\end{align*}
\]

We know that if \(t > \lim x_n\) then \(t\) is not a subsequential limit of \((x_n)\), so there exists \(\epsilon > 0\) and a subsequence of \((x_n)\) lying completely outside the open interval \((t - \epsilon, t + \epsilon)\). In fact, it follows from the following technical lemma that something more is true:

**Lemma 8.7.** Let \((x_n)\) be a bounded real sequence and \(a > \lim x_n\). Then there exists \(k \in \mathbb{N}\) such that \(x_n < a\) for all \(n \geq k\).

**Proof.** No subsequence of \((x_n)\) can lie in \([a, \infty)\), for such a subsequence, hence \((x_n)\) itself, would have a subsequential limit in \([a, \infty)\), which is absurd because every subsequential limit of \((x_n)\) is in \([\lim x_n, \lim x_n]\). Thus only finitely many terms of \((x_n)\) can lie in \([a, \infty)\), and the result follows. \(\square\)

**Exercise 8.8.** State the corresponding result for \(\lim\) inf.

The following exercise shows that \(\lim\) sup and \(\lim\) inf preserve weak inequalities:

**Exercise 8.9.** Let \((x_n)\) and \((y_n)\) be bounded sequences such that \(x_n \leq y_n\) for all \(n\) sufficiently large. Prove:
\[
\begin{align*}
(a) \quad \lim x_n &\leq \lim y_n; \\
(b) \quad \lim x_n &\leq \lim y_n.
\end{align*}
\]

Hint: use the above lemma.

**Exercise 8.10.** Let \(\{x_n\}\) be a bounded sequence of real numbers. For each \(n \in \mathbb{N}\) put \(a_n = \sup\{x_k | k \geq n\}\). Prove that the sequence \(\{a_n\}\) is decreasing and \(\lim x_n = \lim a_n\).
Exercise 8.11. Let \((x_n)\) be a bounded sequence of real numbers, and define a new sequence \((\sigma_n)\) by

\[
\sigma_n = \frac{1}{n} \sum_{i=1}^{n} x_i.
\]

Show that \(\lim \sigma_n \leq \lim sup x_n\). Hint: break the sum into 2 pieces, one with a fixed number of terms and the other with all \(x_n\)'s not much bigger than the lim sup.

Unbounded sequences

It turns out to occasionally be convenient to have lim sup and lim inf available for arbitrary real sequences, bounded or unbounded. We now show how this works for unbounded sequences.

Exercise 8.12. Let \((x_n)\) is a real sequence which is unbounded above. Prove that some subsequence diverges to \(\infty\).

In view of the above exercise, we make the

Definition. If \((x_n)\) is a real sequence which is unbounded above, we define

\[
\lim x_n = \infty.
\]

What if \((x_n)\) is bounded above? Then if there are any subsequential limits at all, it turns out that there is a largest one, so in this situation we make the

Definition. Let \((x_n)\) be a real sequence which is bounded above, and let \(S\) denote the set of subsequential limits of \((x_n)\). If \(S \neq \emptyset\), we define

\[
\lim x_n = \max S.
\]

The only other case is covered by the following exercise:

Exercise 8.13. Let \((x_n)\) be a real sequence which is bounded above. Prove that if \((x_n)\) has no subsequential limits, then in fact \(x_n \to -\infty\).

In view of the above exercise, we make the

Definition. Let \((x_n)\) be a real sequence which is bounded above. If \(x_n \to -\infty\), we define

\[
\lim x_n = -\infty.
\]

A similar analysis leads to the corresponding definition for lim inf:
Definition. Let \((x_n)\) be a real sequence, and let \(S\) denote the set of subsequential limits of \((x_n)\). We define

\[
\lim x_n = \begin{cases} 
\min S & \text{if } (x_n) \text{ is bounded below and } x_n \not\to \infty \\
\infty & \text{if } x_n \to \infty \\
-\infty & \text{if } (x_n) \text{ is unbounded below.}
\end{cases}
\]

We have now defined \(\lim\) sup and \(\lim\) inf for arbitrary real sequences. All the results we proved for \(\lim\) sup and \(\lim\) inf of bounded sequences have true analogues in the unbounded cases, although sometimes the statement must be formulated with care, and the unbounded versions are not used as often. However, for example it’s sometimes useful to keep in mind that \(\lim x_n = \infty\) if and only if \((x_n)\) has a subsequence diverging to \(\infty\), and similarly for \(\lim x_n = -\infty\).

Exercise 8.14. (a) Does the sequence \((-1)^n\) have a convergent subsequence? Why or why not?
(b) Find \(\lim(-1)^n\) and \(\lim(-1)^n\).
(c) Give an example of a sequence \((x_n)\) such that \(\lim x_n = 1\) and \(\lim x_n = 0\).
9. Limits

In preceding sections we’ve studied limits of sequences. Now we look at more general limits. Roughly speaking, we’re making the transition from the discrete to the (possibly) continuous.

**Notation and Terminology.** As before, letters such as $X$, $Y$, and $Z$ will refer to metric spaces, unless otherwise specified.

**Definition.** Let $A \subset X$, $f : A \to Y$, $t \in A'$, and $u \in Y$. $f(x)$ goes to $u$ as $x$ goes to $t$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that $d(f(x), u) < \epsilon$ for all $x \in A \cap B_\delta(t) \setminus \{t\}$.

**Notation and Terminology.** Both

(i) $f(x) \to u$ as $x \to t$ and

(ii) $f(x) \xrightarrow{t} u$

mean $f(x)$ goes to $u$ as $x$ goes to $t$, in which case $u$ is the limit of $f$ at $t$, denoted $\lim_{x \to t} f(x)$.

It’s important to remember that when we let $x \to t$ we don’t allow $x = t$.

**Exercise 9.1.** Prove that:

(a) $\lim_{x \to t} x = t$;

(b) $\lim_{x \to t} c = c$.

The above definition of limit is a little more complicated than the definition of limit of a sequence, mainly because we also have to keep track of the $A$ and the $t$. These two kinds of limit are disjoint in the sense that a limit of a sequence is never a special case of a limit in the above sense (and vice-versa), because, although a sequence is a function, it’s domain is $\mathbb{N}$, which has no cluster points. That being said, later we’ll see “limit of $f$ at infinity”, which actually does include limits of sequences as a special case.

However, it might be surprising that the entire theory of limits can be based on sequences:

**Sequential Characterization of Limits.** Let $A \subset X$, $f : A \to Y$, 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 $\lim_{x \to t} f(x) = \lim f(x_n)$.

\[42\text{and this makes sense since } u \text{ is unique if it exists}\]
Proof. First assume \( \lim_{x \to t} f(x) = u \), and let \( (x_n) \) be a sequence in \( A \setminus \{t\} \) converging to \( t \). Let \( \epsilon > 0 \). Choose \( \delta > 0 \) such that \( d(f(x), u) < \epsilon \) for all \( x \in A \cap B_\delta(t) \setminus \{t\} \). Since \( x_n \to t \) we can choose \( k \in \mathbb{N} \) such that for all \( n \geq k \) we have \( d(x_n, t) < \delta \), hence \( d(f(x_n), u) < \epsilon \) (because \( x_n \neq t \)). Therefore \( f(x_n) \to u \).

Conversely, assume the condition regarding sequences. Our first obstacle is that we haven’t required that all the sequences \( (f(x_n)) \) converge to the same thing; fortunately, this is automatic: 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 \( z_k \to t \) (because both complementary subsequences \( (z_{2n-1}) \) and \( (z_{2n}) \) do), so \( \lim_{k \to \infty} f(z_k) \) exists by hypothesis. Since \( (f(x_n)) \) and \( (f(x'_n)) \) are both subsequences of \( (f(z_k)) \), they have the same limit.

Consequently, we can define \( u \) 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 \( \lim_{x \to t} f(x) = u \). Suppose not. Then there exists \( \epsilon > 0 \) such that for all \( \delta > 0 \) there exists \( x \in A \setminus \{t\} \) such that

\[
d(x, t) < \delta \quad \text{and} \quad d(f(x), u) \geq \epsilon.
\]

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

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

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

The above result had a fairly fussy statement, because we wanted the extra power of not having to know in advance that all the sequences \( (f(x_n)) \) had the same limit — as we took pains to point out in the proof, it’s enough to know that they all converge. Here’s a useful variation:

**Observation.** It follows from the above argument that if there exist sequences \( (x_n) \) and \( (y_n) \) in \( A \setminus \{t\} \) converging to \( t \) such that \( \lim f(x_n) \neq \lim f(y_n) \), then \( \lim_{x \to t} f(x) \) does not exist.

**Exercise 9.2.** Show that \( \lim_{x \to 0} \sin(1/x) \) does not exist.

We’ll use sequences to develop the elementary theory of limits. But first a little notation:

**Notation and Terminology.** If \( f, g : A \to \mathbb{R} \), we write \( f \leq g \) if \( f(x) \leq g(x) \) for all \( x \in A \), and similarly for \( \geq, <, \) and \( > \).
**Squeeze Theorem for Limits.** Let $A \subset X$, $f, g, h: A \to \mathbb{R}$, and $t \in A'$. If $f \leq g \leq h$ and
\[ \lim_{x \to t} f(x) = \lim_{x \to t} h(x) = u, \]
then $\lim_{x \to t} g(x) = u$.

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 $\lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} h(x_n) = u$, so we have $g(x_n) \to u$. Again by the Sequential Characterization of Limits, we conclude that $\lim_{x \to t} g(x) = u$. \qed

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’ll omit the details.

**Proposition 9.3.** Let $A \subset X$, $f = (f_1, \ldots, f_n): A \to \mathbb{R}^n$, $t \in A'$, and $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$. Then $\lim_{x \to t} f(x) = u$ if and only if $\lim_{x \to t} f_i(x) = u_i$ for each $i = 1, \ldots, n$.

Proof. Immediate from the Sequential Characterization of Limits and the corresponding result for sequences. \qed

**Proposition 9.4.** Let $A \subset X$, $f, g: A \to \mathbb{R}$, 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. \qed

Here’s the

**Arithmetic of Limits.** Let $A \subset X$, $f, g: A \to \mathbb{R}$, 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 $\lim g \neq 0$ and $0 \notin \text{ran } g$.

Proof. Immediate from the Sequential Characterization of Limits and the corresponding result for sequences. \qed
When evaluating a limit of the form \( \lim_{x \to t} f(x)/g(x) \), the above proposition tells you to find \( \lim f \) and \( \lim g \) separately, then divide. But as you saw in your calculus course, this is not always possible — for example, what if \( \lim_{x \to t} g(x) = 0 \)? Well, if \( f \not\to 0 \), the best we can hope for is an infinite limit, which we discuss later. But if \( \lim_{x \to t} f(x) = 0 \) also, then we don’t know without further investigation what the deal is with \( \lim f/g \). This is one example of what your calculus book called “indeterminate forms” — the “form” in this case is just denoted by “0/0”, and the “indeterminacy” means that further investigation is necessary. You know from your calculus course that the general remedy is l’Hôpital’s Rule, which we’ll see later. But just to give a trivial example:

**Example.** Let \( n, k \in \mathbb{N} \), and define \( f, g \colon (0, \infty) \to \mathbb{R} \) by

\[
 f(x) = x^n \quad \text{and} \quad g(x) = x^k.
\]

Then \( \lim_{x \to 0} f(x) = \lim_{x \to 0} g(x) = 0 \). If \( n = k \) then \( \lim_{x \to 0} f(x)/g(x) = 1 \), while if \( n > k \) the limit is 0. If \( n < k \) the limit fails to exist, although in a special way if the difference is even — see the infinite limits a little later.

And here’s a less trivial, although not very illuminating, example:

**Example.** \( \lim_{x \to 1} (x^4 - x^3 + x^2 - 1)/(x^5 - x^4 + 2x^2 - 2) = 3/5 \).

The following result has a very awkward appearance, but that’s because that we have to make sure we’re talking about the behavior of the function close to, but not at, the cluster point, and we’re doing it for two functions whose domains are unrelated. Roughly speaking, the result tells us how limits behave under composition of functions. We’ll get cleaner versions of the result once we have continuity available.

**Composition of Limits.** Let \( A \subset X, B \subset Y, t \in A', u \in B', v \in Z, f \colon A \to B \setminus \{u\}, \text{ and } g \colon B \to Z \). If \( \lim_{x \to t} f(x) = u \) and \( \lim_{y \to u} g(y) = v \), then

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

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

In the above proof, we didn’t get to apply a “corresponding result for sequences”, because there wasn’t one, but as a consolation we got to apply the Sequential Characterization of Limits twice.

The following shows that limits preserve weak inequalities:
Proposition 9.5. Let $A \subset X$, $f, g: A \to \mathbb{R}$, 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. $\square$

One-sided limits:

Definition. Let $A \subset \mathbb{R}$, $f: A \to Y$, and $t \in \mathbb{R}$.

(i) If $t \in (A \cap (t, \infty))'$, the right-hand limit of $f$ at $t$ is the limit at $t$ of the restriction $f|_{(A \cap (t, \infty))}$.

(ii) If $t \in (A \cap (-\infty, t))'$, the left-hand limit of $f$ at $t$ is the limit at $t$ of the restriction $f|_{(A \cap (-\infty, t))}$.

Notation and Terminology. Any of the following denote the right-hand limit of $f$ at $t$:

$$f(t+) = \lim_{x \downarrow t} f(x) = \lim_{x \to t^+} f(x),$$

and similarly for the left-hand limit:

$$f(t-) = \lim_{x \uparrow t} f(x) = \lim_{x \to t^-} f(x),$$

Exercise 9.6. Define $f: \mathbb{R} \setminus \{0\} \to \mathbb{R}$ by

$$f(x) = \frac{|x|}{x}.$$ 

Find $f(0^+)$ and $f(0^-)$.

Exercise 9.7. Let $A \subset \mathbb{R}$, $f: A \to Y$, $t \in \mathbb{R}$, and $u \in Y$. Assume $t \in (A \cap (t, \infty))'$. Prove that $\lim_{x \downarrow t} f(x) = u$ if and only if for all $\epsilon > 0$ there exists $\delta > 0$ such that

$$d(f(x), u) < \epsilon \quad \text{whenever} \quad x \in A \text{ and } t < x < t + \delta.$$

It is easy to see that

$$\lim_{x \downarrow t} f(x) = \lim_{x \downarrow t} f(-x),$$

so from any fact concerning right-hand limits we can more-or-less automatically deduce a corresponding fact concerning left-hand limits.

Lemma 9.8. Let $A \subset \mathbb{R}$, $f: A \to Y$, 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 $f(t-) = f(t+) = u$. Let $\epsilon > 0$. By hypothesis we can choose $\delta > 0$ such that $d(f(x), u) < \epsilon$ if $t - \delta < x < t$ or $t < x < t + \delta$, hence whenever $0 < |x - t| < \delta$. \qed

In the above proof we chose $\delta > 0$ to do two consistent jobs (involving $d(f(x), u)$ simultaneously, which is a shortcut we have seen before: we can choose a $\delta$ separately for each task, then take the minimum of the two $\delta$’s.

It’s often hard to determine whether a limit exists, but for monotone functions it’s easy:

**Monotone Limits.** Let $f: [a, b] \to \mathbb{R}$ be increasing and $t \in [a, b]$.

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 $f(s+) \leq f(t-)$.\noindent Similarly for decreasing.

Proof. (i). Assume $a < t \leq b$, and put $u = \sup_{x < t} f(x)$. Then $u \leq f(t)$ since $f$ is increasing. It remains to show $f(t-) = u$. Let $\epsilon > 0$. By Approximation of Suprema we can choose $c < t$ such that $f(c) > u - \epsilon$. Then

$$c < x < t \implies u - \epsilon < f(c) \leq f(x) \leq u \implies |f(x) - u| < \epsilon.$$ 

Thus $f(t-) = u$.

(ii). Similar to (i).

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

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

\noindent Infinite limits:

**Definition.** Let $A \subset X$, $f: A \to \mathbb{R}$, and $t \in A'$.

(i) $f(x)$ goes to $\infty$ as $x$ goes to $t$ if for all $z \in \mathbb{R}$ there exists $\delta > 0$ such that

$$f(x) > z \quad \text{for all } x \in A \cap B_\delta(t) \setminus \{t\}.$$ 

(ii) $f(x)$ goes to $-\infty$ as $x$ goes to $t$ if for all $z \in \mathbb{R}$ there exists $\delta > 0$ such that

$$f(x) < z \quad \text{for all } x \in A \cap B_\delta(t) \setminus \{t\}.$$
Notation and Terminology. The notation for infinite limits is similar to that for ordinary limits, for example we write \( f(x) \xrightarrow{x \to \infty} \infty \) or \( \lim_{x \to \infty} f(x) = \infty \).

We have seen above that many concepts and results for convergence of sequences have analogues for limits of functions. The following is more than just an analogue — it’s a generalization:

**Limits at infinity:**

**Definition.** Let \( A \subset \mathbb{R} \), \( f : A \to Y \), and \( u \in Y \).

(i) If \( A \) is unbounded above, then \( f(x) \) goes to \( u \) as \( x \) goes to \( \infty \) if for all \( \epsilon > 0 \) there exists \( z \in \mathbb{R} \) such that

\[
d(f(x), u) < \epsilon \quad \text{for all} \quad x > z.
\]

(ii) If \( A \) is unbounded below, then \( f(x) \) goes to \( u \) as \( x \) goes to \( -\infty \) if for all \( \epsilon > 0 \) there exists \( z \in \mathbb{R} \) such that

\[
d(f(x), u) < \epsilon \quad \text{for all} \quad x < z.
\]

Notation and Terminology. The notation for limits at infinity follows a familiar pattern: \( f(x) \to u \) as \( x \to \infty \), or \( f(x) \xrightarrow{x \to \infty} u \), and \( u \) is the limit of \( f \) at \( \infty \), or \( u = \lim_{x \to \infty} f(x) \).

Similarly for \( -\infty \).

There are of course other variations involving infinity: for example, \( \lim_{x \to -\infty} f(x) = \infty \) if for all \( z \in \mathbb{R} \) there exists \( w \in \mathbb{R} \) such that

\[
f(x) > z \quad \text{for all} \quad x > w.
\]

Altogether, we can have both \( x \) and \( f \) going to a real number or \( \pm \infty \), giving 9 possibilities; you should write them all to compare and contrast.

**Exercise 9.9.** Write the definition of \( \lim_{x \to -\infty} f(x) = \infty \).

If \( (x_n) \) is a sequence, then in the above definition we can let \( A = \mathbb{N} \), which is certainly unbounded above (by the Archimedean Principle), and then the above definition of \( \lim_{n \to -\infty} x_n \) agrees with the original definition of limit for sequences.

Most results for “ordinary” limits extend to infinite limits and limits at infinity:

(i) The Sequential Characterization of Limits extends to limits at infinity and to infinite limits. For example, if \( \text{dom } f \) is not bounded above, then \( \lim_{x \to \infty} f(x) \) exists if and only if for every sequence \( (x_n) \) in \( \text{dom } f \) diverging to \( \infty \), the sequence \( (f(x_n)) \) converges — in which case \( \lim_{x \to \infty} f(x) \) is the common limit of these sequences \( (f(x_n)) \).
(ii) The basic properties of limits, for example arithmetic and preservation of weak inequalities, extend to limits at infinity.

(iii) The basic properties of infinite limits of sequences extend to infinite limits of functions, for example the limited form of arithmetic: if \( f, g : A \to \mathbb{R} \) and \( \lim_{x \to t} f(x) = \infty \), then:

(a) \( \lim_{x \to t} (f(x) + g(x)) = \infty \) if \( g \) is bounded below;
(b) \( \lim_{x \to t} (f(x)g(x)) = \infty \) if \( \inf_{x \in A} g(x) > 0 \);
(c) \( \lim_{x \to t} g(x)/f(x) = 0 \) if \( g \) is bounded and \( 0 \notin \text{ran } f \).

And \( t = \pm \infty \) is ok here as well. In (b) it’s really only important what happens for \( x \) close to \( t \) (or sufficiently large or small, if we’re considering \( x \to \infty \) or \( x \to -\infty \)). A common special case of (b) occurs when \( \lim_{x \to t} g(x) \) exists and is positive. A similar rule applies if \( \sup g < 0 \), in which case \( fg \to -\infty \), and there are other obvious variations if \( f \to -\infty \).

From your calculus course you probably remember another “indeterminate form”, namely \( \infty/\infty \) — just as for the indeterminate form \( 0/0 \), what’s going on here is that we have a limit of the form \( \lim_{x \to t} f(x)/g(x) \), each of \( f \) and \( g \) is going to either \( \infty \) or \( -\infty \) (not necessarily the same sign for both). Again, further investigation is required to see what the deal is with the limit, and here’s the most trivial example:

**Example.** If \( n, k \in \mathbb{N} \) then

\[
\lim_{x \to \infty} \frac{x^n}{x^k} = \begin{cases} 
\infty & \text{if } n > k \\
1 & \text{if } n = k \\
0 & \text{if } n < k.
\end{cases}
\]

Again, in general the preferred weapon is l’Hôpital’s Rule, coming later.

Here’s a good application of the limited arithmetic for infinite limits:

**Exercise 9.10.** Let \( f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \) be an \( n \)th-degree polynomial. For simplicity, let’s assume \( a_n > 0 \). Prove:

(a) \( \lim_{x \to \infty} f(x) = \infty \);
(b) \( \lim_{x \to -\infty} f(x) = \begin{cases} 
\infty & \text{if } n \text{ is even} \\
-\infty & \text{if } n \text{ is odd}
\end{cases} \)

Hint: factor out \( x^n \).

We can also compose limits of all types (ordinary, infinite, and limits at infinity):

**Example.** Since \( e^x \to \infty \) at \( \infty \) and \( 0 \) at \( -\infty \), we have \( e^{1/x} \to \infty \) as \( x \downarrow 0 \) and \( 0 \) as \( x \uparrow 0 \).
Project 9.11. Let $A \subset X$, $f: A \to \mathbb{R}$, and $t \in A'$. Assume $f$ is bounded. Let $S$ be the set of all subsequential limits of sequences of the form $(f(x_n))$ where $(x_n)$ is a sequence in $A \setminus \{t\}$ converging to $t$.

(a) Prove that $S$ is bounded.

Define:

$$\lim_{x \to t} f(x) = \sup S$$

$$\liminf_{x \to t} f(x) = \inf S,$$

and call them the $\limsup$ and $\liminf$, respectively, of $f$ at $t$.

(b) Prove that $\lim_{x \to t} f(x) \in S$.

Similarly, it can be shown that $\liminf_{x \to t} f(x) \in S$, so in fact the $\limsup$ and $\liminf$ are the largest and smallest subsequential limits of images under $f$ of sequences in $A \setminus \{t\}$ converging to $t$.

(c) Prove that $\lim_{x \to t} f(x)$ exists if and only if

$$\lim_{x \to t} f(x) = \lim_{r \downarrow 0} H(r).$$

Similarly, it can be shown that if we define $h(r) = \inf F_r$, then $h$ is decreasing on $(0, \infty)$ and $\lim_{x \to t} f(x) = \sup h = \lim_{r \downarrow 0} h(r)$.

Define the oscillation\footnote{A slightly different version of oscillation is used to detect continuity.} of $f$ at $t$ as

$$\omega(t) := \sup\{|x - y| \mid x, y \in S\}.$$

(e) Prove that $\omega(t) = \lim_{x \to t} f(x) - \liminf_{x \to t} f(x)$.

(f) Prove that $\lim_{x \to t} f(x)$ exists if and only if $\omega(t) = 0$.

(g) Prove that

$$\omega(t) = \inf \sup_{r > 0}\{|f(x) - f(y)| \mid x, y \in A \cap B_r(t) \setminus \{t\}\}$$

$$= \lim_{r \downarrow 0}\sup_{r > 0}\{|f(x) - f(y)| \mid x, y \in A \cap B_r(t) \setminus \{t\}\}.$$

(h) Prove that

$$\omega(t) = \inf\left\{\epsilon > 0 \mid \text{there exists } r > 0 \text{ such that } |f(x) - f(y)| \leq \epsilon \text{ for all } x, y \in A \cap B_r(t) \setminus \{t\}\right\}.$$
10. Continuity

Continuous functions are the ones for which we can really start to say something interesting. For example, we’ll see that continuous functions are the ones which preserve convergence, for which we can hope to find have maximum and minimum values, and which we have no worries about integrating.

Notation and Terminology. As before, letters such as $X$, $Y$, and $Z$ will refer to metric spaces, unless otherwise specified.

We are most interested in continuity on whole domain, although continuity is really a pointwise property:

Definition. Let $f: X \to Y$ and $t \in X$. $f$ is continuous at $t$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that $d(f(x), f(t)) < \epsilon$ for all $x \in B_\delta(t)$.

Observation. $f$ is continuous at $t$ if and only if for all $\epsilon > 0$ there exists $\delta > 0$ such that $f(B_\delta(t)) \subset B_\epsilon(f(t))$.

Notation and Terminology. (i) $f$ is continuous if it’s continuous at each point of $X$. More generally, if $B \subset X$, we say $f$ is continuous on $B$ if $f$ is continuous at each point of $B$.

(ii) Discontinuous means not continuous.

(iii) $f$ has a discontinuity at $t$, or $t$ is a discontinuity of $f$, if $f$ is discontinuous at $t$.

When we see the above $\epsilon$-$\delta$ definition, we can’t help but think of limits. Indeed, in your calculus course you might have seen a definition of continuity which directly involved limits. And we’ll see in Theorem 10 how limits are related to continuity. However, it turns out that in general this is not as useful as you might expect. Much more powerful is the relation between continuity and convergence of sequences:

Sequential Characterization of Continuity. Let $f: X \to Y$ and $t \in X$. Then $f$ is continuous at $t$ if and only if for every sequence $(x_n)$ in $X$, 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 $X$ converging to $t$. Let $\epsilon > 0$. Choose $\delta > 0$ such that $d(f(x), f(t)) < \epsilon$ for all $d(x, t) < \delta$. Since $x_n \to t$, we can choose $k \in \mathbb{N}$ such that for all $n \geq k$ we have $d(x_n, t) < \delta$, hence $d(f(x_n), f(t)) < \epsilon$. Thus $f(x_n) \to f(t)$.

We prove the converse direction by contrapositive. Suppose $f$ is discontinuous at $t$. We must show that there exists a sequence $(x_n)$...
in $X$ such that $x_n \to t$ but $f(x_n) \not\to f(t)$. By discontinuity we can choose $\epsilon > 0$ such that for all $\delta > 0$ there exists $x \in X$ such that $d(x, t) < \delta$ and $d(f(x), f(t)) \geq \epsilon$. In particular, for all $n \in \mathbb{N}$ we can choose $x_n \in X$ such that

$$d(x_n, t) < \frac{1}{n} \quad \text{and} \quad d(f(x_n), f(t)) \geq \epsilon.$$ 

Then $x_n \to t$ by the Squeeze Theorem, but $f(x_n) \not\to f(t)$, as desired. \hfill $\Box$

You should notice that the above proof was similar to, but easier than, that of the Sequential Characterization of Limits.

**Exercise 10.1.** Let $X$ be a metric space and $A \subset X$, and define $f: X \to \mathbb{R}$ by

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

Prove that the discontinuities of $f$ are precisely at the boundary points of $A$.

**Exercise 10.2.** Let $f: X \to Y$, and let $A \subset X$. Assume that the restriction $f|A$ is continuous.

(a) Give an example where $f$ is not continuous on $A$.

(b) Prove that if $A$ is open, then $f$ is continuous on $A$.

**Exercise 10.3.** In each part, give an example of a function $f: \mathbb{R} \to \mathbb{R}$ with the indicated property:

(a) $f$ is discontinuous everywhere;

(b) $f$ is continuous at only one point.

Just as we did for limits, we can now develop most of the theory of continuity using sequences. Let’s begin with the arithmetic of continuity. But first, here’s some convenient notation:

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

(i) $(f + g)(x) := f(x) + g(x)$;

(ii) $(cf)(x) := cf(x)$ if $c \in \mathbb{R}$;

(iii) $(fg)(x) := f(x)g(x)$;

(iv) $\left(\frac{f}{g}\right)(x) := \frac{f(x)}{g(x)}$ if $0 \notin \text{ran } g$.

**Arithmetic of Continuity.** Let $f, g: X \to \mathbb{R}$ and $t \in X$. 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 \( 0 \notin \text{ran } g \).

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

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

Example. Thus, for example, every rational function is continuous.

The Sequential Characterization of Continuity can be applied to iterative methods of solving equations, for example:

Exercise 10.4. Define a sequence \( (x_n) \) inductively by

\[
x_n = \begin{cases} 
  2 & \text{if } n = 1 \\
  \frac{x_{n-1}}{2} + \frac{1}{x_{n-1}} & \text{if } n > 1 
\end{cases}
\]

Assuming the sequence \( (x_n) \) converges, find it’s limit, and prove you are correct.

Here comes the connection with limits, although it probably doesn’t look as simple as might be expected — the problem is that limits require cluster points and continuity does not. Consequently, we have to consider two cases:

Limit Characterization of Continuity. Let \( f: X \to Y \) and \( t \in X \).

(i) If \( t \notin X' \), then \( f \) is continuous at \( t \).
(ii) If \( t \in X' \), then \( f \) is continuous at \( t \) if and only if

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

Proof. \( ^{44} \)

(i). If \( t \notin X' \) then there exists \( \delta > 0 \) such that \( B_\delta(t) \subset \{t\} \), so \( f(B_\delta(t)) \subset B_\epsilon(f(t)) \) for all \( \epsilon > 0 \).

\(^{44}\)We could prove this by passing through the sequential characterizations, but the following proof is actually a little faster. Moreover, it gives us good practice with the definitions of continuity and limit. Finally, it would be anally retentive to insist upon doing everything using sequences.
(ii). Just observe that for all $\epsilon, \delta > 0$ we have:
\[
d(f(x), f(t)) < \epsilon \quad \text{for } x \in B_\delta(t) \setminus \{t\}
\]
\[
\iff d(f(x), f(t)) < \epsilon \quad \text{for } x \in B_\delta(t)
\]
\[
\square
\]

Case (i) of the above theorem deserves further comment:

**Definition.** A point $x$ in $X$ is *isolated* if $x \notin X'$.

Isolated points are irrelevant as far as limits are concerned, and (i) of the above theorem says functions are automatically continuous at isolated points of their domain.

**Example.** $\mathbb{R}$ has no isolated points, since every real number is a cluster point of $\mathbb{R}$.

**Exercise 10.5.** Prove that $x$ is an isolated point of $X$ if and only if the singleton set $\{x\}$ is open.

**Exercise 10.6.** Prove that in a discrete metric space, every point is isolated.

**Exercise 10.7.** Prove that in the subspace $A = \{1/n \mid n \in \mathbb{N}\}$ of $\mathbb{R}$, every point is isolated.

Between the Sequential and the Limit Characterizations of Continuity, we'll certainly find the Sequential one more useful. However, in some cases limits are quite useful in the study of continuity, for example with monotone functions, as illustrated in the next exercise:

**Exercise 10.8.** Let $f: (a, b) \to \mathbb{R}$ be monotone. Prove that $f$ has only countably many discontinuities. Hint: for each $x \in (a, b)$ consider the interval $(f(x)-, f(x)+)$, and prove that every pairwise-disjoint family of open intervals is countable.

It turns out that we can turn the Limit Characterization of Continuity around and get another technical tool which seems superfluous at first glance, but which we'll need later (Lemma [15]):

**Continuity Characterization of Limits.** Let $A \subset X$, $f: A \to Y$, $t \in A'$, and $u \in Y$. Then $\lim_{x \to t} f(x) = u$ if and only if the function $g: A \cup \{t\} \to Y$ defined by
\[
g(x) = \begin{cases} 
         f(x) & \text{if } x \neq t \\
         u & \text{if } x = t
       \end{cases}
\]
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 \( \lim_{x \to t} f(x) = u \). 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. \( \square \)

In your calculus course you might have made use of the above result to “remove discontinuities”: if the limit of \( f \) at \( t \) exists but is different from the value of \( f \) at \( t \), we can “remove the discontinuity” at \( t \) by redefining \( f(t) = \lim_{x \to t} f(x) \). However, note that if the limit exists but \( t \) is not in the domain of \( f \), it makes no sense to ask whether \( f \) is continuous at \( t \); the best we can do is extend \( f \) to a function which is continuous at \( t \).

Here come the relations among continuity, compositions, and limits:

**Proposition 10.9.**

(i) Let \( f : X \to Y, \ t \in X \), and \( g : Y \to Z \). If \( f \) is continuous at \( t \) and \( g \) is continuous at \( f(t) \), then \( g \circ f \) is continuous at \( t \).

(ii) Let \( A \subset X \), \( f : A \to Y \), \( t \in A' \), \( u \in Y \), and \( g : Y \to Z \). If \( \lim_{x \to t} f(x) = u \) and \( g \) is continuous at \( u \), then \( \lim_{x \to t} g \circ f(x) = g(u) \).

Proof. (i). 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 \( g \). Therefore \( g \circ f \) is continuous at \( t \) by the Sequential Characterization of Continuity.

(ii). If \( x_n \in A \setminus \{t\} \) and \( x_n \to t \), then \( f(x_n) \to u \) by the Sequential Characterization of Limits, so \( g \circ f(x_n) \to g(u) \) by the Sequential Characterization of Continuity. Therefore \( \lim_{x \to t} g \circ f(x) = g(u) \) by the Sequential Characterization of Limits. \( \square \)

We actually used the Sequential Characterization of Continuity a few more times in the above proof than we explicitly mentioned — after a little while it becomes burdensome to read so many repetitious references. After acquiring some familiarity with the uses of the Sequential Characterizations, it’s customary to not specifically mention them when they’re used.

We haven’t used open sets in a while, but here comes the main reason they were invented:

**Open Sets Characterization of Continuity.** Let \( f : X \to Y \). Then \( f \) is continuous if and only if for every open subset \( A \subset Y \), the inverse image \( f^{-1}(A) \) is open in \( X \).

Proof. First assume \( f \) is continuous, and let \( A \subset Y \) be open and \( x \in f^{-1}(A) \). Since \( A \) is open, there exists \( \epsilon > 0 \) such that \( B_\epsilon(f(x)) \subset A \).
By continuity there exists $\delta > 0$ such that $f(B_\delta(x)) \subset B_\epsilon(f(x))$, hence

$$B_\delta(x) \subset f^{-1}(B_\epsilon(f(x))) \subset f^{-1}(A).$$

Therefore $f^{-1}(A)$ is open.

Conversely, assume inverse images of open sets are open, and let $x \in X$ and $\epsilon > 0$. Then $B_\epsilon(f(x))$ is open, so $f^{-1}(B_\epsilon(f(x)))$ is open. Since $x \in f^{-1}(B_\epsilon(f(x)))$, there exists $\delta > 0$ such that $B_\delta(x) \subset f^{-1}(B_\epsilon(f(x)))$. Thus $f(B_\delta(x)) \subset B_\epsilon(f(x))$, and we’ve shown $f$ is continuous at $x$. □
11. Completeness

In this section we'll investigate a curious feature of convergence: if you look at the definition of convergence of sequences, you'll quickly convince yourself that there's no way to rephrase it so that it does not explicitly mention the limit. This brings up a natural question: is there some way to determine whether a sequence converges without finding the limit? It turns out that the answer is yes, and the method involves a test invented by Cauchy, a 19th-Century mathematician among the “Founding Parents”\(^45\) of mathematical analysis. However, the test is only guaranteed to work if the metric space is “complete” — that is, willing to provide us with the limit.

**Notation and Terminology.** As before, the letter \(X\) will refer to a metric space, unless otherwise specified.

**Definition.** A sequence \((x_n)\) in \(X\) is *Cauchy* if for all \(\epsilon > 0\) there exists \(k \in \mathbb{N}\) such that

\[
d(x_n, x_j) < \epsilon \quad \text{for all } n, j \geq k.
\]

The above definition looks a lot like the definition of convergence of the sequence, but does *not* mention a limit. Rather, it says that the sequence “bunches up” if we go “far enough out”\(^46\)

Here’s a “pecking order” of sequences:

**Proposition 11.1.** Every convergent sequence is Cauchy, and every Cauchy sequence is bounded.

**Proof.** Let \((x_n)\) be a sequence in \(X\). First, assume \((x_n)\) converges, and let \(x = \lim x_n\). Let \(\epsilon > 0\). Choose \(k \in \mathbb{N}\) such that \(d(x_n, x) < \epsilon/2\) for all \(n \geq k\). Then for all \(n, j \geq k\),

\[
d(x_n, x_j) \leq d(x_n, x) + d(x, x_j) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,
\]

so \((x_n)\) is Cauchy.

For the other part, assume \((x_n)\) is Cauchy. Choose \(k \in \mathbb{N}\) such that \(d(x_n, x_j) < 1\) for all \(n, j \geq k\). Put \(M = \max\{1, d(x_1, x_k), \ldots, d(x_{k-1}, x_k)\}\) (where without loss of generality \(k > 1\)). Then \(d(x_n, x_k) \leq M\) for all \(n \in \mathbb{N}\), so \((x_n)\) is bounded. \(\square\)

The first part of the above proof is a classic “\(\epsilon/2\)-argument”, using the Triangle Inequality to connect both \(x_n\) and \(x_k\) with \(x\). The second

\(^{45}\)politically correct!

\(^{46}\)In fact, a professor of mine made up the name “far-out-bunching effect”, or “FOBE”, for this property.
part of the proof should remind you of the proof that a convergent
sequence is bounded.

Now, here’s the rub: a Cauchy sequence might not converge:

Exercise 11.2. Let \( X = (0, 2) \). Prove that the sequence \( \frac{1}{n} \) is
Cauchy but divergent in the metric space \( X \).

Also:

Exercise 11.3. Find an example of a continuous function \( f \) and a
Cauchy sequence \( (x_n) \) such that the sequence \( (f(x_n)) \) is not Cauchy.

Thus, although Cauchy sequences “want to converge”, sometimes
the metric space doesn’t let them. Metric spaces lacking this mean-
spiritedness deserve recognition:

Definition. \( X \) is complete if every Cauchy sequence in \( X \) converges.

Exercise 11.4. Prove that every discrete metric space is complete.

We’ve seen before that limits of functions have properties similar
to limits of sequences; as another illustration of this, here’s a Cauchy
criterion:

Exercise 11.5. Let \( A \subset X \), \( t \in A' \), and \( f: A \to Y \). Prove that
\( \lim_{x \to t} f(x) \) exists if and only if for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such
that
\[
d(f(x), f(y)) < \epsilon \quad \text{for all } x, y \in A \cap B_{\delta}(t) \setminus \{t\}.
\]

Among the most important metric spaces are the normed spaces
(and subspaces of them) — think of \( \mathbb{R}^n \) — and the complete ones are
so dandy they are given a name all their own:

Definition. A Banach space is a normed space for which the associated
metric space is complete.

The following result shows that once we find one complete metric
space, we get lots more for free:

Proposition 11.6. Every closed subset of a complete metric space is
complete.

Proof. Assume \( X \) is complete and \( A \subset X \) is closed. Let \( (x_n) \) be a
Cauchy sequence in \( A \). We must show \( (x_n) \) converges in \( A \). Since \( X \)
is complete, there exists \( x \in X \) such that \( x_n \to x \). Since \( A \) is closed,
we must have \( x \in A \). Thus the sequence \( (x_n) \) converges in the metric
space \( A \). Therefore \( A \) is complete. \( \square \)

In fact, the connection between completeness and closedness is closer
(!) than that:
Proposition 11.7. Every complete subset of a metric space is closed.

Proof. Let $X$ be a metric space, and let $A$ be a complete subspace. Let $(x_n)$ be a sequence in $A$, and assume $(x_n)$ converges in $X$, that is, there exists $y \in X$ such that $x_n \to y$. We must show that $y \in A$. Since $(x_n)$ converges, it’s Cauchy. Since $A$ is complete, there exists $x \in A$ such that $x_n \to x$. Since the limit of a convergent sequence is unique, we must have $x = y$, hence $y \in A$, as desired. Thus $A$ is closed. □

We’ll show soon that $\mathbb{R}^n$ is complete, but in preparation for this we need to extend the Bolzano-Weierstrass Theorem:

Bolzano-Weierstrass Theorem. Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Proof. Let $(x_k)$ be a bounded sequence in $\mathbb{R}^n$.

Case 1. $n = 1$. Then the result is just the Bolzano-Weierstrass Theorem for $\mathbb{R}$.

Case 2. $n > 1$. Then each coordinate sequence is bounded in $\mathbb{R}$. For each $k \in \mathbb{N}$ let $x_k = (x_k(1), \ldots, x_k(n))$.[47] By Case 1, we can choose a subsequence $(x_{1,k})$ of $(x_k)$ such that the first coordinates $(x_{1,k}(1))$ converge. Then we can choose a subsequence $(x_{2,k})$ of $(x_{1,k})$ such that the second coordinates $(x_{2,k}(2))$ converge. Of course, $(x_{2,k})$ is also a subsequence of $(x_k)$, and the first coordinates $(x_{2,k}(1))$ converge. Continuing with the third coordinates and so on, we eventually get a subsequence $(x_{n,k})$ of $(x_k)$ such that for each $i = 1, \ldots, n$ the coordinate sequence $x_{n,k}(i)$ converges. Then sequence $(x_{n,k})$ of vectors converges as well. □

Corollary 11.8. Every bounded infinite subset of $\mathbb{R}^n$ has a cluster point.

Proof. Let $A$ be a bounded infinite subset of $\mathbb{R}^n$. Then there is a 1-1 sequence $(x_k)$ in $A$. Since $A$ is bounded, so is the sequence $(x_k)$. By the Bolzano-Weierstrass Theorem, we can pick a subsequential limit $x$ of $(x_k)$. Then either $x$ is a cluster point of the range of $(x_k)$, hence of the superset $A$, or $x_k = x$ for infinitely many $k$. Since $(x_k)$ is 1-1, we could only have $x_k = x$ for at most 1 value of $k$. Therefore, $x$ must be a cluster point of $A$. □

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.

[47]Here we’ll illustrate another strategy for dealing with the notational nightmare of sequences in $\mathbb{R}^n$. Our job is further complicated by the need to choose $n$ successive subsequences — try to do that using the “sub-subscript” (and so on) notation for subsequences!
Corollary 11.9. \( \mathbb{R}^n \) is complete.

Proof. Let \((x_k)\) be a Cauchy sequence in \(\mathbb{R}^n\). Then \((x_k)\) is bounded, so by the Bolzano-Weierstrass Theorem \((x_k)\) has a convergent subsequence. By the elementary lemma below \(^{48}\) \((x_k)\) itself must converge. \(\square\)

Here’s the lemma promised in the preceding proof:

Lemma 11.10. A Cauchy sequence converges if it has a convergent subsequence.

Proof. Let \((x_n)\) be a Cauchy sequence in \(X\), and suppose some subsequence \((x_{n_k})\) converges, with limit \(x\). Let \(\epsilon > 0\). Choose \(l \in \mathbb{N}\) such that \(d(x_n, x_j) < \epsilon/2\) for all \(n, j \geq l\). Also choose \(p \in \mathbb{N}\) such that \(d(x_{n_k}, x) < \epsilon/2\) for all \(k \geq p\). Put \(k = \max\{p, l\}\). Then \(n_k \geq k \geq l\), so for all \(n \geq l\) we have

\[
d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_k}, x) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.\]

Thus \((x_n)\) itself converges. \(\square\)

The above proof was fairly simple, but it’s made fussy by the notational difficulties: we’re controlling part of the sequence using the “main” subscript (when we apply the Cauchy condition), and at the same time we’re controlling the subsequence using the “secondary” subscript; then we make matters worse by using these two controls simultaneously. So, the hardest thing about the proof is keeping track of the notation. But the basic idea of the proof is simple: we can make all the terms of the subsequence eventually close to \(x\), and we can make all the terms of the entire sequence close to each other eventually, so eventually all the terms of the entire sequence are close to \(x\). But the exact order of our choices was crucial.

\(^{48}\)We’re ending this proof by referring to a lemma which follows the proof! This seems to violate the rule that each step in a proof can only use facts which have been previously proven. However, what we’re doing is ok, as long as we are careful to avoid “circular reasoning”, that is, we must be careful in the proof of the following lemma not to use any fact whose truth relies upon the lemma itself. The lemma itself is technical and hard to motivate by itself, while the theorem is important and easily recognized as such. Only after seeing in the proof of the theorem is it finally possible to see that the lemma would be worthwhile. Alternatively, we could have avoided the lemma altogether, by incorporating the argument used in its proof right into the proof of this theorem. However, I separated out the general lemma (and please note that it’s true in an arbitrary metric space, so its scope vastly transcends that of the proof of completeness of \(\mathbb{R}^n\)) so that it could be available for other purposes as a technical tool.
Note that the proof of completeness of \( \mathbb{R}^n \), in particular of \( \mathbb{R} \), required the Completeness Axiom — this is why this axiom is given that name!

The metric space \( \mathbb{R}^n \) has the property that every closed ball is complete. But this doesn’t always happen, in fact it can fail spectacularly:

**Exercise 11.11.** Give an example of a metric space none of whose closed balls (of positive radius) is complete.

The next result is a surprising and deep consequence of completeness: 

**Baire Category Theorem.** Every countable family of dense open sets in a complete metric space has dense intersection.

**Proof.** Let \( \{A_n\}_{n \in \mathbb{N}} \) be a countable family of dense open sets in a complete metric space \( X \). Let \( S \) be a nonempty open subset of \( X \). We must show \( S \cap \bigcap_{1}^{\infty} A_n \neq \emptyset \).

Since \( S \) is nonempty and open, and since \( A_1 \) is dense, \( S \cap A_1 \) is nonempty. Since \( S \) and \( A_1 \) are both open, so it \( S \cap A_1 \). Thus there exist \( x_1 \in X \) and \( r_1 > 0 \) such that \( B_1 := B_{r_1}(x_1) \subset S \cap A_1 \). Shrinking \( r_1 \) if necessary, we can assume both \( r_1 < 1 \) and \( \overline{B_1} \subset S \cap A_1 \). Next, since \( B_1 \) is nonempty and open, and \( A_2 \) is dense and open, \( B_1 \cap A_2 \) is nonempty and open, so there exists \( x_2 \in X \) and \( r_2 > 0 \) such that \( B_2 := B_{r_2}(x_2) \subset B_1 \cap A_2 \). Shrinking \( r_2 \) if necessary, we can assume both \( r_2 < 1/2 \) and \( \overline{B_2} \subset B_1 \cap A_2 \). Continuing, we inductively choose \( x_n \in X \) and \( r_n \in (0, 1/n) \) such that, with \( B_n := B_{r_n}(x_n) \), we have \( \overline{B_n} \subset B_{n-1} \cap A_n \). We now have a sequence \( (x_n) \) in \( X \). For all \( k \in \mathbb{N} \) we have \( x_n \in B_k \) for all \( n \geq k \).

Claim: \( (x_n) \) is Cauchy. Let \( \epsilon > 0 \). Choose \( k \in \mathbb{N} \) such that \( 1/k < \epsilon/2 \). Let \( n, j \geq k \). By construction, both \( x_n \) and \( x_j \) are in \( B_k \), which is an open ball with radius \( r_k < 1/k \). Thus \( d(x_n, x_j) < 2r_k < \epsilon \), proving the claim. Since \( X \) is complete, \( x := \lim_{n} x_n \) exists.

We finish by showing \( x \in S \cap \bigcap_{1}^{\infty} A_n \). Let \( i \in \mathbb{N} \). Then \( x_n \in \overline{B_i} \) for all \( n \geq i \), so \( x \in \overline{B_i} \) since \( \overline{B_i} \) is closed. Since \( \overline{B_i} \subset \overline{B_1} \) and \( B_1 \subset S \), we also have \( x \in S \), and this concludes the proof.

Believe it or not, the Baire Category Theorem is most often used in the following severely crippled form: instead of asserting that the intersection is dense, a lot of the time all that’s needed is that it’s nonempty. Actually, this latter fact is most often used in the equivalent form obtained by taking complements. Recall that the interior of \( A \) is the largest open set contained in \( A \). It follows from the definitions that a set is dense if and only if its complement has empty interior. So, the complement of an open dense set is a closed set with empty interior. By De Morgan’s Laws, an intersection is nonempty if and only if the
union of the complements is a proper subset. Thus we get the following very weak corollary, which is sometimes also called the Baire Category Theorem: A nonempty complete metric space is not a countable union of closed sets with empty interior. Actually, this is usually stated in slightly different form, which we’ll record officially in a moment, but first we need a

**Definition.** A subset $A$ of a metric space is *nowhere dense* if the closure $\overline{A}$ has empty interior, that is, $(\overline{A})^o = \emptyset$.

**Little Baire Category Theorem.** A nonempty complete metric space is not a countable union of nowhere dense sets.

**Proof.** If $\{A_n\}_{n=1}^\infty$ is a countable family of nowhere dense sets in a nonempty complete metric space $X$, then the family $\{\overline{A_n}\}_{n=1}^\infty$ of closures comprises closed sets with empty interior; as we discussed above, $X \neq \bigcup_{n=1}^\infty \overline{A_n}$, so certainly $X \neq \bigcup_{n=1}^\infty A_n$. $\square$

Here are a couple of applications:

**Exercise 11.12.** Prove that $\mathbb{R}^3$ is not a countable union of planes.

**Exercise 11.13.** Prove that a nonempty complete metric space with no isolated points is uncountable.
12. Separability

Notation and Terminology. As before, the letter $X$ will refer to a metric space, unless otherwise specified.

Definition. $X$ is separable if it has a countable dense subset.

Example. $\mathbb{R}^n$ is separable, because $\mathbb{Q}^n$ is dense. To see this, let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, for each $i = 1, \ldots, n$ there exists a sequence $(x_i^{(k)})$ in $\mathbb{Q}$ converging to $x_i$. For each $k \in \mathbb{N}$ put $x^{(k)} = (x_1^{(k)}, \ldots, x_n^{(k)})$. Then $(x^{(k)})$ is a sequence in $\mathbb{Q}^n$ converging to $x$. Thus $\mathbb{Q}^n$ is dense, and since it’s a finite product of countable sets, it’s countable.

Definition. A base of $X$ is a family $B$ of open subsets of $X$ such that every open subset of $X$ is the union of some subfamily of $B$.

Observation. A family $B$ of open subsets of $X$ is a base of $X$ if and only if for every open set $A$ and every element $x \in A$, there exists $U \in B$ such that $x \in U \subset A$.

Exercise 12.1. Prove that in any metric space $X$, the family $\{B_r(x) \mid x \in X, r > 0\}$ of open balls is a base.

Theorem 12.2. $X$ is separable if and only if it has a countable base.

Proof. First assume $X$ is separable, and choose a countable dense subset $D$. Claim: the family $B := \{B_{1/n}(t) \mid n \in \mathbb{N}, t \in D\}$ is a base of $X$. This will suffice, because the function

$$(n, t) \mapsto B_{1/n}(t) : \mathbb{N} \times D \to B$$

is onto, and $\mathbb{N} \times D$ is countable since it’s a finite product of countable sets. Let $A$ be an open subset of $X$, and let $x \in A$. We must show there exists $n \in \mathbb{N}$ and $t \in D$ such that $x \in B_{1/n}(t) \subset A$. Choose $r > 0$ such that $B_r(x) \subset A$, then $n \in \mathbb{N}$ such that $2/n < r$, and finally $t \in D$ such that $d(t, x) < 1/n$. Then $x \in B_{1/n}(t)$, and for all $y \in B_{1/n}(t)$ we have

$$d(y, x) \leq d(y, t) + d(t, x) < \frac{1}{n} + \frac{1}{n} = \frac{2}{n} < r,$$

hence $B_{1/n}(t) \subset B_r(x) \subset A$.

Conversely, assume $B$ is a countable base, and without loss of generality each $A \in B$ is nonempty. For each $A \in B$ choose a point $x_A \in A$. Then $D := \{x_A \mid A \in B\}$ is countable, so it suffices to show it’s dense. Let $C$ be a nonempty open subset of $X$. We can choose $A \in B$ such that $A \subset C$, and then we have $x_A \in D \cap C$. Thus $D \cap C \neq \emptyset$, so we have shown $D$ is dense. \qed
Thus, in a separable space, there is a countable family of open sets such that every open set is a union of sets from this family. It seems not too great a leap to hope that in a separable space uncountable unions of open sets are unnecessary, that is, every union of open sets can be reduced to a countable union — the following result confirms this important fact:

**Lindelöf's Theorem.** Let $X$ be separable, and let $\mathcal{F}$ be a family of open subsets of $X$. Then there exists a countable subfamily $\mathcal{G} \subset \mathcal{F}$ such that

$$\bigcup \mathcal{G} = \bigcup \mathcal{F}.$$ 

*Proof.* Choose a countable base $\mathcal{B}$ of $X$. For each set $A \in \mathcal{F}$, since $A$ is open and since $\mathcal{B}$ is a base we can choose a subfamily $\mathcal{S}_A \subset \mathcal{B}$ such that $A = \bigcup \mathcal{S}_A$. Then

$$\mathcal{S} := \bigcup_{A \in \mathcal{F}} \mathcal{S}_A$$

is countable, being a subfamily of $\mathcal{B}$. Each set $C \in \mathcal{S}$ arose as part of a union which gives some set in the family $\mathcal{F}$, so in particular we can choose a set $A_C \in \mathcal{F}$ such that $C \subset A_C$. Put

$$\mathcal{G} = \{A_C\}_{C \in \mathcal{S}},$$

a countable subfamily of $\mathcal{F}$. We have

$$\bigcup \mathcal{F} = \bigcup_{A \in \mathcal{F}} A = \bigcup_{A \in \mathcal{F}} \bigcup_{C \in \mathcal{S}_A} C = \bigcup_{A \in \mathcal{F}} \bigcup_{C \in \mathcal{S}_A} C = \bigcup_{A \in \mathcal{F}} \bigcup_{C \in \mathcal{S}_A} A_C = \bigcup_{C \in \mathcal{S}} A_C = \bigcup \mathcal{G} \subset \bigcup \mathcal{F},$$

so we must have equality throughout. \qed

In the above proof, the manipulations with families of sets, both indexed and nonindexed, were a little delicate. It’s not surprising that the proof was a little hard; the result is named after someone (Lindelöf), and this is usually a signal that the result is deep and/or hard to prove.

If we have a separable space, it’s natural to ask what subspaces are also separable. At first glance it might seem like a very hard question — a countable dense subset of the ambient space might have nothing to do with the subspace (they can be disjoint!), so how can we get a countable dense subset of the subspace? The amazing fact is that there is no problem: *every* subspace is separable:

**Exercise 12.3.** Prove that every subspace of a separable metric space is separable. Hint: prove first that if $\mathcal{B}$ is a base for a metric space $X$, then
and if $A$ is a subspace of $X$, then the family $\{A \cap U \mid U \in \mathcal{B}\}$ is a base for $A$.

We know $\mathbb{R}$ is separable, and our favorite countable dense subset is $\mathbb{Q}$. By the preceding exercise, we know that the subspace $\mathbb{R} \setminus \mathbb{Q}$ is also separable. But the exercise didn’t give us a method of finding a countable dense subset. In fact, it might seem difficult to imagine a countable set of irrational numbers which is dense in the irrationals, because the irrationals are so, well, irrational!

However, the following exercise will show how to find a countable set of irrational numbers which is dense in $\mathbb{R}$, hence in the subspace $\mathbb{R} \setminus \mathbb{Q}$.

**Exercise 12.4.** Find a countable set of irrational numbers which is dense in $\mathbb{R}$.
In your calculus course, you saw a lot of results about closed bounded intervals. In some cases you probably didn’t see the proof[49] in most cases the result was true because the interval was “compact” in the sense we’ll study in this section. For the purposes of analysis, being compact is almost as good as being finite (which is very good indeed).

Notation and Terminology. As before, letters such as $X$ and $Y$ will refer to metric spaces, unless otherwise specified.

For us the most useful form of compactness is in the following

**Definition.**[50] $X$ is *sequentially compact* if every sequence in $X$ has a convergent subsequence.

Sequentially compact spaces have lots of nice properties:

**Proposition 13.1.** Every sequentially compact metric space is complete and separable.

*Proof.* Let $X$ be sequentially compact. Every Cauchy sequence has a convergent subsequence, hence converges since it’s Cauchy. Thus $X$ is complete.

For separability, we first verify the following claim: if $n \in \mathbb{N}$, then there is a finite subset $A_n$ of $X$ such that $X = \bigcup_{x \in A_n} B_{1/n}(x)$. Suppose not. Choose $x_1 \in X$, and for $k > 1$ inductively choose $x_k \in \left( \bigcup_{i=1}^{k-1} B_{1/n}(x_i) \right)^c$.

We get a sequence $(x_k)$ in $X$ such that $d(x_k, x_j) \geq 1/n$ whenever $k \neq j$. Thus $(x_k)$ has no Cauchy subsequence, contradicting sequential compactness. This proves the claim.

Put $A = \bigcup_{1}^{\infty} A_n$. Then $A$ is countable since it’s a countable union of finite sets. We show that $A$ is dense: let $x \in X$ and $\epsilon > 0$. We must show that there exists $y \in A$ such that $d(x, y) < \epsilon$. Choose $n \in \mathbb{N}$ such that $1/n < \epsilon$, and then choose $y \in A_n$ such that $x \in B_{1/n}(y)$. Then $y \in A$ and $d(x, y) < 1/n < \epsilon$. □

**Corollary 13.2.** Every sequentially compact subset of a metric space is closed.

[49]“beyond the scope of the course”

[50]Here’s an example of a phenomenon which occurs all over the place in modern mathematics: a theorem becomes a definition — in this case the Bolzano-Weierstrass Theorem. In fact, sometimes the property defining sequential compactness is called the Bolzano-Weierstrass Property.
Proof. Let \( A \subset X \) be sequentially compact. Then \( A \) is complete, hence closed. \( \square \)

The following result shows that once we find one sequentially compact metric space, we get lots more for free:

**Proposition 13.3.** Every closed subset of a sequentially compact metric space is sequentially compact.

**Proof.** Assume \( X \) is sequentially compact and \( A \subset X \) is closed. Let \( (x_n) \) be a sequence in \( A \). Since \( X \) is sequentially compact, there is a subsequence \( (y_k) \) which converges in \( X \). Since \( A \) is closed, we have \( \lim y_k \in A \). We have shown \( A \) is sequentially compact. \( \square \)

Here comes an unexpected approach to compactness:

**Definition.**

(i) A *cover* of \( X \) is a family \( \mathcal{F} \) of sets such that \( X = \bigcup \mathcal{F} \). We also say \( \mathcal{F} \) *covers* \( X \).

(ii) A cover \( \mathcal{F} \) of \( X \) is *open* if each \( A \in \mathcal{F} \) is open in \( X \).

(iii) A *subcover* of a cover \( \mathcal{F} \) of \( X \) is a subfamily of \( \mathcal{F} \) which still covers \( X \).

**Definition.** \( X \) is *compact* if every open cover of \( X \) has a finite subcover.

In plainer (but somewhat more cumbersome) language, the condition for compactness is that whenever \( \mathcal{F} \) is a family of open sets with \( X = \bigcup \mathcal{F} \), then there exists a finite subfamily \( \mathcal{G} \subset \mathcal{F} \) such that \( X = \bigcup \mathcal{G} \).

The definition of compact takes a while to digest.

**Exercise 13.4.** Let \( (x_n) \) be a sequence in a metric space \( X \), and let \( x \in X \). Assume \( x_n \to x \). Use open covers to prove that the subspace

\[
A := \{x\} \cup \{x_n \mid n \in \mathbb{N}\}
\]

is compact.

We will get something useful in the definition of compactness by taking complements, but first, a definition to ease the way:

**Definition.** Let \( \mathcal{S} \) be a family of sets. We say \( \mathcal{S} \) has the *finite intersection property* if \( \bigcap \mathcal{F} \neq \emptyset \) for every finite subfamily \( \mathcal{F} \subset \mathcal{S} \).

---

51 This is pretty nonsensical — why are we calling the cover, which is after all a set (of sets), “open”? But this is one of those cases where the terminology is so ingrained we can’t do anything about it. Rest assured that if I had the power to design the terminology from scratch, some things would be different.

52 I saw a cartoon once showing a student puzzling over a table-top full of cosmetic-type compacts, one of which was open, and the caption was “Kevin still didn’t understand compact sets; the open cover seemed completely irrelevant.”
Proposition 13.5. Let $\mathcal{S}$ be a family of closed sets in a compact metric space $X$. If $\mathcal{S}$ has the finite intersection property, then $\bigcap \mathcal{S} \neq \emptyset$.

Proof. This follows from the definition of compactness upon taking complements and then the contrapositive. \hfill \Box

Theorem 13.6. For a metric space $X$, the following are equivalent:

(i) $X$ is compact;
(ii) every infinite subset has a cluster point;
(iii) $X$ is sequentially compact.

Proof. (i) implies (ii). Arguing by contradiction, suppose $A \subset X$ is infinite and $A' = \emptyset$. Then for every $x \in X$ there exists $r_x > 0$ such that $B_{r_x}(x) \cap A$ is finite. The family $\{B_{r_x}(x)\}_{x \in X}$ is an open cover of $X$, so by compactness there exists a finite subset $F \subset X$ such that $X = \bigcup_{x \in F} B_{r_x}(x)$. But then

$$A = \bigcup_{x \in F} (A \cap B_{r_x}(x))$$

is a finite union of finite sets, hence is finite, which is a contradiction.

(ii) implies (iii). We prove the contrapositive: suppose $(x_n)$ is a sequence with no convergent subsequence, and put $A = \text{ran}(x_n)$. Then $A$ is infinite, otherwise for some $x \in A$ the set $\{n \in \mathbb{N} \mid x_n = x\}$ would be infinite, making $x$ a subsequential limit. Also, $A' = \emptyset$, again since there are no subsequential limits.

(iii) implies (i). Assume $X$ is sequentially compact, and let $\{A_i\}_{i \in I}$ be an open cover of $X$. Since $X$ is separable, by Lindelöf’s Theorem we can assume $I = \mathbb{N}$. Arguing by contradiction, suppose $\{A_n\}_{n \in \mathbb{N}}$ has no finite subcover. Then for each $n \in \mathbb{N}$ we can choose

$$x_n \in \left( \bigcup_{i=1}^n A_k \right)^c.$$

There exists a subsequence $(x_{n_k})$ converging to some $x \in X$. Since the $A_n$’s cover $X$, there exists $n \in \mathbb{N}$ such that $x \in A_n$. Since $A_n$ is open, there exists $r > 0$ such that $B_r(x) \subset A_n$. Since $x_{n_k} \to x$, there exists $l \in \mathbb{N}$ such that $x_{n_k} \in B_r(x)$ for all $k \geq l$. But then there exists $i > n$ such that

$$x_i \in B_r(x) \subset A_n,$$

contradicting

$$x_i \notin A_1 \cup \cdots \cup A_n \cup \cdots \cup A_i.$$

\hfill \Box
The above result unshackles us from the term “sequentially compact”.\footnote{In fact, I had considered using “compact” for what we’ve defined as “sequentially compact”, since the two conditions are equivalent anyway; however, you should be familiar with the term sequentially compact.} Here’s an example of how convenient the open covers can be:

**Proposition 13.7.** Every compact metric space is bounded.

*Proof.* Suppose $X$ is unbounded. Choose $x \in X$. Then $X \neq B_n(x)$ for all $n \in \mathbb{N}$. Since $X = \bigcup_{n=1}^{\infty} B_n(x)$, we have an open cover $\{B_n(x)\}_{n \in \mathbb{N}}$ with no finite subcover. Therefore $X$ is not compact. \qed

Thus every compact subset of a metric space is closed and bounded.

The converse is false, even in a complete space:

**Exercise 13.8.** Let $X$ be an infinite discrete metric space. Prove that $X$ is complete and bounded, but not compact.

However some metric spaces are nice enough to disallow the above unpleasant behavior. For example:

**Heine-Borel Theorem.** Every closed and bounded subset of $\mathbb{R}^n$ is compact.

*Proof.* Let $A \subset \mathbb{R}^n$ be closed and bounded. Let $(x_k)$ be a sequence in $A$. Then $(x_k)$ is a bounded sequence in $\mathbb{R}^n$, so has a subsequence $(y_j)$ which converges in $\mathbb{R}^n$, by the Bolzano-Weierstrass Theorem. Since $A$ is closed, we have $\lim y_j \in A$. Thus $A$ is compact. \qed

Compact subsets of $\mathbb{R}$ are particularly nice:

**Theorem 13.9.** Every nonempty compact subset of $\mathbb{R}$ has both a maximum and a minimum.

*Proof.* Let $A \subset \mathbb{R}$ be nonempty and compact. Since $A$ is compact, it’s closed and bounded. Since $A$ is nonempty and bounded, $\sup A$ and $\inf A$ exist. Since $A$ is closed, $\sup A,\inf A \in A$. Thus $\sup A = \max A$ and $\inf A = \min A$. \qed

**Example.** We now describe the *Cantor Set*, one of the most important examples in analysis. It’s a compact subset of $\mathbb{R}$ with amazing properties.

The description will be a little complex, so it behooves us to simplify. For this purpose let’s agree upon what we mean by “removing the open middle third” from a closed interval: if $I$ is a closed interval of positive length $l > 0$ and left endpoint $a$, so that $I = [a, a + l]$, we define

$$M(I) = \left[ a, a + \frac{l}{3} \right] \cup \left[ a + \frac{2l}{3}, a + l \right]$$
Note that $M(I)$ is a union of two disjoint closed intervals of length $l/3$.

Now we begin: we’ll define the Cantor Set by an inductive process, and here’s the beginning: define

$$F_0 = I^0 = [0, 1],$$

and note that $F_0$ is a closed set comprising a single closed interval of length 1.

To make things a little clearer let’s do the next couple of steps explicitly, before describing the general inductive step: for the first step, define

$$F_1 = M(I^0) = \bigcup_{k=1}^{2} I^1_k,$$

a union of 2 disjoint closed intervals $I^1_1, I^1_2$ of length $1/3$.

Here’s the second step: define

$$F_2 = \bigcup_{k=1}^{2} M(I_k^1) = \bigcup_{k=1}^{2^2} I^2_k,$$

a union of $2^2$ disjoint closed intervals $I^2_1, \ldots, I^2_{2^2}$ of length $1/3^2$.

Note that $F_0, F_1, F_2$ are nonempty closed sets with

$$F_0 \supset F_1 \supset F_2.$$ 

Here’s the inductive step: let $n \in \mathbb{N}$, and suppose we have a nonempty closed set $F_n$ which is a union of $2^n$ disjoint closed intervals $I^n_1, \ldots, I^n_{2^n}$ of length $1/3^n$. Then define

$$F_{n+1} = \bigcup_{k=1}^{2^n} M(I^n_k) = \bigcup_{k=1}^{2^{n+1}} I^{n+1}_k,$$

a union of $2^{n+1}$ disjoint closed intervals $I^{n+1}_1, \ldots, I^{n+1}_{2^{n+1}}$ of length $1/3^{n+1}$.

Then $F_{n+1}$ is a nonempty closed set, and

$$F_n \supset F_{n+1}.$$ 

Now the Cantor Set is defined as:

$$C := \bigcap_{n=0}^{\infty} F_n.$$ 

$C$ is nonempty because $\{F_n\}_{n=1}^{\infty}$ is a family of closed subsets of the compact set $[0, 1]$ and has the finite intersection property.

Thus: the Cantor Set $C$ is a closed subset of $[0, 1]$, hence is compact and complete. We’ll explore some of the amazing properties of $C$ in exercises to come.
It might seem difficult to find actual elements of the Cantor set. Here are infinitely many:

**Exercise 13.10.** Let \( x \) be an endpoint of one of the closed intervals \( I_k^n \) used in the definition of the Cantor Set \( C \). Prove that \( x \in C \).

**Exercise 13.11.** Prove that the Cantor Set is nowhere dense. Since \( C \) is closed, you only have to show the interior \( C^\circ \) is empty. Hint: the lengths of the closed intervals comprising the sets \( F_n \) go to 0 as \( n \to \infty \).

Thus the Cantor Set is very “thin”. However, it’s in some ways quite substantial.

**Exercise 13.12.** Prove that the Cantor Set has no isolated points, that is, \( C' = C \). Hint: for all \( x \in C \) and all \( n \in \mathbb{N} \), one of the closed intervals \( I_1^n, \ldots, I_{2^n} \) contains \( x \).

It might seem at first glance that the Cantor Set can contain nothing else but the endpoints of the closed intervals \( I_k^n \). However:

**Exercise 13.13.** Prove that the Cantor Set is uncountable.

The Cantor Set is substantial in other ways, for example:

**Exercise 13.14.** Prove that \( C + C = [0, 2] \), where \( C + C := \{ x + y \mid x, y \in C \} \). Hint: let \( p \in [0, 2] \), and prove by induction that for all \( n \in \mathbb{N} \) there exist \( x_n, y_n \in F_n \) such that \( x_n + y_n = p \); for this you might want to prove a general lemma something like: if \( I \) and \( J \) are closed intervals of equal length then \( M(I) + M(J) = I + J \). Then you can pass to subsequences.

The following result is one of the many manifestations of: “continuous functions preserve many nice things”.

**Theorem 13.15.** A continuous image of a compact space is compact.

**Proof.** Let \( X \) and \( Y \) be metric spaces with \( X \) compact, and let \( f: X \to Y \) be continuous. Let \( (y_n) \) be a sequence in \( f(X) \). Then for each \( n \in \mathbb{N} \) there exists \( x_n \in X \) such that \( y_n = f(x_n) \). By compactness, there exists a subsequence \( (x_{n_k}) \) and an element \( x \in X \) such that \( x_{n_k} \to x \). Then \( y_{n_k} = f(x_{n_k}) \to f(x) \) by continuity. Thus the subsequence \( (y_{n_k}) \) converges in \( f(X) \), and we have shown \( f(X) \) is compact. \( \square \)

---

\(^{54}\)Recall that \( M(I) \) means \( I \) minus the open middle third.
The next couple of results show that continuity and compactness together are a kind of magic combination which yields many nice consequences. But first:

**Definition.** Let \( f : X \to \mathbb{R} \) and \( x \in X \). \( f \) has a **maximum** at \( x \) if \( f(x) = \max \text{ran } f \), and similarly for **minimum**.

**Notation and Terminology.** Let \( f \) be real-valued.

1. \( \max f := \max \text{ran } f \), and similarly for \( \min f \).
2. While we’re at it: \( \sup f := \sup \text{ran } f \). More generally, \( \sup_{x \in A} f(x) \). Similarly for \( \inf \).

**Extreme Value Theorem.** Every continuous real-valued function on a nonempty compact metric space has both a maximum and a minimum.

**Proof.** Let \( f : X \to \mathbb{R} \) be continuous, where \( X \) is nonempty and compact. Then \( \text{ran } f \) is a nonempty compact subset of \( \mathbb{R} \), hence has both a maximum and a minimum. \( \square \)

Here comes one of your favorite theorems from your calculus course:

**Corollary 13.16.** If \( f : [a, b] \to \mathbb{R} \) is continuous, then \( f \) has both a maximum and a minimum.

**Proof.** This follows immediately from the Extreme Value Theorem since \([a, b]\) is compact. \( \square \)

**Exercise 13.17.** Let \( f : \mathbb{R} \to \mathbb{R} \) be continuous, and assume

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

Prove that \( f \) has a minimum.

Here’s a variation on continuity that we’ll find very important:

**Definition.** \( f : X \to Y \) is **uniformly continuous** if for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that

\[
d(f(x), f(y)) < \epsilon \quad \text{whenever } \, d(x, y) < \delta.
\]

Note that \( f \) is continuous if and only if for all \( x \in X \) and \( \epsilon > 0 \) there exists \( \delta > 0 \) such that for all \( y \in X \) we have \( d(x, y) < \delta \) implies \( d(f(x), f(y)) < \epsilon \). Let’s symbolically compare the definitions of continuity and uniform continuity:

**continuity:**

\[
(\forall x \in X)(\forall \epsilon > 0)(\exists \delta > 0)(\forall y \in X) d(x, y) < \delta \implies d(f(x), f(y)) < \epsilon
\]

**uniform continuity:**

\[
(\forall \epsilon > 0)(\exists \delta > 0)(\forall x \in X)(\forall y \in X) d(x, y) < \delta \implies d(f(x), f(y)) < \epsilon
\]
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. Here’s an important situation in which the two kinds of continuity are actually equivalent:

**Theorem 13.18.** If $X$ is compact and $f: X \to Y$ is continuous, then $f$ is in fact uniformly continuous.

**Proof.** We argue by contradiction. Suppose not. Then we can choose $\epsilon > 0$ such that for all $\delta > 0$ there exist $x, y \in X$ such that $d(x, y) < \delta$ and $d(f(x), f(y)) \geq \epsilon$. In particular, for all $n \in \mathbb{N}$ there exist $x_n, y_n \in X$ such that

$$d(x_n, y_n) < \frac{1}{n} \quad \text{and} \quad d(f(x_n), f(y_n)) \geq \epsilon.$$ 

Choose a convergent subsequence $(x_{n_k})$ of $(x_n)$, and put $x = \lim x_{n_k}$. Since $d(x_{n_k}, y_{n_k}) \to 0$, we also have $y_{n_k} \to x$. Since $f$ is continuous,

$$f(x_{n_k}) \to f(x) \quad \text{and} \quad f(y_{n_k}) \to f(x).$$

But then

$$\epsilon \leq d(f(x_{n_k}), f(y_{n_k})) \to d(f(x), f(x)) = 0,$$

a contradiction. \hfill \Box

We could have avoided proof by contradiction in the above argument, but it would have turned out to require quite fussy manipulations with $\epsilon$’s and $\delta$’s, so the above proof was actually a little shorter and (I think) easier to follow.

We should say a few more words about a couple of the steps in the above proof. First, we asserted that $y_{n_k} \to x$ because $x_{n_k} \to x$ and $d(x_{n_k}, y_{n_k}) \to 0$. You should justify this:

**Exercise 13.19.** Let $(x_n)$ and $(y_n)$ be sequences in a metric space $X$, and assume $d(x_n, y_n) \to 0$. Prove that if $(x_n)$ converges, then so does $(y_n)$, with the same limit.

Second, we had $f(x_{n_k}) \to f(x)$ and $f(y_{n_k}) \to f(x)$, and we asserted $d(f(x_{n_k}), f(y_{n_k})) \to d(f(x), f(x))$; this is justified by Lemma 7.8.

**Exercise 13.20.** Prove that $x^2$ is not uniformly continuous on $\mathbb{R}$.

**Exercise 13.21.** (a) Prove that if $f, g: X \to \mathbb{R}$ are uniformly continuous, then $f + g$ is also uniformly continuous.

(b) Give an example where $fg$ is not uniformly continuous.
Exercises 13.22. Prove that if $f, g: X \to \mathbb{R}$ are bounded and uniformly continuous, then $fg$ is uniformly continuous.

Exercise 13.23. Let $f: [0, 1) \to \mathbb{R}$ be continuous, and assume
\[ \lim_{x \to 1} f(x) \]
exists. Prove that $f$ is uniformly continuous.

Exercise 13.24. Let $f: (0, 1) \to \mathbb{R}$ be uniformly continuous. Prove that $f$ is bounded.

Exercise 13.25. Prove that if $f: X \to Y$ is uniformly continuous and $(x_n)$ is a Cauchy sequence in $X$, then the image $(f(x_n))$ is also Cauchy.

Exercise 13.26. Let $A$ be dense in $X$ and $f: A \to Y$ uniformly continuous. Prove that if $Y$ is complete then there exists a uniformly continuous extension of $f$ to $X$.  


14. Connectivity

As we mentioned at the beginning of the preceding section, in your calculus course you saw a lot of results about closed bounded intervals. We said that most of the results relied upon compactness; in a lot of the other cases the crucial property was connectedness.

Example. Let \( A = (-\infty, 0), B = (0, \infty), \) and \( C = [0, \infty) \). Then \( A \) is disjoint from both \( B \) and \( C \), but there’s a difference: within \( A \), you can get close to a particular point in \( C \) (namely, 0), but you can’t get close to any point of \( B \). Roughly speaking, in some sense the sets \( A \) and \( B \) “disconnect” the subspace \( A \cup B \), but \( A \) and \( C \) don’t disconnect \( \mathbb{R} \). We need to formulate this precisely, in any metric space:

Definition. A metric space is connected if it’s not a union of two nonempty disjoint open sets.

Otherwise the metric space is disconnected.

Notation and Terminology. As before, letters such as \( X \) and \( Y \) will refer to metric spaces, unless otherwise specified.

What about subspaces? The definition would have us working with relatively open subsets, which is a little bit of a pain. Here’s a handy characterization of connectedness for subspaces:

Exercise 14.1. Let \( S \) be a subspace of a metric space \( X \). Prove that \( S \) is disconnected if and only if there exist nonempty subsets \( A, B \subset S \) such that \( S = A \cup B \) and

\[
A \cap \overline{B} = \overline{A} \cap B = \emptyset,
\]

where the closures are computed in the ambient space \( X \).

Exercise 14.2. Find all connected subsets of \( \mathbb{Z} \).

It’s easy to show a space is disconnected: all you have to do is find two nonempty disjoint complementary open sets. But how do we show a space is connected? To apply the definition, we’d have to assume we had two complementary open sets and use the properties of the particular metric space to deduce that at least one of the sets is empty. But there is also the general method of somehow finding a few connected spaces and showing how this implies other spaces are connected. Here’s a handy tool:

Proposition 14.3. Let \( \mathcal{F} \) be a family of subsets of \( X \). If each \( A \in \mathcal{F} \) is connected, and if \( \bigcap \mathcal{F} \neq \emptyset \), then \( \bigcup \mathcal{F} \) is connected.
**Proof.** Without loss of generality \( X = \bigcup \mathcal{F} \). Choose \( x \in \bigcap \mathcal{F} \). Suppose \( X = \overline{B} \cup \overline{C} \) for nonempty disjoint open sets \( B, C \). Without loss of generality \( x \in B \). For each \( A \in \mathcal{F} \), \( A = (A \cap B) \cup (A \cap C) \), a union of two disjoint open sets in \( A \). Since \( A \) is connected, \( A \cap B \) and \( A \cap C \) cannot both be nonempty. Since \( x \in A \cap B \), we must have \( A \cap C = \emptyset \). Then \( C = X \cap C = \bigcup_{A \in \mathcal{F}} (A \cap C) = \emptyset \), a contradiction. Therefore such \( B, C \) cannot exist, so \( X \) is connected. \( \square \)

And here’s another:

**Exercise 14.4.** Let \( A \) be a connected subset of a metric space \( X \), and let \( A \subset B \subset \overline{A} \). Prove that \( B \) is connected. Hint: you might find it helpful to do it first under the assumption that \( B = X \).

In a typical metric space, it’s impossible to visualize all the connected subsets. But in the real line it’s easy:

**Characterization of Connected Subsets of \( \mathbb{R} \).** A subset of \( \mathbb{R} \) is connected if and only if it’s an interval.

**Proof.** Let \( S \subset \mathbb{R} \). First assume \( S \) is not an interval. Then there exist \( x < y < z \) with \( x, z \in S \) and \( y \notin S \). Then

\[
S = (S \cap (-\infty, y)) \cup (S \cap (y, \infty)),
\]

a union of nonempty disjoint relatively open sets in \( S \), so \( S \) is disconnected.

Conversely, assume \( S \) is an interval. We must show \( S \) is connected. We argue by contradiction: suppose \( S \) is disconnected. Then there exist nonempty subsets \( A, B \subset S \) such that \( S = A \cup B \) and

\[
A \cap \overline{B} = \overline{A} \cap B = \emptyset.
\]

Pick \( a \in A \) and \( b \in B \), and without loss of generality \( a < b \). Since \( S \) is an interval containing both \( a \) and \( b \), it contains the closed interval \([a, b]\). Put \( C = A \cap [a, b] \). \( C \) is nonempty and bounded above, so \( t := \sup C \) exists, and \( a \leq t \leq b \). Then \( t \in \overline{C} \subset \overline{A} \), so \( t \notin B \). Since \( S = A \cup B \), we have \( t \in A \). Thus \( t < b \) since \( b \in B \). Since \( A \cap \overline{B} = \emptyset \), we have \( t \notin \overline{B} \). Thus there exists \( \epsilon > 0 \) such that \((t - \epsilon, t + \epsilon) \cap B = \emptyset \). Shrink \( \epsilon \) if necessary so that \( t + \epsilon < b \). Then \((t, t + \epsilon) \subset [a, b] \subset S \), and \((t, t + \epsilon) \cap B = \emptyset \), so \((t, t + \epsilon) \subset A \). Pick \( x \in (t, t + \epsilon) \). Then \( x \in C \) and \( x > t = \sup C \), which is a contradiction. \( \square \)
Why did we discard the open interval \((t - \epsilon, t + \epsilon)\) in favor of the smaller \((t, t + \epsilon)\) in the above proof? Because we didn’t know that \(t > a\), and we wanted an interval guaranteed to be contained in \([a, b]\). And besides, we only cared about the part to the right of \(t\) anyway.

**Exercise 14.5.** Find the connected subsets of \(\mathbb{Q}\).

And now a miracle occurs: in a typical metric space it’s impossible to visualize all open sets, but it turns out that in the real line there are so few directions to go that we can determine all the open sets:

**Characterization of Open Subsets of \(\mathbb{R}\).** Every open subset of \(\mathbb{R}\) is a countable union of pairwise disjoint open intervals.

**Proof.** Let \(A \subset \mathbb{R}\) be open. For each \(x \in A\) let \(I_x\) be the union of all the open intervals which contain \(x\) and are subsets of \(A\). Then \(I_x\) is an open interval since it’s a union of open connected sets with nonempty intersection. Moreover, \(A = \bigcup_{x \in A} I_x\). We show the family \(F := \{I_x \mid x \in A\}\) is pairwise disjoint. Let \(x, y \in A\), and assume \(I_x \cap I_y \neq \emptyset\). We must show \(I_x = I_y\), and by symmetry it suffices to show \(I_x \subset I_y\). Since \(I_x\) and \(I_y\) are open intervals with nonempty intersection, \(I_x \cup I_y\) is an open interval. Since \(y \in I_x \cup I_y\), we get \(I_x \cup I_y \subset I_y\), hence \(I_x \subset I_y\).

Finally, we show the family \(F\) is countable. \(\mathbb{R}\) is separable, so by Lindelöf’s Theorem the open cover \(F\) of \(A\) has a countable subcover. But since the family \(F\) is pairwise disjoint, no member of it can be deleted without changing the union. Therefore \(F\) must itself be countable. \(\square\)

Here’s an alternative argument showing the family \(F\) in the above proof is countable: since each \(I \in F\) is nonempty and open, it contains a rational number \(r_I\). Since \(F\) is pairwise disjoint, if \(I, J \in F\) with \(I \neq J\) then \(r_I \neq r_J\). Thus the map \(I \mapsto r_I\) from \(F\) to \(\mathbb{Q}\) is 1-1. Since \(\mathbb{Q}\) is countable, so is \(F\). This is a little more elementary, but our application of Lindelöf’s Theorem was more efficient.\(^{55}\)

Here’s another instance of “continuous functions preserving many nice things”:

**Theorem 14.6.** A continuous image of a connected set is connected.

**Proof.** Let \(f: X \rightarrow Y\) be continuous and \(X\) connected. Without loss of generality \(f\) is onto. Suppose \(Y = A \cup B\) for disjoint open sets \(A, B\). Then \(f^{-1}(A)\) and \(f^{-1}(B)\) are disjoint open sets with union \(X\), so by connectivity one of them is empty. Since \(f\) is onto, one of \(A = f(f^{-1}(A))\) or \(B = f(f^{-1}(B))\) is empty. Therefore \(Y\) is connected. \(\square\)

\(^{55}\)And it was also more “elegant” — this is a valuable quality in a proof.
**Intermediate Value Theorem.** If $X$ is connected and $f : X \to \mathbb{R}$ is continuous, then for all $x, y \in X$ and for all $t$ between $f(x)$ and $f(y)$, there exists $z \in X$ such that $f(z) = t$.

*Proof.* Since $X$ is connected, so is ran $f$ by continuity. Thus ran $f$ is an interval, so has the required property. \qed

Here comes another of your favorite theorems from your calculus course:

**Corollary 14.7.** If $f : [a, b] \to \mathbb{R}$ is continuous and $t$ is between $f(a)$ and $f(b)$, then there exists $z \in [a, b]$ such that $f(z) = t$.

*Proof.* This follows immediately from the Intermediate Value Theorem since $[a, b]$ is connected. \qed

Thus, a continuous image of an interval is an interval. For monotone functions, there is a sort of converse:

**Exercise 14.8.** Let $f : (a, b) \to \mathbb{R}$ be monotone, and suppose its range is dense in the interval with endpoints $f(a^+)$ and $f(b^-)$. Prove that $f$ is continuous.

In most applications of the preceding exercise, we happen to know that the range is an interval. There are similar versions for other types of intervals — it’s often used for closed intervals, for instance.

**Exercise 14.9.** Let $f : [0, 1] \to [0, 1]$ be continuous. Prove that there exists $x \in [0, 1]$ such that $f(x) = x$. Hint: consider $g(x) = f(x) - x$.

**Exercise 14.10.** Let $A$ be a nonempty, compact, and connected subset of $\mathbb{R}^2$. Prove that there exist $a, b \in \mathbb{R}$, with $a \leq b$, such that for each $c \in \mathbb{R}$ the line $x + y = c$ has nonempty intersection with $A$ if and only if $c \in [a, b]$.

**Exercise 14.11.** Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Suppose there exist sequences $(a_n)$ and $(b_n)$ diverging to $\infty$ such that $f(a_n) \to 0$ and $f(b_n) \to 4$. Prove that there exists a sequence $(c_j)$ such that $c_j \to \infty$ and $f(c_j) = 2$ for all $j$.

What about $\mathbb{R}^n$? Here we don’t have the luxury of knowing what all the connected subsets are (nor the open ones, for that matter), but at least we can say that $\mathbb{R}^n$ itself, and many other subsets, are connected:

**Definition.** A subset $A$ of $\mathbb{R}^n$ is *convex* if

$$(1 - t)x + ty \in A \quad \text{for all } x, y \in A, t \in [0, 1].$$

**Corollary 14.12.** Every convex subset of $\mathbb{R}^n$ (in particular, $\mathbb{R}^n$ itself) is connected.
Proof. Let \( A \subset \mathbb{R}^n \) be convex. If \( A = \emptyset \), then \( A \) is trivially connected, so assume \( A \neq \emptyset \), and pick \( x \in A \). Then for all \( y \in A \) put \( L_y = \{(1 - t)x + ty \mid t \in [0, 1]\} \). Then \( L_y \subset A \), and \( A = \bigcup_{y \in A} L_y \). Each \( L_y \) is connected since it’s the image of the connected subset \([0, 1]\) of \( \mathbb{R} \) under the continuous function \( t \mapsto (1 - t)x + ty \). Since \( x \in \bigcap_{y \in A} L_y \), we conclude that \( A \) is connected. \( \square \)

The essential idea of the above proof was the recognition that a convex set is a union of “closed intervals” with one endpoint fixed.

In the case \( n = 2 \), here’s another source of connected subsets: Let \( I \) be an interval, and let \( f : I \to \mathbb{R} \) be continuous. Then the graph of \( f \), namely the subset \( \{(x, f(x)) \mid x \in I\} \) of \( \mathbb{R}^2 \), is connected, since the function \( x \mapsto (x, f(x)) : I \to \mathbb{R}^2 \) is continuous. What if \( f \) is discontinuous? Could the graph still be connected? Surprisingly, yes: Define \( f : (0, 1] \to \mathbb{R} \) by \( f(x) = \sin(1/x) \), and let \( A \) be the graph of \( f \). Then of course \( A \) is connected since \( f \) is continuous. But if we pick our favorite number \( c \) in the interval \([-1, 1]\) and extend \( f \) to a function \( g \) on the closed interval \([0, 1]\) by defining \( g(0) = c \), then the graph of this extension \( g \) is still connected, although \( g \) is discontinuous at 0 (no matter what we pick for \( c \), because \( \lim_{x \to 0} g(x) \) does not exist). You’ll justify this in the following exercise:

**Exercise 14.13.** Keep the above notation regarding the functions \( f \) and \( g \), and the graph \( A \) of \( f \).

(a) Prove that every point of the set \( \{(0, y) \mid -1 \leq y \leq 1\} \) is in the closure of \( A \).

(b) Show how part (a) implies that the graph of the function \( g \) is connected.

Here’s another manifestation of: “there aren’t many directions in which to go in the real line”:

**Theorem 14.14.** If \( f : [a, b] \to \mathbb{R} \) is continuous and 1-1, then:

(i) \( f \) is strictly monotone;

(ii) \( f([a, b]) \) is a closed bounded interval \([c, d]\);

(iii) \( f^{-1} : [c, d] \to [a, b] \) is strictly monotone the same way \( f \) is;

(iv) \( f^{-1} \) 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 \( a \leq x < y \leq b \). We must show \( f(x) < f(y) \).

First we show \( f(a) < f(y) \). Argue by contradiction: suppose \( f(a) > f(y) \) (note that this is the only way \( f(a) < f(y) \) can be false, since \( a < y \)
and \( f \) is 1-1). Then \( f(a) \) is between \( f(y) \) and \( f(b) \), and \( f \) is continuous on the interval \([y, b]\), so by the Intermediate Value Theorem there exists \( z \in [y, b] \) such that \( f(z) = f(a) \), which is a contradiction since \( a < z \) and \( f \) is 1-1. This argument only depended upon \( a < y \leq b \).

A similar argument applied to the interval \([a, y]\) and using \( a \leq x < y \) would show \( f(x) < f(y) \), as desired.

(ii). Since \([a, b]\) is compact and connected, so is \( f([a, b]) \), hence \( f([a, b]) \) is a closed bounded interval \([c, d]\).

(iii). Recall that we know by now that \( f \) is strictly monotone, and without loss of generality we’ve focused on the strictly-increasing case. 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). Just note that \( f^{-1}: [c, d] \to \mathbb{R} \) is monotone and its range is an interval, so \( f^{-1} \) must be continuous, by (a variation of) Exercise 14.8.

In the above proof we reduced the task of showing continuity at \( t \in \mathbb{R} \) to showing continuity “from the right”. This is elementary, and is similar to one-sided limits. Let’s formalize it:

**Definition.** Let \( X \) be a metric space, \( A \subset \mathbb{R}, t \in A \), and \( f: A \to X \). We say \( f \) is continuous from the right at \( t \) if the restriction
\[
f \mid (A \cap [t, \infty))
\]
is continuous, and similarly for continuous from the left.

**Exercise 14.15.** Let \( X \) be a metric space, \( A \subset \mathbb{R}, t \in A \), and \( f: A \to X \). Prove:

(a) \( f \) is continuous from the right at \( t \) if and only if for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that
\[
d(f(x), f(t)) < \epsilon \quad \text{whenever} \quad t \in A \text{ and } t \leq x < t + \delta.
\]

(b) Prove that \( f \) is continuous at \( t \) if and only if it’s continuous from both the right and left at \( t \).

The above theorem (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) \( f^{-1}: (c, d) \rightarrow (a, b) \) 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’s strictly monotone on an open interval.
15. Differentiation

The elementary theory of derivatives for functions of one real variable is pretty easy, and most of it will look quite familiar from your calculus course. We’ll develop the theory in a fairly efficient way, with not much comment along the way.

Standing Hypothesis. Throughout this section, all our functions \( f \) will satisfy \( \text{dom } f, \text{ran } f \subset \mathbb{R} \) unless otherwise specified.

Definition. 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 differentiable at \( t \) if \( f'(t) \) exists.

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

(iv) We also write
\[
\frac{d}{dx} f(x) = f'(x).
\]

Thus, by definition a differentiable function must have open domain.

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

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

Intuitively, we think of continuous functions as having “unbroken” graphs, and differentiable functions as having “smooth” graphs. Smooth is definitely better than unbroken:

Proposition 15.1. 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 \( t \). \( \square \)

Here’s the arithmetic of derivatives:

Proposition 15.2 (Arithmetic of Derivatives). 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 \(0 \not\in \text{ran} \, g\).

Proof. (i). We have
\[
\frac{(f + g)(x) - (f + g)(t)}{x - t} = \frac{f(x) - f(t)}{x - t} + \frac{g(x) - g(t)}{x - t}
\]
\[
\xrightarrow{x \to t} f'(t) + g'(t).
\]

(ii). We have
\[
\frac{(fg)(x) - (fg)(t)}{x - t} = \frac{f(x)g(x) + f(x)g(t) - f(t)g(t)}{x - t}
\]
\[
= \frac{f(x)g(x) - g(x) - f(t)g(t)}{x - t} + \frac{f(x) - f(t)}{x - t} g(t)
\]
\[
\xrightarrow{x \to t} f(t)g'(t) + f'(t)g(t),
\]
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 \((\frac{1}{g})(t) = \frac{-g'(t)}{g(t)^2}\):
\[
\frac{1}{g}(x) - \frac{1}{g}(t)
\]
\[
= \frac{g(t) - g(x)}{(x - t)g(x)g(t)} \xrightarrow{x \to t} \frac{-g'(t)}{g(t)^2},
\]
by continuity of \(g\) at \(t\). \(\square\)

**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
\[
\frac{d}{dx} x^n = \frac{d}{dx} x^{-k} = -kx^{-k-1} = -kx^{-k-1} = n x^{n-1}.
\]

**Example.** Thus, for example, every rational functional function is differentiable.
Exercise 15.3. Give an example of a function \( f: \mathbb{R} \to \mathbb{R} \) which is differentiable at only one point.

In calculus you used the derivative to get the “tangent line” to a graph — let’s formalize this:

**Definition.** If \( f \) is differentiable at \( a \), the linear approximation\(^{\text{56}} \) to \( f \) at \( a \) is the function

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

The linear approximation to \( f \) at \( a \) is the only linear function having the same value and derivative as \( f \) at \( a \).

The following is a technical tool which can be useful to verify differentiability:

**Continuity Characterization of Differentiability.** \( f \) is differentiable at \( t \) if and only if there exists a function \( q \) which is continuous at \( t \) and satisfies

\[
f(x) = f(t) + q(x)(x - t) \quad \text{for all } x \in \text{dom } f,
\]

in which case \( f'(t) = q(t) \).

**Proof.** For \( x \in \text{dom } f \setminus \{t\} \), Equation \((3)\) is equivalent to

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

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

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

is continuous at \( t \). Of course, in this case Equation \((3)\) holds for all \( x \in \text{dom } f \), and we also have \( q(t) = f'(t) \).

To show \( f \) is differentiable at \( t \), it suffices to verify Equation \((3)\) for \( x \neq t \).

Here comes the relation between derivatives and composition. It’s surprisingly tricky to prove — the “obvious” idea doesn’t work.

\(^{56}\)The adjective “linear” is used because the graph is a straight line; in linear algebra you study a slightly different concept of linear function. In one-variable calculus the term “linear function” refers to a polynomial of degree at most 1.
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. Using the Continuity Characterization of Differentiability, we can 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

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

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

$$(g \circ f)'(t) = r \circ f(t)q(t) = g'(f(t))f'(t).$$ \qed

We’ve seen that differentiable functions are continuous. However, their derivatives might themselves be discontinuous:

**Exercise 15.4.** Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 
  x^2 \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\
  0 & \text{if } x = 0.
\end{cases}$$

Prove:

(a) $f$ is differentiable;
(b) $f'$ is discontinuous at 0.

Here’s how you found extreme values in your calculus course:

**Critical Point Lemma.** Let $f : (a, b) \to \mathbb{R}$ 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 $x \in (a, t)$,

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

and letting $x \uparrow t$ we get $f'(t) \geq 0$. On the other hand,

$$\frac{f(x) - f(t)}{x - t} \leq 0 \quad \text{for all } x \in (t, b),$$

and letting $x \downarrow t$ we get $f'(t) \leq 0$. Therefore, we must have $f'(t) = 0$. \qed
Exercise 15.5. Let $f, g: (a, b) \to \mathbb{R}$ and $a < c < b$. Suppose $f \leq g$ on $(a, b)$, $f$ and $g$ are both differentiable at $c$, and $f(c) = g(c)$. Use the Critical Point Lemma to prove that $f'(c) = g'(c)$.

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, $f'(c) = 0$. □

As an immediate corollary of Rolle’s Theorem (but which is not really worth recording as one of our official results, although we’ll 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.

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. □

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 $g(x) = x$. □

Here’s the pecking order among Rolle’s Theorem, Cauchy’s Mean Value Theorem, and the Mean Value Theorem: the Mean Value Theorem is a special case of Cauchy’s Mean Value Theorem, which is obvious from the above proof. However, that being said, it’s a good idea to keep in mind that the Mean Value Theorem is used far more often than the more general Cauchy’s Mean Value Theorem. Rolle’s Theorem is a
special case of the Mean Value Theorem, although we actually used it to prove Cauchy’s Mean Value Theorem!

The Mean Value Theorem can be used to prove inequalities:

*Example.* Let \( f(x) = \log x - x + 1 \), and fix \( x > 1 \). Then \( f \) is continuous on \([1, x]\) and differentiable on \((1, x)\), and \( f(1) = 0 \), so by the Mean Value Theorem there exists \( c \in (1, x) \) such that
\[
 f(x) = f'(c)(x - 1).
\]

Since
\[
 f'(c) = \frac{1}{c} - 1 < 0
\]
because \( c > 1 \), we have \( f(x) < 0 \). Conclusion:
\[
 \log x < x - 1 \quad \text{for all } x > 1.
\]

**Exercise 15.6.** Use the Mean Value Theorem to prove that
\[
 \sqrt{1 + x} < 1 + \frac{x}{2} \quad \text{for all } x > 0.
\]

The Mean Value Theorem can also be used to control how quickly \( f \) changes values:

**Exercise 15.7.** Let \( f : (a, b) \to \mathbb{R} \) be differentiable. Prove that if \( f' \) is bounded, then \( f \) is uniformly continuous. Hint: Mean Value Theorem.

Here’s how you used derivatives\(^{57}\) in your calculus course to help graph functions:

**Proposition 15.8.** If \( f \) is continuous on \([a, b]\) and differentiable on \((a, b)\), 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) \( f' \neq 0 \) on \((a, b)\) implies \( f \) is strictly monotone on \([a, b]\);
(iv) \( f' \geq 0 \) on \((a, b)\) implies \( f \) is increasing on \([a, b]\);
(v) \( f' \leq 0 \) on \((a, b)\) implies \( f \) is decreasing on \([a, b]\);
(vi) \( f' = 0 \) on \((a, b)\) implies \( f \) is constant on \([a, b]\).

**Proof.** In each case the desired behavior follows immediately from the Mean Value Theorem (more precisely, in (iii) we get \( f \) is 1-1, hence strictly monotone by an earlier theorem). \( \square \)

In the above proposition, we could replace \([a, b]\) by \((a, b)\) throughout (and the continuity hypothesis could then be deleted.

*Example.* If \( n \in \mathbb{N} \) then the function \( x \mapsto x^n \) is strictly increasing on \((0, \infty)\) if \( n \) is even, and on \( \mathbb{R} \) if \( n \) is odd.

\(^{57}\)well, the first derivative, anyway
**Example.** Using only mental arithmetic, try to work out which is bigger: $\pi^e$ or $e^\pi$. Give up? Define $f(x) = (\log x)/x$. Then $f'(x) = (1 - \log x)/x^2$, which is 0 when $x = e$. Since $f'$ changes sign from positive to negative there, $f(e) = 1/e$ is the maximum value of $f$. Now, we know $\pi > e$, so $f(\pi) < f(e)$, that is, $\frac{\log \pi}{\pi} < \frac{1}{e}$, or $e \log \pi < \pi$, hence $\pi^e < e^\pi$.

The above results, together with the theorem and discussion at the end of the preceding lecture, can be used to prove the existence of roots (see the discussion at the end of Lecture 3); for example:

**Exercise 15.9.** Prove that for all $x \geq 0$ there exists a unique $\sqrt{x} \geq 0$ such that $\sqrt{x}^2 = x$, and more over the function $x \mapsto \sqrt{x}$ is continuous and strictly increasing. Hint: show that the function $f : [0, \infty) \to [0, \infty)$ defined by $f(x) = x^2$ is strictly increasing, continuous, and onto.

More generally, the same sort of arguments can be used to show that the $n$-th root function $x \mapsto \sqrt[n]{x}$ is continuous and strictly increasing on $[0, \infty)$ if $n$ is even and on all of $\mathbb{R}$ if $n$ is odd.

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

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

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

**Exercise 15.10.** Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^3 + 3x^2 - 36x + 5$.

(a) Prove that $f$ is 1-1 on the interval $[-1, 1]$.

(b) Find $(f^{-1})'(5)$.

The above theorem can be used to extend the Power Rule to rational exponents:

**Power Rule.** $\frac{d}{dx} x^r = rx^{r-1}$ for all $r \in \mathbb{Q}$.

---

58They’re pretty close — the difference is about 0.7.
Proof. We already know it’s true for \( r \in \mathbb{N} \). Of course, \( x \) must be suitably restricted, depending upon the exponent \( r \): if \( r < 1 \) we avoid \( x = 0 \), and if the denominator of \( r \) is even we avoid \( x \leq 0 \).

You will prove it for \( x > 0 \) in the exercise following this proof.

Similar arguments would work for \( r \) with odd denominator and \( x < 0 \).

Finally, if \( r > 1 \) with odd denominator then the derivative at \( x = 0 \) is

\[
\lim_{t \to 0} \frac{t^r}{t} = \lim_{t \to 0} t^{r-1} = 0 = r x^{r-1}|_{x=0}.
\]

\( \square \)

Now you must do your duty and fill in the (advertised) gap in the above proof:

**Exercise 15.11.** Establish the Power Rule \( \frac{d}{dx} x^r = rx^{r-1} \) for \( x > 0 \) as follows:

(a) When \( r = 1/n \) for \( n \in \mathbb{N} \), use the Inverse Function Theorem.

(b) When \( r = k/n \) for \( k, n \in \mathbb{N} \), use (a) and the Chain Rule.

Here comes l’Hôpital’s Rule, an extremely handy tool for finding limits involving the “indeterminate forms” 0/0 and \( \infty/\infty \). The proof of the second case is surprisingly fussy:

**L’Hôpital’s Rule.** Let each of \( a \) and \( u \) be either a real number or \( \pm \infty \). Assume that \( f \) and \( g \) are both differentiable on either open interval with one endpoint \( a \), or an open interval with \( a \) removed, and that either \( \lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0 \) or \( \lim_{x \to a} g(x) = \pm \infty \). If

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

then

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

Proof. This one involves quite fussy manipulations with inequalities, so it behooves us to simplify as much as possible. So, we’ll only do a representative part of a special case; the other parts and other possible cases are similar.

We’ll only deal with one side of \( a \) (and remember that \( a \) could be \( \pm \infty \) — our proof won’t care). Specifically, we’ll deal with the right-hand limit, assuming \( f \) and \( g \) are differentiable on some open interval with left endpoint \( a \).

Also, if for example \( u \) is a real number, then we would have to show we can get \( f/g \) within some \( \epsilon \) of \( u \); we’ll only show we can make it less than \( u + \epsilon \), which is representative. But actually \( u \) could be \( -\infty \), so to
MAT 472 LECTURES

get an argument which works even in this case we’ll instead show we can make \( f/g \) less than any real number which is greater than \( u \).

So, without loss of generality it suffices to show that if \( u < v \) then there exists \( b > a \) such that

\[
\frac{f(x)}{g(x)} < v \quad \text{for all } x \in (a, b).
\]

Pick \( s \in (u, v) \), and choose \( b_1 > a \) such that

\[
\frac{f'(x)}{g'(x)} < s \quad \text{for all } x \in (a, b_1).
\]

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

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

Case 1. \( \lim f = \lim g = 0 \). Letting \( x \downarrow a \) in \((4)\) gives

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

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

\[
g(x), g(x) - g(y) > 0 \quad \text{for all } x \in (a, b_2).
\]

Multiplying \((4)\) by \( \frac{g(x) - g(y)}{g(x)} \) gives

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

so

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

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

Similarly for \( \lim g = -\infty \), \( \lim_{x \uparrow a} \frac{f(x)}{g(x)} < v \) or \( v < u \). \( \square \)

It’s 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.

Example. l’Hôpital’s Rule gives us the asymptotic behavior of rational functions: let \( f(x) = \sum_0^n a_n x^n \) and \( g(x) = \sum_0^k b_k x^k \), with \( a_n, b_k \neq 0 \). Then applying l’Hôpital’s Rule \( \min\{n, k\} \) times shows that

\[
\lim_{x \to \pm\infty} \frac{f(x)}{g(x)} = \lim_{x \to \pm\infty} \frac{a_n}{b_k} x^{n-k}
\]
Of course, this limit could be computed using other methods; l’Hôpital’s Rule is a unifying tool.

Example. l’Hôpital’s Rule gives
\[ \lim_{{x \to 0}} \frac{\sin x}{x} = 1, \]
although in fact this follows from \( \cos 0 = 1 \), since \( \cos = \sin' \). In general, if \( f'(a) \) exists, then \( \lim_{{x \to a}} \frac{f(x) - f(a)}{x - a} = f'(a) \) by definition; l’Hôpital’s Rule could be used if \( f' \) is continuous at \( a \). The next exercise gives a useful variation on this.

Exercise 15.12. Let \( f: (a, b) \to \mathbb{R} \) and \( t \in (a, b) \). Suppose \( f \) is continuous at \( t \) and differentiable on \((a, b) \setminus \{t\}\), and \( \lim_{{x \to t}} f'(x) \) exists. Prove:

(a) \( f \) is differentiable at \( t \).
(b) \( f' \) is continuous at \( t \).

Thus, although derivatives need not be continuous, they can’t have any “removable” discontinuities. They also can’t have “jump” discontinuities, since a similar argument shows that if \( f: (a, b) \) is differentiable, \( t \in (a, b) \), and \( f'(t+) \) exists, then in fact \( f'(t) = f'(t+) \), and similarly from the left.

They also possess an intermediate value property:

Exercise 15.13. Prove Darboux’s Intermediate Value Theorem for Derivatives: if \( f: (a, b) \to \mathbb{R} \) is differentiable and \( a < s < t < b \), then for every \( m \) strictly between \( f'(s) \) and \( f'(t) \) there exists \( c \in (s, t) \) such that \( f'(c) = m \). Hint: first assume \( m = 0 \); for the general case consider \( g(x) = f(x) - mx \).

Exercise 15.14. Define \( f: \mathbb{R} \to \mathbb{R} \) by
\[
 f(x) = \begin{cases} 
 \frac{1}{1+e^{1/x^2}} & \text{if } x \neq 0 \\
 0 & \text{if } x = 0 
\end{cases}
.
\]
Find \( f'(0) \).

l’Hôpital’s Rule is effective on many “exponential” indeterminate forms:

Example. Consider \( \lim_{{x \to 0}} x^x \), of the form “0^0”. We have
\[
 \log x^x = x \log x = \frac{\log x}{1/x},
\]
giving the indeterminate form $\infty/\infty$. Apply l'Hôpital's rule to this:

$$\frac{d}{dx}(\log x) = \frac{1}{x}, \quad \frac{d}{dx}(1/x) = -1/x^2$$

Thus $x^x \to 0$ by l'Hôpital’s Rule, so

$$x^x = e^{\log x^x} \xrightarrow{x \to 0} e^0 = 1$$

by continuity.

Here are some more standard applications of l'Hôpital’s Rule:

**Exercise 15.15.** Show that

$$\lim_{x \to \infty} (1 + \frac{t}{x})^x = e^t \quad \text{for all } t \in \mathbb{R}.$$  

**Exercise 15.16.** Show that the sequence $(n^{1/n})$ converges to 1 by showing that the function $x^{1/x}$ goes to 1 as $x \to \infty$.

**Example.** It follows from the preceding exercise that $x^{1/n} \to 1$ for any $x \geq 1$, but in fact it’s easier to verify this limit directly, and even for all $x > 0$: $(\log x)/n \to 0$, so $x^{1/n} \to e^0 = 1$.

**Exercise 15.17.** Prove that if $f$ is any rational function then

$$\lim_{x \to \infty} f(x)e^{-x} = 0.$$  

**Higher Order Derivatives:**

**Definition.**

(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

$$\frac{d^n}{dx^n} f(x) = f^{(n)}(x),$$

and say $f$ is $n$-times differentiable at $a$ if $f^{(n)}(a)$ exists.

By definition, a function which is $n$-times differentiable at $a$ is $(n-1)$-times differentiable in some open interval containing $a$.

**Exercise 15.18.** Let $f : (a, b) \to \mathbb{R}$ and $t \in (a, b)$, and assume $f''(t)$ exists. Prove that

$$f''(t) = \lim_{h \to 0} \frac{f(t + h) - 2f(t) + f(t - h)}{h^2}.$$  

Hint: l'Hôpital’s Rule.
In calculus you used the second derivative to describe the “concavity” of the graph. The general idea was that if $f''$ is positive then the graph of $f$ is “concave up”\footnote{actually called “convex” in more advanced texts} meaning the graph lies above its tangent line. Actually, slightly more is true:

**Exercise 15.19.** Let $f''(a) > 0$. Prove that there exist $c, \delta > 0$ such that

$$f(x) \geq f(a) + f'(a)(x - a) + c(x - a)^2 \quad \text{if } |x - a| < \delta.$$  

Hint: use l’Hôpital’s Rule to show that the fraction

$$\frac{f(x) - f(a) - f'(a)(x - a)}{(x - a)^2}$$

has a positive limit at $a$. 
16. Integration

In this section we cover the elementary theory of the Riemann integral. You should really be learning the Lebesgue integral, because it’s much more powerful. On the other hand the Riemann integral in one real variable is much easier to develop, and you need to be familiar with it anyway, because it appears so often in mathematical literature. And besides, here we’ll rigorously justify all the results about integration you saw in your calculus course.

Standing Hypothesis. Throughout this section, all our functions $f$ will satisfy $\text{dom} f, \text{ran} f \subset \mathbb{R}$ unless otherwise specified.

It’ll take a little while to get to the definition of the integral — there are lots of preparations to get out of the way. But it’ll be worth it.

Partitions. Let $-\infty < a < b < \infty$.

(i) A partition of $[a, b]$ is 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) If $P = \{x_i\}_{0}^{n} \in \mathcal{P}[a, b]$, the norm of $P$ is

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

where

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

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

Thus, a partition of an interval is just a finite subset which includes the endpoints.

Upper and lower sums. Let $f$ be bounded on $[a, b]$ and $P = \{x_i\}_{0}^{n} \in \mathcal{P}[a, b]$.

(i) The upper sum of $f$ associated to $P$ is

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

where

$$M_i = M_i(f) := \sup_{[x_{i-1}, x_i]} f.$$

---

60 Actually, in other areas of mathematics, the words “partition” and “norm” have completely different meanings.
(ii) The lower sum of $f$ associated to $P$ is

$$L(P) = L(f, P) := \sum_{1}^{n} m_i \Delta x_i,$$

where

$$m_i = m_i(f) := \inf_{[x_{i-1}, x_i]} f.$$

**Exercise 16.1.** Let $f$ be bounded on $[a, b]$ and $P = \{x_i\}_{i=0}^{n} \in \mathcal{P}[a, b]$. Prove that

$$M_i - m_i = \sup_{x,y \in [x_{i-1}, x_i]} |f(x) - f(y)|$$

for all $i = 1, \ldots, n$.

**Lemma 16.2.**

(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 that $Q$ contains one more element than $P$, so that there exists $s \notin P$ such that $Q = P \cup \{s\}$. Then there exists a unique $i$ such that $x_{i-1} < s < x_i$. 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

$$L(P) \leq L(R) \leq U(R) \leq U(Q). \quad \square$$

**The integral.** A bounded function $f$ on $[a, b]$ 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 the Riemann integral of $f$ on $[a, b]$.

**Notation and Terminology.** $\int_{a}^{b} f$ or $\int_{a}^{b} f(x) \, dx$ denote the Riemann integral, $f$ is the integrand, $a$ the lower limit of integration, and $b$ the upper limit of integration.
Exercise 16.3. Prove that $\int_a^b 1 = b - a$.

It follows immediately from Lemma 16.2 that

$$-\infty < \sup_P L(P) \leq \inf_P U(P) < \infty.$$ 

**Definition.** Let $f$ be bounded on $[a, b]$ and $P = \{x_i\}_0^n \in \mathcal{P}[a, b]$. 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 a Riemann sum for $f$ associated to $P$. We denote the set of all Riemann sums for $f$ associated to $P$ by $\mathcal{R}(P) = \mathcal{R}(f, P)$.

Remember: Riemann sums, just like upper and lower sums, are numbers, so $\mathcal{R}(P)$ is a subset of $\mathbb{R}$.

For all $S \in \mathcal{R}(P)$, $L(P) \leq S \leq U(P)$. 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 16.4.** Let $f$ be bounded on $[a, b]$ and $P = \{x_i\}_0^n \in \mathcal{P}[a, b]$. 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

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

Then

$$\sum_{i=1}^n f(t_i) \Delta x_i > U(P) - \frac{\epsilon}{b - a} \sum_{i=1}^n \Delta x_i = U(P) - \epsilon,$$

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 16.5.** If \( f \) is bounded on \([a, b]\), then the following are equivalent:

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

(ii) **Darboux’s Theorem.** There exists \( I \in \mathbb{R} \) such that for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that

\[
|S - I| < \epsilon \quad \text{whenever} \quad \|P\| < \delta, S \in \mathcal{R}(P).
\]

(iii) There exists \( I \in \mathbb{R} \) such that for all \( \epsilon > 0 \) there exists \( P \in \mathcal{P}[a, b] \) such that

\[
|S - I| < \epsilon \quad \text{for all} \quad S \in \mathcal{R}(P).
\]

(iv) **Riemann’s Criterion.** For all \( \epsilon > 0 \) there exists \( P \in \mathcal{P}[a, b] \) such that \( U(P) - L(P) < \epsilon \).

More precisely, Darboux’s Theorem is the equivalence (i) \( \iff \) (ii), and Riemann’s Criterion is the implication (iv) \( \implies \) (i). In (ii), from the context it’s clear that we also restrict \( P \in \mathcal{P}[a, b] \). In (ii)–(iii), of course \( I = \int_a^b f \).

**Proof.** (i) implies (ii). This is the hard part. 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 \( M = \sup_{[a,b]} |f|, \quad P = \{x_i\}_{i=0}^n \in \mathcal{P}[a,b], \quad s \in [a,b], \) and \( Q = P \cup \{s\} \), then

\[
U(P) - U(Q) \leq 2M\|P\|.
\]
Let \( x_{i-1} < s < x_i \), and put

\[
M' = \sup_{[x_{i-1}, s]} f, \quad M'' = \sup_{[s, x_i]} f, \quad \Delta x'_i = s - x_{i-1}, \quad \text{and} \quad \Delta x''_i = x_i - s.
\]
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) implies (iv). Let \( \epsilon > 0 \). Choose \( P \in \mathcal{P}[a,b] \) such that \( |S - I| < \epsilon/3 \) for all \( S \in \mathcal{R}(P) \). Then \( |S - T| < 2\epsilon/3 \) for all \( S, T \in \mathcal{R}(P) \). Thus
\[ U(P) - L(P) = \sup_{S,T \in \mathcal{R}(P)} |S - T| \leq \frac{2\epsilon}{3} < \epsilon. \]

(iv) implies (i). Assume \( f \) is not integrable. Then
\[ \epsilon := \inf_P U(P) - \sup_P L(P) > 0, \]
and for every \( P \in \mathcal{P}[a,b] \),
\[ U(P) - L(P) \geq \epsilon. \]
We have shown that (iv) is false if (i) is, which is the contrapositive of the desired implication.

In the proof that (iii) implies (iv), why did we choose to make \( |S - T| < \epsilon/3 \)? We chose \( \epsilon/3 \) in our intermediate estimate so that we would end up with the desired quantity strictly less than \( \epsilon \), not just less than or equal. There was nothing magical about the fraction \( 1/3 \), but since we took a supremum of estimates, the strict inequalities become weak ones, and we wanted to make sure we got something less than or equal to a multiple of \( \epsilon \) which was in turn strictly less than \( \epsilon \).

In Darboux’s Theorem, the forward implication (i) \( \implies \) (ii) is the “deep” part, and shows that the integral is a certain kind of “limit of Riemann sums as the norm of the partitions goes to 0” — but we won’t
make this precise. However, it’s worth mentioning that Darboux’s Theorem is the basis for many interesting limits of the following form: if $f$ is integrable on $[a, b]$, then

$$\lim_{n \to \infty} \frac{b-a}{n} \sum_{i=1}^{n} f\left(a + \frac{b-a}{n} i\right) = \int_{a}^{b} f.$$

When we want to show a function is integrable and its integral is a certain value, we don’t use the full power (ii) $\implies$ (i) of Darboux’s Theorem; instead the implication (iii) $\implies$ (i) in the above theorem is more suitable — it’s enough to find a single partition for which all Riemann sums are sufficiently close to the desired value.

**Exercise 16.6.** Let $f$ be integrable on $[a, b]$. Let $(P_n)$ be a sequence of partitions of $[a, b]$ such that $\|P_n\| \to 0$, and for each $n \in \mathbb{N}$ let $S_n$ be a Riemann sum for $f$ associated to the partition $P_n$. Prove that $S_n \to \int_{a}^{b} f$.

Interestingly, Riemann’s Criterion makes no use of Riemann sums. Riemann’s Criterion is frequently used when all we want to show is that a function is integrable, but don’t care about it’s value. For example:

**Theorem 16.7.** 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. Let $\epsilon > 0$. Choose $\delta > 0$ such that $|f(x) - f(y)| < \frac{\epsilon}{2(b-a)}$ whenever $|x - y| < \delta$.

Now choose $P = \{x_i\}_{0}^{n} \in \mathcal{P}[a, b]$ such that $\|P\| < \delta$. Then

$$M_i - m_i \leq \frac{\epsilon}{2(b-a)}$$

for all $i$,

so

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

$$= \frac{\epsilon(b-a)}{2(b-a)} = \frac{\epsilon}{2} < \epsilon.$$ 

Therefore $f$ is integrable, by Riemann’s Criterion. □

Again, in the above proof we chose an intermediate upper bound of $\epsilon/(2(b-a))$ because we knew we would be taking a supremum, so the best we could hope for was a weak inequality, and we wanted the upper bound at that point to be strictly less than $\epsilon$ — in fact, it was $\epsilon/2$. 

It’s important to keep in mind that the above result is not an if and only if: an integrable function need not be continuous. Let’s begin to explore this a little:

**Exercise 16.8.** Let $f: [a, b] \to \mathbb{R}$ be bounded, and assume $f$ is integrable on $[a, t]$ for every $t \in (a, b)$. Prove that $f$ is integrable on $[a, b]$.

**Observation.** A similar result is true for the lower limit: if $f: [a, b] \to \mathbb{R}$ is bounded, and integrable on $[t, b]$ for every $t \in (a, b)$, then $f$ is integrable on $[a, b]$.

Thus, if $f$ is bounded on $[a, b]$ and continuous except perhaps at one endpoint, then it’s integrable. For example, the function $f$ taking values $\sin(1/x)$ for $x \neq 0$ and 0 at 0 is integrable\(^{61}\) on $[0, 1]$. However, the function can’t be too discontinuous:

**Exercise 16.9.** Prove that 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.

But a certain amount of discontinuity can be allowed under the right conditions, for example:

**Exercise 16.10.** Prove that if $f: [a, b] \to \mathbb{R}$ is monotone, then it is integrable.

Here’s the arithmetic of integrals:

**Proposition 16.11.** 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). Let $\epsilon > 0$. Choose $P \in \mathcal{P}[a, b]$ such that if $t_i \in [x_{i-1}, x_i]$ for all $i$ then

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

\(^{61}\)although we it would be difficult to compute the integral!
To see that this is possible, by Darboux’s Theorem we can choose \( \delta_1, \delta_2 > 0 \) such that
\[
|S - \int_a^b f| < \epsilon/2 \quad \text{whenever} \quad \|P\| < \delta_1, S \in \mathcal{R}(f, P)
\]
\[
|T - \int_a^b g| < \epsilon/2 \quad \text{whenever} \quad \|P\| < \delta_2, T \in \mathcal{R}(g, P),
\]
and then we can take any \( P \in \mathcal{P}[a, b] \) with \( \|P\| < \min{\delta_1, \delta_2} \). Then
\[
\left| \sum (f + g)(t_i) \Delta x_i - \left( \int_a^b f + \int_a^b g \right) \right| \\
\leq \left| \sum f(t_i) \Delta_i - \int_a^b f \right| + \left| \sum g(t_i) \Delta_i - \int_a^b g \right| < \epsilon.
\]
(ii). Let \( \epsilon > 0 \). 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
\[
\left| \sum cf(t_i) \Delta x_i - c \int_a^b f \right| = |c| \left| \sum f(t_i) \Delta x_i - \int_a^b f \right| \\
\leq \frac{|c| \epsilon}{|c| + 1} < \epsilon. \quad \square
\]

In the proof of (ii) above, again we chose an intermediate estimate so that the final estimate would have \( \epsilon \) as a strict upper bound. Why did we use \( |c| + 1 \) rather than just \( |c| \)? Well, \( c \) could have been 0, and we had to avoid division by 0; we could have dispensed with the case \( c = 0 \) easily enough, but it was just as fast to use the above trick.

Also note that we found it convenient to use Riemann sums throughout the above proof, rather than upper and lower sums.

**Proposition 16.12.** 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.** We have to prove two things: an if and only if statement, and an equation. First we’ll prove the forward direction of the if and only if: assume \( f \) is integrable on \( [a, c] \). Let \( \epsilon > 0 \). Choose \( P \in \mathcal{P}[a, c] \) such that \( U(P) - L(P) < \epsilon \). Without loss of generality \( b \in P \), since

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

$$U(P') - L(P'), U(P'') - L(P'') < \epsilon.$$

Now we'll prove in one whack the converse direction of the if and only if statement and the desired equation: assume $f$ is integrable on both $[a, b]$ and $[b, c]$. Let $\epsilon > 0$. Choose $P' \in \mathcal{P}[a, b]$, $P'' \in \mathcal{P}[b, c]$ such that

$$\left| S' - \int_a^b f \right| < \frac{\epsilon}{2} \quad \text{for all } S' \in \mathcal{R}(P')$$

$$\left| S'' - \int_b^c f \right| < \frac{\epsilon}{2} \quad \text{for all } S'' \in \mathcal{R}(P'').$$

Put $P = P' \cup P''$, and let $S \in \mathcal{R}(P)$. Then $P \in \mathcal{P}[a, c]$, and moreover, since $b \in P$, there exist $S' \in \mathcal{R}(P')$, $S'' \in \mathcal{R}(P'')$ such that $S = S' + S''$, 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 16.5 tells us $f$ is integrable on $[a, c]$, with $\int_a^c f = \int_a^b f + \int_b^c f$. \hfill $\square$

In the above proof we used Riemann’s Criterion for the first part, and Riemann sums for the second part. Remember, if we only have to prove a function is integrable, Riemann’s Criterion is usually the way to go, but if we also have to prove that the integral has a certain value, then Riemann sums are probably better.

**Observation.** We’ve seen before that if $f : [a, b] \to \mathbb{R}$ is bounded, and has a single discontinuity at one of the endpoints, then it’s integrable; the above result and induction show that this is still ok if we allow finitely many discontinuities, anywhere in the interval.

Now we see that integrals preserve weak inequalities — unsurprisingly, since an integral is a kind of limit:

**Proposition 16.13.** 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 $P \in \mathcal{P}[a, b]$, $\int_a^b f \leq U(f, P) \leq U(g, P)$. Thus $\int_a^b f \leq \int_a^b g$. \hfill $\square$

**Exercise 16.14.** Let $f : [a, b] \rightarrow [0, \infty)$ be continuous, and suppose there exists $K \in \mathbb{R}$ such that $f(x) \leq K \int_a^x f$ for all $x \in [a, b]$. Prove that $f(x) = 0$ for all $x \in [a, b]$. Hint: integrate the inequality $f \leq \max f$, and choose a sufficiently fine partition.

**Exercise 16.15.** Let $f : [a, b] \rightarrow [0, \infty)$ be continuous, and assume $\max f = 1$. Prove that $\left( \int_a^b f^n \right)^{1/n} \rightarrow 1$.

The following result is a kind of “continuous triangle inequality”:

**Proposition 16.16.** If $f$ is integrable on $[a, b]$, then so is $\lvert f \rvert$, and

$$\left\lvert \int_a^b f \right\rvert \leq \int_a^b \lvert f \rvert.$$ 

**Proof.** Let $\epsilon > 0$. 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

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

so

$$M_i(\lvert f \rvert) - m_i(\lvert f \rvert) \leq M_i(f) - m_i(f).$$

Hence

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

Thus $\lvert f \rvert$ is integrable.

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

$$-\int_a^b \lvert f \rvert \leq \int_a^b f \leq \int_a^b \lvert f \rvert.$$ 

Hence

$$\left\lvert \int_a^b f \right\rvert \leq \int_a^b \lvert f \rvert.$$ 

\hfill $\square$

**Proposition 16.17.** If $f$ and $g$ are integrable on $[a, b]$, then so is $fg$. 

---

**Proof.** For every $P \in \mathcal{P}[a, b]$, $\int_a^b f \leq U(f, P) \leq U(g, P)$. Thus $\int_a^b f \leq \int_a^b g$. \hfill $\square$
Proof. We use a trick: it’s pretty easy for $f^2$, and then a trick gives the general case:

Case 1. $f = g$. Put $M = \sup |f|$, and without loss of generality $M > 0$. Let $\epsilon > 0$. Choose $P = \{x_i\}^n_1 \in \mathcal{P}[a,b]$ 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. □

In the above proof we used the device of saying “without loss of generality $M > 0$” because we wanted to divide by it; as we’ve seen before, an alternative would be to use $\epsilon/(2M + 1)$ in the intermediate estimate. Again, the ultimate goal was to get something strictly less than $\epsilon$.

Also in the above proof, we referred to $\sup |f|$ — from the context it was implicit that we were taking the sup over $[a,b]$.

Finally, observe that the manipulations with $M(f^2), \ldots$ were similar to those with $M(|f|), \ldots$ in an earlier proof.

After looking at the proof of the following result, it might seem more appropriate to call it the “Intermediate Value Theorem for Integrals”; we’ll comment later why “Mean Value” is used instead.

**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

$$\int_a^b fg = f(c) \int_a^b g.$$

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. \( \square \)

The Mean Value Theorem for Integrals is most often used for the case \( g = 1 \): if \( f \) is continuous on \( [a, b] \) then there exists \( c \in [a, b] \) such that \( \int_{a}^{b} f = f(c)(b - a) \). This may be interpreted as saying that the integral actually coincides with a particularly simple Riemann sum.

In the above proof, it’s clear from the context that \( \min f \) and \( \max f \) refer to the \( \min \) and \( \max \) over \( [a, b] \).

The following result shows that integrating “smoothes” functions — integrable functions are bounded, but need not be continuous, however their integrals are continuous:

**Theorem 16.18.** If \( f \) is integrable on \( [a, b] \), then the function \( x \mapsto \int_{a}^{x} f \) is continuous on \( [a, b] \).

**Proof.** Put \( M = \sup |f| \). Let \( \epsilon > 0 \). Choose \( \delta > 0 \) such that \( M\delta < \epsilon \). Then for all \( a \leq y \leq x \leq b \) and \( x - y < \delta \) we have:
\[
\left| \int_{a}^{x} f - \int_{a}^{y} f \right| = \left| \int_{y}^{x} f \right| \leq \int_{y}^{x} |f| \leq M(x - y) < \epsilon.
\]

Here’s a useful application of the above theorem:

**Observation.** Let \( f : [a, b] \to \mathbb{R} \) be bounded. Recall that if \( f \) is integrable on \( [a, t] \) for every \( t \in (a, b) \), then \( f \) is integrable on \( [a, b] \): we can now say what the integral is:
\[
\int_{a}^{b} f = \lim_{t \uparrow b} \int_{a}^{t} f.
\]

Similarly at the left endpoint: if \( f \) is integrable on \( [t, b] \) for every \( t \in (a, b) \), then
\[
\int_{a}^{b} f = \lim_{t \downarrow a} \int_{t}^{b} f.
\]
Example. Suppose $f$ is identically $c$ on $[a, b)$. Then

$$\int_a^b f = \lim_{t \uparrow b} \int_a^t c = \lim_{t \uparrow b} c(t - a) = c(b - a).$$

Similarly if $f$ is $c$ on $(a, b]$. If we only know $f$ is $c$ on $(a, b)$ (but we have no knowledge of the values of $f$ at the endpoints), then we can split the interval at some intermediate point, use the above analysis on each subinterval, and add the results, again getting $\int_a^b f = c(b - a)$.

**Step functions.** A $f : [a, b] \to \mathbb{R}$ is a step function if there exists a partition $P = \{x_i\}_{0}^{n} \in \mathcal{P}[a, b]$ such that $f$ is constant on each open subinterval $(x_{i-1}, x_i)$ for $i = 1, \ldots, n$. If $f$ is identically $c_i$ on $(x_{i-1}, x_i)$ for $i = 1, \ldots, n$, then

$$\int_a^b f = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f = \sum_{i=1}^{n} c_i \Delta x_i,$$

by the preceding example.

The definition of $\int_a^b f$ required $a < b$; it’s frequently convenient to allow more flexibility:

**Definition.** Define

$$\int_a^b f = \begin{cases} -\int_b^a f & \text{if } a > b \\ 0 & \text{if } a = b. \end{cases}$$

**Corollary 16.19.** Let $f$ be integrable on a closed interval $I$. Then

$$\int_a^c f = \int_a^b f + \int_b^c f$$

for all $a, b, c \in I$.

**Proof.** This follows from considering the cases determined by the possible relative positions of $a, b, c$. \hfill \Box

Most of our results on integrals have true analogues if the limits of integration are equal or backwards. However, when we write $\int_a^b f$ without specifying the ordering of $a, b$, by default we intend that $a < b$!

**Exercise 16.20.** Let $a < c < b$, and let $f : [a, b] \to \mathbb{R}$ be integrable. Suppose that $\int_a^c f > 0$ and $\int_a^b f < 0$. Prove that there exists $t \in (c, b)$ such that $\int_a^t f = 0$.

Here comes the pillar of calculus (hence its name), Newton’s and Leibniz’ (independent) discovery that integration is the inverse of differentiation. You used it most often in your calculus course to evaluate integrals using “antiderivatives” (also called “indefinite integrals”).
Actually, there are two closely related forms of the Fundamental Theorem, which we prove independently:

**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

\[
\left| \frac{F(x) - F(t)}{x-t} - f(t) \right| = \left| \int_t^x f(s) \frac{ds}{x-t} - \int_t^x f(t) \frac{ds}{x-t} \right|
\]

\[
= \left| \int_t^x (f(s) - f(t)) \frac{ds}{x-t} \right|
\]

\[
\leq \sup \left\{ |f(s) - f(t)| \left| s \text{ between } t \text{ and } x \right| \right\} \frac{|x-t|}{|x-t|}
\]

\[
= \sup \left\{ |f(s) - f(t)| \mid s \text{ between } t \text{ and } x \right\} \quad x \to t
\]

\[
\to 0.
\]

(ii). Let $\epsilon > 0$. Choose $P = \{x_i\}_{i=0}^n \in \mathcal{P}[a, b]$ such that $|S - \int_a^b f| < \epsilon$ for all $S \in \mathcal{R}(P)$. By the Mean Value Theorem (for derivatives), 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

\[ |G(b) - G(a) - \int_a^b f| < \epsilon. \]

Letting $\epsilon \to 0$ gives the conclusion.

**Exercise 16.21.** Let $f: [a, b] \to \mathbb{R}$ be continuous and nonnegative. Prove that if $\int_a^b f = 0$ then $f$ is identically 0. Hint: show that the function $F(x) = \int_a^x f$ is increasing.
Exercise 16.22. In each part below, find the derivative of the function $f : \mathbb{R} \to \mathbb{R}$ defined by the given formula, and justify your method.

(a) 

$$f(x) = \int_0^{e^{3x}} \sqrt{t^8 + t + 5} \, dt.$$ 

(b) 

$$f(x) = \int_{\log(x^2+1)}^0 \sqrt{t^8 + t + 5} \, dt.$$ 

(c) 

$$f(x) = \int_{\log(x^2+1)}^{e^{3x}} \sqrt{t^8 + t + 5} \, dt.$$ 

Exercise 16.23. Let $f : [a, b] \to \mathbb{R}$ be continuous, and assume that there exist $s, t \in \mathbb{R}$ such that $s \neq t$ and 

$$s \int_a^x f + t \int_x^b f = 0 \quad \text{for all } x \in (a, b).$$ 

Prove that $f(x) = 0$ for all $x \in [a, b]$.

Combining (i) and (ii) in the Fundamental Theorem of Calculus, if $f$ is continuous on $(a, b)$, then the function $F$ is continuous on $[a, b]$ and differentiable on $(a, b)$, so the Mean Value Theorem (for derivatives) says there exists $c \in (a, b)$ such that 

$$\int_a^b f = F(b) - F(a) = F'(c)(b - a) = f(c) \int_a^b 1,$$ 

which is a special case of the Mean Value Theorem for Integrals — and that’s why the latter result is given it’s name.

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 — this a minor problem for us, so let’s clear it up right now: we have never allowed ourselves to consider differentiating a function at an endpoint of its domain. How to get around this? Well, it makes sense to restrict a differentiable function to a closed interval $[a, b]$ contained in its domain. When we want to speak of a derivative on a closed interval $[a, b]$, let’s agree once and for all that we have in mind that the function is differentiable on some open interval containing $[a, b]$.

It’s interesting to note, however, that if we are considering a function $f$ which is differentiable on $[a, b]$, we don’t need any information of
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 $x \downarrow a$.

Continuing this train of thought, in (i) of the Fundamental Theorem, if in fact $f$ is continuous on $[a, b]$, the result tells us $F' = f$ on $(a, b)$; it’s not hard to see that this extends to the endpoints (with our new convention): we can say $F' = f$ on $[a, b]$ if $f$ is continuous there. How do we extend to the endpoints here? Just extend $f$ to a continuous function on an open interval containing $[a, b]$, for example by making it constant to the left of $a$ and the right of $b$.

Anyway, just remember that when we talk about differentiating a function $f$ on a closed interval $[a, b]$, it tacitly means $f$ extends to a differentiable function on an open interval containing $[a, b]$.

In the next two theorems we take advantage of this so that we can state our hypotheses in a less fussy — and more convenient — way.

The Fundamental Theorem of Calculus leads us to expect “integral analogues” of results for derivatives; we’ve already seen analogues of a couple of the arithmetic properties, involving sums and constant multiples. What about the Product and Quotient Rules? It turns out that there’s no useful integral analogue of the Quotient Rule, but here comes the analogue of the Product Rule:

**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,

$$f(b)g(b) - f(a)g(a) = \int_a^b (fg)' = \int_a^b fg' + \int_a^b f'g,$$

implying the desired equality. $\square$

And here’s the analogue of the Chain Rule:

**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.** Since $\phi$ is differentiable on $[a, b]$, it’s continuous there, so $\phi([a, b])$ is a compact interval. Define $F: \phi([a, b]) \to \mathbb{R}$ by $F(x) = \int_{\phi(a)}^x f$. 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 $\phi$ and $f$ are 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:

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

**Improper integrals**

Now for a rather curious variation on Riemann integrals: “improper” integrals. You’ll remember from your calculus course that these come in two kinds: either the integrand is unbounded or a limit of integration is infinite. It turns out that we can deal with both kinds in a unified way.

First, let’s observe one more thing about Riemann integrals: let $-\infty < a < b < \infty$, and suppose $f: [a, b) \to \mathbb{R}$ is bounded, and integrable on $[a, t]$ for every $t \in (a, b)$. If $f(b)$ is defined, we’ve seen before that $f$ is integrable on $[a, b]$ and $\int_a^b f = \lim_{t \uparrow b} \int_a^t f$. What if $b$ is not in the domain of $f$? Hopefully you’ve decided right away that there is no need for a separate definition for such a situation: we should just imagine that we’ve extended $f$ to $[a, b]$ by defining $f(b)$ any way we want (it doesn’t matter what value we give it), and then as before we have $\int_a^b f = \lim_{t \uparrow b} \int_a^t f$, which only involves the original function $f$ defined on the right-half-open interval $[a, b)$ — and of course similarly at the left endpoint. Conclusion: we do not want to call such integrals improper, even though the original function is not defined at one endpoint.

Here come the improper ones\(^{62}\)

**Definition.** (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

$$\int_a^b f = \lim_{t \uparrow b} \int_a^t f.$$

\(^{62}\)but I have to comment that I find this definition ultimately unsatisfying (although I like it better than any other I’ve read), however I have come to the sad conclusion that with the accepted terminology there’s no way to define improper integrals in a logically reassuring way. Roughly speaking, improper integrals are certain limits of Riemann integrals.
(ii) Let $f$ be Riemann integrable on $[t, b]$ for every $t \in (a, b)$. If either $f$ is unbounded on $(a, b]$ or $a = -\infty$ we define
\[ \int_{a}^{b} f = \lim_{t \downarrow a} \int_{t}^{b} f. \]

(iii) In either of the two above cases we call $\int_{a}^{b} f$ an improper integral, and if the limit exists we say $\int_{a}^{b} f$ converges, or $f$ is improperly integrable on $(a, b)$; otherwise we say $\int_{a}^{b} f$ diverges.

(iv) More generally, if $a = x_0 < x_1 < \cdots < x_n = b$, and if for each $i = 1, \ldots, n$ either $f$ is Riemann integrable on $[x_{i-1}, x_i]$ or the integral $\int_{x_{i-1}}^{x_i} f$ is improper, with at least one of the integrals $\int_{x_{i-1}}^{x_i} f$ for $i = 1, \ldots, n$ really being improper, then we call $\int_{a}^{b} f$ an improper integral, and moreover if for every $i = 1, \ldots, n$ either $f$ is Riemann integrable on $[x_{i-1}, x_i]$ or improperly integrable on $(x_{i-1}, x_i)$, then we say $f$ is improperly integrable on $(a, b)$ and define
\[ \int_{a}^{b} f = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f. \]

In case (i), if $c \in (a, b)$ then $\int_{a}^{b} f$ exists if and only if $\int_{c}^{b} f$ does, in which case $\int_{a}^{b} f = \int_{c}^{a} f + \int_{c}^{b} f$. Similarly in case (ii).

**Examples.**

(i) Both $\int_{1}^{\infty} x \, dx$ and $\int_{1}^{\infty} 1/x^2 \, dx$ are improper integrals; the first diverges and the second converges.

(ii) $\int_{0}^{\infty} 1/x \, dx$ is improper because both $\int_{1}^{\infty} 1/x \, dx$ and $\int_{1}^{\infty} 1/x \, dx$ are (albeit for different reasons).

(iii) $\int_{-\infty}^{-1} 1/x \, dx$ is improper because $\int_{-\infty}^{-1} 1/x \, dx$, $\int_{-1}^{0} 1/x \, dx$, and $\int_{0}^{\infty} 1/x \, dx$ are.

**Exercise 16.24.** Show that:

(a) $\int_{1}^{\infty} x \, dx$ diverges;

(b) $\int_{1}^{\infty} 1/x^2 \, dx$ converges;

(c) $\int_{0}^{1} 1/\sqrt{x} \, dx$ converges (and find its value).

**Exercise 16.25.** Show that:

(a) $\int_{-1}^{1} 1/x^2 \, dx$ diverges;

(b) $\int_{-1}^{1} 1/x^{1/3} \, dx$ converges (and find it’s value);

**Exercise 16.26.** Show that $\int_{-\infty}^{\infty} 1/(x^2 + 1) \, dx = \pi$. 
The arithmetic of improper integrals follows from that of Riemann integrals and of limits:

**Proposition 16.27.** If each of $f$ and $g$ is either Riemann or improperly integrable, 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 $c \in \mathbb{R}$.

More precisely, we’re assuming that both $f$ and $g$ are improperly integrable on $(a, b)$, or $-\infty < a < b < \infty$ and one of $f$ or $g$ is Riemann integrable on $[a, b]$ while the other is either improperly integrable on $(a, b)$ or Riemann integrable on $[a, b]$ — note that this actually allows both to be Riemann integrable, although we don’t really get a new proof of the arithmetic properties of Riemann integrals.

**Proof.** If necessary, partition the interval $[a, b]$, so that without loss of generality both $f$ and $g$ are integrable on $[a, t]$ for all $t \in (a, b)$. For (i), we have

$$\int_a^b (f + g) = \lim_{t \uparrow b} \int_a^b (f + g) = \lim_{t \uparrow b} \left( \int_a^b f + \int_a^b g \right) = \lim_{t \uparrow b} \int_a^b f + \lim_{t \uparrow b} \int_a^b g = \int_a^b f + \int_a^b g.$$

For (ii), we have

$$\int_a^b cf = \lim_{t \uparrow b} \int_a^t cf = \lim_{t \uparrow b} c \int_a^t f = c \lim_{t \uparrow b} \int_a^t f = c \int_a^b f. \quad \Box$$

**Exercise 16.28.** Let $f$ be a nonnegative improperly integrable function on $[0, \infty)$. Prove that if $f$ is uniformly continuous, then

$$\lim_{x \to \infty} f(x) = 0.$$

**Comparison Theorem for Improper Integrals.** Let $\int_a^b f$ and $\int_a^b g$ be improper integrals, and assume $\int_a^b g$ converges and $0 \leq f \leq g$. Then $\int_a^b f$ converges, and

$$\int_a^b f \leq \int_a^b g.$$
Proof. If necessary, partition the interval $[a, b]$, so that without loss of generality $f$ and $g$ are integrable on $[a, t]$ for all $t \in (a, b)$. Since $0 \leq f \leq g,$

\[
\int_a^t f \leq \int_a^t g \leq \int_a^b g \quad \text{for all } t \in (a, b).
\]

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

\[
\int_a^b f = \lim_{t \uparrow b} \int_a^t f \leq \int_a^b g.
\]

Example. $\int_1^\infty \frac{1}{x^p} \, dx$ 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 $p = 1$,

\[
\lim_{t \to \infty} \log t \bigg|_1^t = \infty,
\]

so $\int_1^\infty \frac{1}{x} \, dx$ does not exist. Finally, for $p < 1$ we have $\frac{1}{x^p} \geq \frac{1}{x}$ for all $x \geq 1$, so by the Comparison Theorem $\int_1^\infty \frac{1}{x^p} \, dx$ fails to exist.

Corollary 16.29. Let $\int_a^b f$ be improper, and assume $\int_a^b |f|$ converges. Then so does $\int_a^b f$, and

\[
\left| \int_a^b f \right| \leq \int_a^b |f|.
\]

Proof. If necessary, partition the interval $[a, b]$, so that without loss of generality $f$ is integrable on $[a, t]$ for all $t \in (a, b)$. 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 $a < t < b$ then

\[
\left| \int_a^t f \right| \leq \int_a^t |f|,
\]

so letting $t \uparrow b$ we get

\[
\left| \int_a^b f \right| \leq \int_a^b |f|. \quad \square
\]
Combining the above corollary with the Comparison Theorem, we get:

**Corollary 16.30.** If $\int_a^b f$ is improper, $g$ is improperly integrable on $(a, b)$, and $|f| \leq g$, then $f$ is improperly integrable.

**Exercise 16.31.** Show that $\int_\pi^\infty \frac{\sin x}{x} \, dx$ converges. Hint: integration by parts.
17. LOG AND EXP

The “Natural” Logarithm Function. Define \( \log : (0, \infty) \to \mathbb{R} \) by
\[
\log x = \int_1^x \frac{dt}{t}.
\]

**Proposition 17.1.** (i) \( \log' (x) = \frac{1}{x} \);
(ii) \( \log 1 = 0 \);
(iii) \( \log xy = \log x + \log y \);
(iv) \( \frac{\log y}{y} = \log x - \log y \);
(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
\[
\frac{\log y}{y} + \log y = \log \left( \frac{x}{y} \right) = \log x.
\]
(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’s 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 \( \lim_{x \to 1^0} \log x = -\infty \).
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.

In the proof of (vii) above, we used the letter $n$ to (lazily) signal that we were considering a sequence.

**Exercise 17.2.** Prove that

$$\lim_{n \to \infty} \sum_{i=1}^{n} \frac{1}{n+i} = \log 2.$$ 

**The Exponential Function.** Define $\exp : \mathbb{R} \to (0, \infty)$ by $\exp = \log^{-1}$.

**Proposition 17.3.**

(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

$$\frac{\exp(x)}{\exp(y)} = \exp\left(\log \frac{\exp(x)}{\exp(y)}\right)$$

$$= \exp\left(\log \exp(x) - \log \exp(y)\right)$$

$$= \exp(x - y).$$

(v). This follows from Theorem 14.14 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 $(0, \infty)$. \qed
**Arbitrary Powers.** For each $x > 0$ and $t \in \mathbb{R}$ define

$$x^t = \exp(t \log x).$$

**Proposition 17.4.** 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

$$\frac{d}{dx} x^t = \frac{d}{dx} \exp(t \log x) = \exp(t \log x) \frac{t}{x} = tx^{t-1}. \quad \square$$

**The Number $e$.**

$e := \exp(1)$.

**Proposition 17.5.** $e^x = \exp(x)$ for all $x \in \mathbb{R}$.

*Proof.* $e^x = \exp(x \log e) = \exp(x). \quad \square$
Now we come to a surprisingly important area of analysis, arising from the desire to add up infinitely many things. Of course, this doesn’t really make sense; to put it on a firm footing we need to involve sequences. Once you see the following definition, you’ll probably wonder why we have a separate concept called series — in fact, however, often it’s much more effective to use series than sequences.

**Definition.** Given a real sequence \((a_n)\), define a new sequence \((s_k)\) by

\[ s_k = \sum_{n=1}^{k} a_n. \]

The series with \(n\)th term \(a_n\) is the sequence \((s_k)\).

**Notation and Terminology.** \(\sum_{n=1}^{\infty} a_n\) denotes the series with terms \((a_n)\), and the \(k\)th partial sum of the series is \(\sum_{n=1}^{k} a_n\).

Thus, a series is a sequence, so we can apply all the terminology and results concerning sequences. In particular, a series converges if and only if its sequence of partial sums does.

**Notation and Terminology.** The sum of a series \(\sum_{n=1}^{\infty} a_n\) is the limit of its sequence of partial sums, and the sum is denoted \(\sum_{n=1}^{\infty} a_n\).

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 \(\sum_{n=1}^{\infty} a_n\) by

\[ a_n = \begin{cases} s_1 & \text{if } n = 1 \\ s_n - s_{n-1} & \text{if } n > 1. \end{cases} \]

Just as with sequences, it’s 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’s 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

\[ {64} \text{Again, I find this definition unsatisfying, although better than most attempts I've read in print. Some authors define a series as something like an “expression of the form } \sum_{n=1}^{\infty} a_n, \text{” but this is logically suspect — what does it mean to define something to be the same as the notation for it?} \]

\[ {65} \text{and here's a particularly confusing duplication of notation!} \]
the general theory we usually assume the starting point of the series is 1.

**Example.** If $|x| < 1$, the geometric series $\sum_{n=0}^{\infty} x^n$ converges, and the sum is
\[ \sum_{n=0}^{\infty} x^n = \frac{1}{1 - x}. \]
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’s sometimes convenient to abuse the terminology by referring to a series of the form $\sum c x^n$ (no matter what the starting point is) as a geometric series as well.

**Example.** The series $\sum_{n=1}^{\infty} 1/(n^2+n)$ converges, because the $k$th partial sum is
\[ s_k = \sum_{n=1}^{k} \frac{1}{n^2+n} = \sum_{n=1}^{k} \left( \frac{1}{n} - \frac{1}{n+1} \right) = 1 - \frac{1}{k+1} \xrightarrow{k \to \infty} 1. \]
Again, we got lucky enough to find a closed form for $s_k$. In this case the partial sums are *telescoping*.

**Example.** The harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges, because
\[ s_k = \sum_{n=1}^{k} \frac{1}{n} \geq \int_{1}^{k+1} \frac{dx}{x} = \log(k+1) \to \infty. \]
We’ll generalize the technique of this example a little later, in the Integral Test.

**Cauchy Criterion for Series.** $\sum a_n$ converges if and only if for all $\epsilon > 0$ there exists $l \in \mathbb{N}$ such that
\[ \left| \sum_{j}^{k} a_n \right| < \epsilon \quad \text{for all } k \geq j \geq l. \]

**Proof.** This follows from the corresponding result for sequences, since
\[ |s_k - s_{j-1}| = \left| \sum_{j}^{k} a_n \right|. \]

**Corollary 18.1.** If $\sum a_n$ converges, then $a_n \to 0$. 
Proof. This follows immediately from the preceding theorem, since

\[ a_n = \sum_{k=n}^{n} a_k. \]  \hfill \Box

Warning: 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 \( \sum \frac{1}{n} \).

Example. 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 \( x^n \not\to 0 \).

Here’s the arithmetic of convergent series:

**Proposition 18.2.** If \( \sum a_n \) and \( \sum b_n \) both converge, then:

(i) \( \sum (a_n + b_n) = \sum a_n + \sum b_n \);
(ii) \( \sum c a_n = c \sum a_n \) if \( c \in \mathbb{R} \).

Proof. (i). \( \sum_{k=1}^{n} (a_n + b_n) = \sum_{k=1}^{n} a_n + \sum_{k=1}^{n} b_n \to \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n \).
(ii). \( \sum_{k=1}^{n} c a_n = c \sum_{k=1}^{n} a_n \to c \sum_{n=1}^{\infty} a_n. \)  \hfill \Box

Here’s a comment, involving manipulating series, which is somehow related to (i) above: we can’t “remove parentheses” from a series, for example

\[ (1 - 1) + (1 - 1) + (1 - 1) + \cdots = 0 + 0 + 0 + \cdots = 0, \]

but the series

\[ 1 - 1 + 1 - 1 + 1 - 1 + \cdots \]

diverges because the terms don’t go to 0. However, it’s ok to “insert parentheses” in a convergent series:

**Exercise 18.3.** Let \( \sum_{n=1}^{\infty} a_n \) converge, and let \( (n_1, n_2, n_3, \ldots) \) be a strictly increasing sequence of positive integers. Put \( n_0 = 0 \). For each \( j \in \mathbb{N} \) define

\[ b_j = a_{n_{j-1}+1} + \cdots + a_{n_j}. \]

Prove that \( \sum_{j=1}^{\infty} b_j \) converges and

\[ \sum_{j=1}^{\infty} b_j = \sum_{n=1}^{\infty} a_n. \]

Most of the time, if a series converges we won’t be able to compute the actual sum; the best we can usually hope for is to answer the “yes/no” question: does the series converge? This question turns out to be quite important, and a lot of tests have been devised to answer it, of which we’ll only see a few.
Comparison Test.\footnote{\label{fn:Comparison Test 1}Frequently $a_n \geq 0$, in which case of course the absolute values can be omitted.} If $|a_n| \leq b_n$ for all $n$ and $\sum b_n$ converges, then $\sum a_n$ converges and $\left|\sum a_n\right| \leq \sum b_n$.

Proof. Let $\epsilon > 0$. 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,

$$\left|\sum_{1}^{k} a_n\right| \leq \sum_{1}^{k} b_n \quad \text{for all } k \in \mathbb{N}.$$  

Letting $k \to \infty$, we get

$$\left|\sum_{1}^{\infty} a_n\right| \leq \sum_{1}^{\infty} b_n. \quad \square$$

Definition. $\sum a_n$ converges absolutely if $\sum |a_n|$ converges.

Corollary 18.4. \footnote{\label{fn:Corollary 18.4}This may be viewed as a “triangle inequality for series.”} If $\sum a_n$ converges absolutely, then $\sum a_n$ converges and $\left|\sum a_n\right| \leq \sum |a_n|$.

Proof. This follows immediately from the Comparison Test. \quad \square

Thus, in the Comparison Test we can conclude that $\sum a_n$ converges absolutely. Moreover, just as for sequences, a series $\sum_{n=1}^{\infty} a_n$ converges if and only if every tail $\sum_{n=k}^{\infty} a_n$ does. Consequently, if we only have $|a_n| \leq b_n$ for large $n$ (meaning there exists $k \in \mathbb{N}$ such that the inequality holds for all $n \geq k$) then we can still conclude $\sum a_n$ converges absolutely, although we no longer have the inequality $\left|\sum a_n\right| \leq \sum b_n$.

Exercise 18.5. (a) Prove that the series

$$\sum_{n=1}^{\infty} \left(\frac{1}{n} - \int_{n}^{n+1} \frac{1}{x} \, dx\right)$$

converges.

(b) Use the result of part (a) to show that

$$\lim_{k \to \infty} \left(\sum_{n=1}^{k} \frac{1}{n} - \log(k+1)\right)$$

exists.
The limit in part (b) of the preceding exercise is Euler's constant; it’s still unknown whether it’s irrational.

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.

A lot of the time, it’s a little tricky to get the inequalities in the Comparison Theorem to work out right; here’s a handy alternative:

**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.*

**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. 

**Exercise 18.6.** Show that the series

$$\sum_{n=2}^{\infty} \frac{3n^3 - 5n + 1}{7n^4 - 6n^2 + 3n - 8}$$

diverges.

However, sometimes the Comparison Test applies but the Limit Comparison Test doesn’t:

**Exercise 18.7.** In each part, determine whether the series converges:

(a) $$\sum_{n=1}^{\infty} \frac{2 + \sin n}{n^2}$$

(b) $$\sum_{n=1}^{\infty} \frac{2 + \cos n}{\sqrt{n}}$$

We know that integrals are additive over partitions of the interval; here’s an “infinite version” of this:

**Exercise 18.8.** Let $f : [a, \infty) \to \mathbb{R}$ be integrable on $[a, t]$ for all $t > a$, and suppose $(a_n)_{n=0}^{\infty}$ is a strictly increasing sequence diverging to $\infty$ such that $a_0 = a$.

(a) Prove that if $\int_{a}^{\infty} f$ converges, then so does the series $\sum_{n=1}^{\infty} \int_{a_{n-1}}^{a_n} f$.

(b) Assuming $f$ is nonnegative, prove the converse of part (a).

Here’s an opportunity to apply the above exercise:
Exercise 18.9. Give an example of a continuous nonnegative function $f$ on $[1, \infty)$ such that $\int_1^\infty f$ converges but $f(x) \not\to 0$ as $x \to \infty$.

But here’s the real reason we introduced the exercise before the last one: another convergence test for series, sort of a different kind of comparison test:

**Integral Test.** Let $f: [1, \infty) \to \mathbb{R}$ be decreasing and nonnegative. Then the series $\sum_{n=1}^\infty f(n)$ converges if and only if the improper integral $\int_1^\infty f$ does.

Note that since $f$ is monotone, it is integrable on $[1, t]$ for all $t \geq 1$, so the improper integral makes sense.

**Proof.** By the above exercise, since $f \geq 0$ the integral $\int_1^\infty f$ converges if and only if the series $\sum_{n=1}^\infty \int_n^{n+1} f$ does. We’ll apply the Comparison Test to show that this latter series converges if and only if the series $\sum_{n=1}^\infty f(n)$ does. Since $f$ is decreasing, we have

$$f(n+1) \leq \int_n^{n+1} f \leq f(n) \quad \text{for all } n \in \mathbb{N}.$$ 

Thus if $\sum f(n)$ converges then so does $\sum \int_n^{n+1} f$ by the Comparison Test. Conversely, if $\sum \int_n^{n+1} f$ converges, then the above inequalities and the Comparison Test tell us that the series $\sum_{n=1}^\infty f(n+1)$ does also. But this is the tail $\sum_{2}^\infty f(n)$ of the series $\sum_{1}^\infty f(n)$, so the latter series converges as well. $\Box$

**Example.** The $p$-series $\sum_{1}^\infty 1/n^p$ 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 $1/n^p$ don’t go to 0, so the series diverges.

The harmonic series is the special case $p = 1$.

Exercise 18.10. Prove that if $\sum_{1}^\infty a_n^2$ converges then so does $\sum_{1}^\infty a_n/n$. Hint: Cauchy-Schwarz Inequality.

In the Integral Test, the left endpoint of the interval $[1, \infty)$ can be replaced by any positive number; thus it’s enough for $f$ to be eventually decreasing and nonnegative (meaning there exists $a > 0$ such that the desired property holds on $[a, \infty)$), and it’s enough to use the improper integral $\int_a^\infty f$ for any $a > 0$.  

---

68 but not uniformly so, by a previous exercise!
Example. We’ll determine all values of $p$ for which the series
\[ \sum_{n=2}^{\infty} \frac{1}{n(\log n)^p} \]
converges.

If $p > 0$ then the function $1/x(\log x)^p$ is decreasing, and a change of variables gives
\[ \int_{e}^{\infty} \frac{1}{x(\log x)^p} \, dx = \int_{1}^{\infty} \frac{1}{w^p} \, du, \]
which we know converges if and only if $p > 1$. Thus by the Integral Test the given series converges if $p > 1$ and diverges if $0 < p \leq 1$.

On the other hand, if $p \leq 0$, then $1/n(\log n)^p \geq 1/n$, so the given series diverges by the Comparison Test, because the harmonic series $\sum 1/n$ diverges.

Conclusion: the given series converges if and only if $p > 1$.

As the preceding exercise illustrates, the Integral Test is occasionally handy for series involving $\log$. However, sometimes it’s easier to use something else:

Example. We’ll determine all values of $p$ for which the series
\[ \sum_{n=1}^{\infty} \frac{\log n}{n^p} \]
converges.

We have $\log n/n^p \geq 1/n^p$ for large enough $n$, so if $p \leq 1$ then the given series diverges because the $p$-series does.

On the other hand, let $p > 1$. Then $p = 1 + 2a$ with $a > 0$, so
\[ \frac{\log n}{n^p} = \frac{\log n}{n^a} \cdot \frac{1}{n^{1+a}}. \]
The fraction $\log n/n^a$ goes to 0 as $n \to \infty$ by l’Hôpital’s Rule, so for large enough $n$ we have $\log n/n^p \leq 1/n^{1+a}$. Since $\sum 1/n^{1+a}$ converges because it is a $p$-series with $p = 1 + a > 1$, the given series converges by the Comparison Test.

Conclusion: the given series converges if and only if $p > 1$. This example is interesting because it tells us multiplying by the unbounded function $\log n$ has no effect on the convergence of the $p$-series (although of course it’ll affect the sum).

Definition. $\sum a_n$ converges conditionally if it converges but not absolutely.

\[ ^{69}i.e., \, n \geq 3 \]
Conditionally convergent series are delicate (for example, see the remark following the Rearrangement Theorem below); the only general test we’ll prove which has the power to recognize conditional convergence is the Alternating Series Test below. But first:

**Definition.** \( \sum a_n \) is alternating if \( a_n a_{n+1} < 0 \) for all \( n \).

We’ve made a choice here: we insist that the terms of an alternating series strictly alternate in sign; in particular, none of the terms can be zero. Some books have less restrictive definitions of alternating series, but ours gives what we need. Thus, every alternating series can be written either in the form \( \sum_{n=0}^{\infty} (-1)^n p_n \) or \( \sum_{n=1}^{\infty} (-1)^n p_n \), with \( p_n > 0 \) for all \( n \), and the starting point for \( n \) depends upon whether the first term of the alternating series is positive or negative. Alternatively (!) we could start the series at \( n = 1 \) and still make the first term positive by writing \( \sum_{n=1}^{\infty} (-1)^{n+1} p_n \).

**Alternating Series Test.** If \( \sum a_n \) is an alternating series such that \( a_n \to 0 \) and \( |a_n| \geq |a_{n+1}| \) for all \( n \), then the series converges to a sum between 0 and \( a_1 \), strictly between if the above inequalities are strict.

**Proof.** Without loss of generality \( a_1 > 0 \) (otherwise just delete the first term). Then the odd terms are positive, the even terms are negative, and \( a_n + a_{n+1} \) is nonnegative for \( n \) odd and nonpositive for \( n \) even. For any \( k \in \mathbb{N} \),

\[
\sum_{1}^{k} a_n = (a_1 + a_2) + (a_3 + a_4) + \cdots + \begin{cases} a_{k-1} + a_k & \text{if } k \text{ even} \\ a_k & \text{if } k \text{ odd} \end{cases}
\geq 0.
\]

On the other hand,

\[
\sum_{1}^{k} a_n = a_1 + (a_2 + a_3) + (a_4 + a_5) + \cdots + \begin{cases} a_{k-1} + a_k & \text{if } k \text{ odd} \\ a_k & \text{if } k \text{ even} \end{cases}
\leq a_1.
\]

Note that the above reasoning shows that

\[
\left| \sum_{j}^{k} a_n \right| \leq |a_j| \quad \text{if } k \geq j.
\]

Let \( \epsilon > 0 \). Since \( a_n \to 0 \), there exists \( l \in \mathbb{N} \) such that \( |a_n| < \epsilon \) for all \( n \geq l \). Then

\[
\left| \sum_{j}^{k} a_n \right| \leq |a_j| < \epsilon \quad \text{if } k \geq j \geq l.
\]
Therefore, \( \sum a_n \) converges, and then the inequality

\[
0 \leq \sum_{1}^{k} a_n \leq a_1 \quad \text{for all } k
\]

gives

\[
0 \leq \sum_{1}^{\infty} a_n \leq a_1.
\]

Moreover, if \((|a_n|)\) is strictly decreasing, the even partial sums \(s_{2k}\) are strictly increasing and the odd partial sums \(s_{2k+1}\) are strictly decreasing, so the sum is strictly between 0 and \(a_1\).

\[\square\]

**Exercise 18.11.** Prove that the *alternating harmonic series* \(\sum_{1}^{\infty} (-1)^{n+1}/n\) converges conditionally, and the sum is strictly between 0 and 1.

The following result shows that absolutely convergent series are robust in the sense that the terms can be rearranged willy-nilly:

**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 \(\sum a_{f(n)} = \sum a_n\). Let \(\epsilon > 0\). Choose \(j \in \mathbb{N}\) such that \(\sum_{j+1}^{\infty} |a_n| < \epsilon/2\). Now choose \(l \in \mathbb{N}\) such that

\[
f(\{1, \ldots, l\}) \supset \{1, \ldots, j\}.
\]

\[\text{Actually, we’ll see later that the sum is log 2.}\]
Then $k \geq l$ implies
\[
\left| \sum_{1 \leq n \leq k} a_f(n) - \sum_{1 \leq n} a_n \right| \leq \left| \sum_{1 \leq n \leq k} a_f(n) - \sum_{1 \leq n} a_n \right| + \left| \sum_{1 \leq n} a_n \sum_{1 \leq n} a_n \right|
\]
\[
= \left| \sum_{n \leq k \atop f(n) > j} a_f(n) \right| + \left| \sum_{j+1} a_n \right|
\]
\[
\leq \sum_{n \leq k \atop f(n) > j} |a_f(n)| + \sum_{j+1} |a_n|
\]
\[
\leq 2 \sum_{j+1} |a_n|
\]
\[
< \epsilon.
\]

What if $\sum a_n$ converges conditionally? Then it turns out that there exist divergent rearrangements, and moreover there are rearrangements converging to any real number we want. This is called Riemann’s Rearrangement Theorem; since we don’t need it, we don’t 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 $3s/2$.

The next two results involve our old friends lim sup and lim inf. There are weaker versions which use regular limits.

**Ratio Test.** Assume $a_n \neq 0$ for all $n$.

(i) If $\lim \left| \frac{a_{n+1}}{a_n} \right| < 1$ then $\sum a_n$ converges absolutely, while

(ii) if $\lim \left| \frac{a_{n+1}}{a_n} \right| > 1$ then $\sum a_n$ diverges.

**Proof.** (i). If $\lim \left| \frac{a_{n+1}}{a_n} \right| < t < 1$, then there exists $k \in \mathbb{N}$ such that $n \geq k$ implies
\[
\left| \frac{a_{n+1}}{a_n} \right| < t.
\]

Then
\[
|a_{k+j}| \leq t|a_{k+j-1}| \leq t^2|a_{k+j-2}| \leq \cdots \leq t^j|a_k| \quad \text{for all } j \in \mathbb{N}.
\]
Thus the tail \( \sum_{k+1}^{\infty} a_n \) converges absolutely by comparison to the geometric series \( \sum t^n|a_k| \), so \( \sum_{1}^{\infty} a_n \) 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 \( a_n \neq 0 \), thus \( \sum a_n \) diverges. \( \square \)

In the above theorem it often happens that \( \rho := \lim |a_{n+1}/a_n| \) exists, and then the Ratio Test says \( \sum a_n \) converges absolutely if \( \rho < 1 \) and diverges if \( \rho > 1 \). However, if \( \rho = 1 \) the series could either converge or diverge. For example, the harmonic series \( \sum 1/n \) diverges, while the series \( p \)-series \( \sum 1/n^2 \) converges, but in both cases the ratios \( a_{n+1}/a_n \) converge to 1.

The following result is similar to, but a little better than, the Ratio Test:

**Root Test.** Put \( \rho = \lim |a_n|^{1/n} \).

(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 \( a_n \neq 0 \), so \( \sum a_n \) diverges. \( \square \)

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.

Here’s why I said the Root Test is a little better than the Ratio Test:

**Exercise 18.12.** Consider the series \( \sum_{n=1}^{\infty} a_n \), where
\[
a_n = \begin{cases} 
3^{-n} & \text{if } n \text{ is odd} \\
2^{-n} & \text{if } n \text{ is even}.
\end{cases}
\]

(a) Show that the Ratio Test gives no information.
(b) Prove that the Root Test shows that the series converges.

Here are some series taken out of context, so you must decide which tests to use:

**Exercise 18.13.** Find all nonzero real numbers \( p \) for which the series \( \sum_{n=1}^{\infty} p^n n^p \) converges absolutely, converges conditionally, or diverges.
Exercise 18.14. Let $a_n > 0$ for all $n \in \mathbb{N}$. Prove that if $\sum a_n$ diverges, then so does

$$\sum \frac{a_n}{1 + a_n}.$$ 

Exercise 18.15. Does the series

$$\sum_{n=1}^{\infty} 2^n e^{-n}$$

converge?

Exercise 18.16. Does the series

$$\sum_{n=1}^{\infty} n^n e^{-n}$$

converge?

Exercise 18.17. Does the series

$$\sum_{n=1}^{\infty} e^{-\log n}$$

converge?

Exercise 18.18. Does the series

$$\sum_{n=1}^{\infty} (\log n)e^{-\sqrt{n}}$$

converge? (This one’s tricky.)

Exercise 18.19. Does the series

$$\sum_{n=1}^{\infty} n! e^{-n}$$

converge?

Exercise 18.20. Does the series

$$\sum_{n=1}^{\infty} n! e^{-n^2}$$

converge?

Exercise 18.21. Does the series

$$\sum_{n=1}^{\infty} \frac{2^n n!}{n^n}$$

converge?
Exercise 18.22. Does the following series converge?

\[ \sum_{n=1}^{\infty} \frac{\sqrt{n+1} - \sqrt{n}}{n} \]

Exercise 18.23. Let \((a_n)\) be a decreasing sequence of positive numbers. Prove that if \(\sum a_n\) converges, then \(na_n \to 0\). Hint: consider partial sums of a sufficiently small tail \(\sum_{k+1}^{\infty} a_n\). (This one’s tricky.)
19. SEQUENCES OF FUNCTIONS

In some of the most important metric spaces (for example, $C[0, 1]$, as we’ll see later), the points are functions. In such spaces we can consider convergence of sequences of functions, and other metric concepts such as compactness. However, the elementary theory of sequences and series of functions is more easily developed from scratch.

Notation and Terminology. As before, letters such as $X$ and $Y$ will refer to metric spaces, unless otherwise specified.

How can we talk about convergence of a sequence of functions? If you were invited to ponder this question with no prior knowledge of what we’ll end up doing, you might think: well, if we evaluate all the functions at a point, we get a sequence of values, and if these values are in a metric space then we know what to do:

**Definition.** Let $f, f_1, f_2, \ldots : X \to Y$. Then $(f_n)$ converges pointwise to $f$ if for every $x \in X$, $f_n(x) \to f(x)$.

**Notation and Terminology.** $f_n \to f$ and $f = \lim f_n$ mean $(f_n)$ converges pointwise to $f$.

However, it happens that this simple-minded notion of convergence isn’t very useful, since it doesn’t have any good properties.

**Example.** For each $n \in \mathbb{N}$ define $f_n : [0, 1] \to \mathbb{R}$ by $f_n(x) = x^n$. Then each $f_n$ is continuous, but the pointwise limit is the function $f : [0, 1] \to \mathbb{R}$ defined by

\[
    f(x) = \begin{cases} 
        0 & \text{if } 0 \leq x < 1 \\
        1 & \text{if } x = 1
    \end{cases}
\]

is discontinuous at 1.

**Example.** For each $n \in \mathbb{N}$ define $f_n : [0, 1] \to \mathbb{R}$ by

\[
    f_n(x) = \begin{cases} 
        n & \text{if } 0 < x < \frac{1}{n} \\
        0 & \text{if } x = 0 \text{ or } \frac{1}{n} \leq x \leq 1
    \end{cases}
\]

Then each $f_n$ has integral 1, but the pointwise limit is 0, which has integral 0. (To verify that $f_n \to 0$, let $x \in [0, 1]$. If $x = 0$, then $f_n(x) = 0$ for all $n$, while if $x > 0$ then by the Archimedean Principle there exists $k \in \mathbb{N}$ such that $1/k < x$, and then $f_n(x) = 0$ for all $n \geq k$.)

**Exercise 19.1.** Find a sequence $(f_n)$ of continuous functions on $[0, 1]$ which converge pointwise to 0 but for which $\int_0^1 f_n \neq 0$. 

---

71 no offense, if you thought of it independently
Example. Let \((q_n)\) be a sequence whose range is the set of rational numbers in \([0, 1]\), and for each \(n\) define \(f_n : [0, 1] \to \mathbb{R}\) by

\[
f_n(x) = \begin{cases} 
1 & \text{if } x = q_1, \ldots, q_n \\
0 & \text{if not.}
\end{cases}
\]

Then each \(f_n\) is integrable (being continuous except at the finitely many points \(q_1, \ldots, q_n\)), but the pointwise limit is the Dirichlet function \(f : [0, 1] \to \mathbb{R}\) defined by

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

which is not integrable.

Example. For each \(n \in \mathbb{N}\) define \(f_n : \mathbb{R} \to \mathbb{R}\) by

\[
f_n(x) = \frac{nx}{1 + n^2 x^2}.
\]

Then \(f_n \to 0\) pointwise, but \(f_n'(0) \neq 0\). This is a special case of the general result you’ll verify in the following exercise.

Exercise 19.2. Let \(g : \mathbb{R} \to \mathbb{R}\). Suppose \(g\) is differentiable, \(g(0) = 0\), \(\lim_{x \to \pm \infty} g(x) = 0\), and \(g'(0) \neq 0\). For each \(n \in \mathbb{N}\) define \(f_n : \mathbb{R} \to \mathbb{R}\) by

\[
f_n(x) = g(nx).
\]

Prove that \(f_n \to 0\) pointwise, but \(f_n'(0) \neq 0\).

On the other hand, the following more sophisticated type of convergence\(^{72}\) behaves very well:

Definition. Let \(f, f_1, f_2, \ldots : X \to Y\). Then \((f_n)\) converges uniformly to \(f\) if for all \(\epsilon > 0\) there exists \(k \in \mathbb{N}\) such that

\[
d(f_n(x), f(x)) < \epsilon \quad \text{for all } n \geq k, x \in X.
\]

Notation and Terminology. \(f_n \to f\) uniformly means \((f_n)\) converges uniformly to \(f\).

It follows almost immediately from the definition that \(f_n \to f\) uniformly if and only if \(\sup_{x \in X} d(f_n(x), f(x)) \xrightarrow{n \to \infty} 0\).

As we did when we encountered the definition of uniform continuity, let’s symbolically compare the pointwise and uniform notions of

\(^{72}\)whose importance was not recognized until analysis was put on a firm foundation toward the end of the 19th century
convergence:

\[\forall x \in X \left( \forall \epsilon > 0 \left( \exists k \in \mathbb{N} \left( \forall n \in \mathbb{N} \right) n \geq k \implies d(f_n(x), f(x)) < \epsilon \right) \right)\]

uniform convergence:

\[\forall \epsilon > 0 \left( \exists k \in \mathbb{N} \left( \forall x \in X \left( \forall n \in \mathbb{N} \right) n \geq k \implies d(f_n(x), f(x)) < \epsilon \right) \right)\]

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.

**Exercise 19.3.** Prove that \(x^{1/n} \to 1\) uniformly for \(x \in (a, 1]\) if \(0 < a < 1\) but not if \(a = 0\).

Just as for “ordinary” convergence, there’s a way to detect uniform convergence without knowing the limit:

**Uniform Cauchy Criterion.** Let \(f_1, f_2, \ldots : X \to Y\), and assume that \(Y\) is complete. Then \((f_n)\) converges uniformly if and only if for all \(\epsilon > 0\) there exists \(k \in \mathbb{N}\) such that

\[d(f_n(x), f_j(x)) < \epsilon \quad \text{for all } n, j \geq k, x \in X.\]

**Proof.** First assume \(f_n \to f\) uniformly. Let \(\epsilon > 0\). Choose \(k \in \mathbb{N}\) such that \(d(f_n(x), f(x)) < \epsilon/2\) for all \(n \geq k, x \in X\). Then for all \(n, j \geq k\) and \(x \in X\),

\[d(f_n(x), f_j(x)) \leq d(f_n(x), f(x)) + d(f(x), f_j(x)) < \epsilon.\]

Conversely, assume \((f_n)\) satisfies the Uniform Cauchy Criterion. Then for each \(x \in X\), the sequence \((f_n(x))\) is Cauchy, so, since \(Y\) is complete, there exists \(f(x) \in Y\) such that \(f_n(x) \to f(x)\). Let \(\epsilon > 0\). Choose \(k \in \mathbb{N}\) such that \(d(f_n(x), f_j(x)) < \epsilon/2\) for all \(n, j \geq k, x \in X\). Fix \(n \geq k\) and \(x \in X\), and let \(j \to \infty\) to get

\[d(f_n(x), f(x)) \leq \frac{\epsilon}{2} < \epsilon.\]  

Now we’ll start to see all the good things that uniform convergence gives us. Roughly speaking, uniform convergence allows us to switch the order of two limits. Our first result along these lines says that a uniform limits preserve continuity:

**Theorem 19.4.** Let \(f, f_1, f_2, \ldots : X \to Y\) and \(t \in X\). If \(f_n \to f\) uniformly and each \(f_n\) is continuous at \(t\), then \(f\) is also continuous at \(t\).
Proof. Let $\epsilon > 0$. Choose $k \in \mathbb{N}$ such that $d(f_k(x), f(x)) < \epsilon/3$ for all $x \in X$. Now choose $\delta > 0$ such that $d(f_k(x), f_k(t)) < \epsilon/3$ for all $d(x, t) < \delta$. Then

$$d(f(x), f(t)) \leq d(f(x), f_k(x)) + d(f_k(x), f_k(t)) + d(f_k(t), f(t)) < \epsilon$$

for all $(x, t) < \delta$. □

Here's a variation:

Exercise 19.5. Let $X$ and $Y$ be metric spaces, $f, f_1, f_2, \ldots : X \to Y$, and $x_1, x_2, \ldots \in X$. Suppose $f_n \to f$ uniformly and $f(x_n) \to y$. Prove that $f_n(x_n) \to y$.

And here's another:

Exercise 19.6. Let $A \subset X, f, f_1, f_2, \ldots : A \to Y$, and $t \in A'$. Suppose $f_n \to f$ uniformly, and for each $n \in \mathbb{N}$ the limit $\lim_{x \to t} f_n(x)$ exists. Prove that $\lim_{x \to t} f(x)$ exists.

Our next result says that uniform limits preserve integrals:

**Theorem 19.7.** 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. Let $\epsilon > 0$. Choose $k \in \mathbb{N}$ such that

$$|f_n(x) - f_j(x)| < \frac{\epsilon}{2(b-a)} \quad \text{for all } n, j \geq k, x \in [a, b].$$

Then

$$\left| \int_a^b f_n - \int_a^b f_j \right| \leq \int_a^b |f_n - f_j| \leq \frac{\epsilon}{2(b-a)}(b-a) = \frac{\epsilon}{2} < \epsilon$$

for all $n, j \geq k$. Thus $I := \lim_n \int_a^b f_n$ exists.

Now we show $f$ is integrable and $\int_a^b f = I$. Let $\epsilon > 0$. 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 \( P = \{x_i\}_{i=0}^n \) such that if \( t_i \in [x_{i-1}, x_i] \) for all \( i \) then

\[
\left| \sum_i f_k(t_i) \Delta x_i - \int_a^b f_k \right| < \frac{\epsilon}{3}.
\]

Then

\[
\left| \sum f(t_i) \Delta x_i - I \right| \leq \left| \sum f(t_i) \Delta x_i - \sum f_k(t_i) \Delta x_i \right|
+ \left| \sum f_k(t_i) \Delta x_i - \int_a^b f_k \right| + \left| \int_a^b f_k - I \right|
< \sum |f(t_i) - f_k(t_i)| \Delta x_i + \frac{2\epsilon}{3}
< \frac{\epsilon}{3(b - a)} (b - a) + \frac{2\epsilon}{3}
= \epsilon. \quad \square
\]

However, the above result does not hold for improper integrals:

**Exercise 19.8.** Find an example of a sequence \((f_n)\) of improperly integrable functions on \([0, \infty)\) such that \(f_n \to 0\) uniformly but \(\int_0^\infty f_n \not\to 0\).

**Exercise 19.9.** Let

\[ f_n(x) = nxe^{-nx} \quad \text{for } n \in \mathbb{N}. \]

(a) Prove that \(\int_0^1 f_n \to 0\).

(b) Does \(f_n \to 0\) uniformly on \([0, 1]\)?

Now for differentiability. Here we have much worse luck: it’s definitely false that \(f_n \to f\) uniformly implies \(f_n' \to f'\):

**Exercise 19.10.** Let \(f_n(x) = \tan^{-1}(nx)/n\). Prove that \(f_n \to 0\) uniformly on \(\mathbb{R}\) but \(f_n'(0) \not\to 0\).

As the above exercise shows, differentiability is much more delicate than continuity or integrability. What we want is a sufficient condition to ensure that if \(f_n \to f\) and each \(f_n\) is differentiable, then \(f\) is differentiable and \(f_n' \to f'\). In the following theorem, it almost seems like we’re assuming the conclusion; it takes a little contemplation to see that the result gives some information:

**Theorem 19.11.** 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 the convergence \( f_n \to f \) is uniform.

Proof. By the Continuity Characterization of Differentiability, for each \( n \) there is a function \( q_n \) such that
\[
    f_n(x) = f_n(c) + q_n(x)(x - c) \quad \text{for all } x \in (a, b),
\]
and moreover \( q_n \) is continuous at \( c \) and \( q_n(c) = f'_n(c) \).

Claim: \((q_n)\) converges uniformly. Let \( \epsilon > 0 \). Choose \( k \in \mathbb{N} \) such that
\[
    |f'_n(x) - f'_j(x)| < \epsilon \quad \text{for all } n, j \geq k, x \in (a, b).
\]
If \( x \neq c \) then
\[
    q_n(x) - q_j(x) = \frac{(f_n - f_j)(x) - (f_n - f_j)(c)}{x - c} = f'_n(s) - f'_j(s)
\]
for some \( s \), by the Mean Value Theorem applied to the function \( f_n - f_j \). Thus \(|q_n(x) - q_j(x)| < \epsilon\) for all \( n, j \geq k \) and \( x \neq c \). On the other hand, for \( x = c \) we have \(|q_n(c) - q_j(c)| = |f'_n(c) - f'_j(c)| < \epsilon \), and our claim is verified.

Thus the sequence \((q_n)\) converges uniformly to a function \( q \) which is continuous at \( c \) since each \( q_n \) is. Letting \( n \to \infty \) in (5), we see that \((f_n)\) converges pointwise to a function \( f \) which satisfies
\[
    f(x) = f(c) + q(x)(x - c) \quad \text{for all } x \in (a, b).
\]
Since \( q \) is continuous at \( c \), \( f \) is differentiable there (by the Continuity Characterization of Differentiability again), and
\[
    f'(c) = q(c) = \lim_{n \to \infty} q_n(c) = \lim_{n \to \infty} f'_n(c).
\]

Now for a trick: since we know now that \( f_n \to f \), we can play the same game with \( c \) replaced by any point \( t \in (a, b) \), and we find that \( f \) is differentiable at \( t \) and \( f'_n(t) \to f'(t) \).

For the other part, assuming \((a, b)\) is bounded, we have
\[
    |f_n(x) - f(x)| \leq |f_n(c) - f(c)| + |q_n(x) - q(x)||x - c| \\
    \leq |f_n(c) - f(c)| + |q_n(x) - q(x)|(b - a),
\]
which converges uniformly to 0 since \(|q_n(x) - q(x)|\) does (and of course \(|f_n(c) - f(c)| \to 0\)). \(\square\)

Exercise 19.12. The last part of the above proof left a minor detail for you to check, but let’s strip away some of the notation and prove a general lemma: suppose \((g_n)\) is a uniformly convergent sequence of real-valued functions on some set \( A \), \((t_n)\) is a convergent sequence of
\footnote{Actually, uniformly, but you don’t get any extra points for pointing this out.}
real numbers, and \( d \) is a real number. Prove that the sequence \( (h_n) \)
defined by \( h_n(x) = t_n + dg_n(x) \) converges uniformly on \( A \).

In fact, uniformly convergent sequences of real-valued functions are
closed under addition and scalar multiplication, but not multiplication
of functions:

**Exercise 19.13.** Find an example of a uniformly convergent sequence
\( (f_n) \) of real-valued functions such that \( (f_n^2) \) doesn’t converge uniformly.

However, it follows from the next exercise that this sort of unpleasantness can be avoided by restricting the functions appropriately:

**Exercise 19.14.** Let \( c, d \in \mathbb{R} \) with \( c < d \), and let \( f, f_1, f_2, \ldots : X \to [c, d] \). Also let \( g : [c, d] \to \mathbb{R} \) be continuous. Suppose \( f_n \to f \) uniformly. Prove that \( g \circ f_n \to g \circ f \) uniformly.

In our above theorem on uniform convergence and derivatives, we
showed that \( (f_n) \) converges uniformly if the interval \((a, b)\) is bounded.
This is necessary:

**Exercise 19.15.** Find an example of a sequence \( (f_n) \) of real-valued
functions on \( \mathbb{R} \) such that:

- \( (f'_n) \) converges uniformly on \( \mathbb{R} \);
- \( (f_n) \) converges pointwise on \( \mathbb{R} \);
- \( (f_n) \) does not converge uniformly on \( \mathbb{R} \).

**Definition.** \( C_b(X) \) denotes the vector space of bounded continuous
real-valued functions on \( X \) (with the usual pointwise operations). For
\( f \in C_b(X) \), define

\[
\|f\| = \sup\{|f(x)| \mid x \in X\}.
\]

If \( X \) is compact, just write \( C(X) \). If \( X = [a, b] \) is a compact interval,
just write \( C[a, b] \).

**Theorem 19.16.** \( C_b(X) \) is a Banach space. Moreover, convergence in
the metric space \( C_b(X) \) is the same as uniform convergence.

**Proof.** First we show that \( \| \cdot \| \) is a norm. Since the absolute value
function on \( \mathbb{R} \) is nonnegative, \( \|f\| \geq 0 \) for all \( f \). Moreover, if \( \|f\| = 0 \)
then \( f(x) = 0 \) for all \( x \in X \), so \( f = 0 \). Next, if \( f \in C_b(X) \) and \( c \in \mathbb{R} \) then

\[
\|cf\| = \sup\{|cf(x)| \mid x \in X\}
\]

\[
= \sup\{|c||f(x)| \mid x \in X\}
\]

\[
= |c| \sup\{|f(x)| \mid x \in X\}
\]

\[
= |c|\|f\|.
\]
MAT 472 LECTURES

For the triangle inequality, if \( f, g \in C_b(X) \) then
\[
\|f + g\| = \sup\{|(f + g)(x)| \mid x \in X\}
= \sup\{|f(x) + g(x)| \mid x \in X\}
\leq \sup\{|f(x)| + |g(x)| \mid x \in X\}
\leq \sup\{|f(x)| \mid x \in X\} + \sup\{|g(x)| \mid x \in X\}
= \|f\| + \|g\|.
\]

Thus \( C_b(X) \) is a normed space. Let \( f, f_1, f_2, \ldots \in C_b(X) \). Since \( f_n \to f \) uniformly if and only if \( \sup_{x \in X} |f_n(x) - f(x)| \to 0 \), convergence in the metric space \( C_b(X) \) is the same as uniform convergence.

Now we show \( C_b(X) \) is complete. Let \( (f_n) \) be a Cauchy sequence in \( C_b(X) \). Claim: for each \( x \in X \), the sequence \( (f_n(x)) \) of real numbers is Cauchy. Fix \( x \in X \), and let \( \epsilon > 0 \). Choose \( k \in \mathbb{N} \) such that \( \|f_n - f_j\| < \epsilon \) for all \( n, j \geq k \). Since \( |f_n(x) - f_j(x)| \leq \|f_n - f_j\| \), the claim follows. Since \( \mathbb{R} \) is complete, there exists \( f(x) \) such that \( f_n(x) \to f(x) \). We have defined a function \( f : X \to \mathbb{R} \). We must show \( f \in C_b(X) \) and \( f_n \to f \) in \( C_b(X) \). Since \( (f_n) \) is Cauchy, it’s bounded. Thus there exists \( M \in \mathbb{R} \) such that \( \|f_n\| \leq M \) for all \( n \in \mathbb{N} \). Then
\[
|f(x)| = \lim_{n \to \infty} |f_n(x)| \leq M \quad \text{for all } x \in X.
\]

Thus \( f \) is bounded. We next show that \( f_n \to f \) uniformly. Let \( \epsilon > 0 \). Choose \( k \in \mathbb{N} \) such that \( \|f_n - f_j\| < \epsilon/2 \) for all \( n, j \geq k \). Let \( x \in X \). Then
\[
|f_n(x) - f_j(x)| \leq \|f_n - f_j\| < \epsilon/2 \quad \text{for all } n, j \geq k,
\]
so for all \( n \geq k \) we have
\[
|f_n(x) - f(x)| = \lim_{j \to \infty} |f_n(x) - f_j(x)| \leq \frac{\epsilon}{2} < \epsilon,
\]
and we have shown \( f_n \to f \) uniformly. Since each \( f_n \) is continuous, so is \( f \). Therefore \( f \in C_b(X) \), and \( f_n \to f \) in the metric space \( C_b(X) \) since the convergence is uniform. \( \square \)

The metric spaces of the form \( C_b(X) \) are among the most important in modern analysis. They take some getting used to, because they have unexpected properties. For example: it’s easy to visualize finite discrete spaces\(^\text{74}\) but infinite discrete spaces might seem rather contrived; however, they can be “visualized” inside spaces like \( C_b(X) \), for example:

\(^\text{74}\)The set of standard basis vectors in \( \mathbb{R}^n \), regarded as a subspace, has the discrete metric.
Exercise 19.17. Find an example of an infinite subset $F$ of $C[0,1]$ such that $\|f - g\| = 1$ whenever $f$ and $g$ are distinct elements of $F$.

In the Banach space $\mathbb{R}^n$, every closed bounded set is compact, by the Heine-Borel Theorem. This is definitely false in the Banach space $C[0,1]$: 

Exercise 19.18. Prove that there exists a closed bounded subset of $C[0,1]$ which is not compact. Hint: you may use the result of the preceding exercise.

In fact, the (deep) Arzela-Ascoli Theorem below will give a characterization of the compact subsets of $C[0,1]$ (in fact $C(X)$ for any compact metric space $X$), but to state it we’ll need a couple of new concepts:

Definition. Let $A$ be a set of real-valued functions on $X$. Then $A$ is:

(i) equicontinuous if for all $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $f \in A$, $d(x, y) < \delta$;

(ii) pointwise bounded if for every $x \in X$ the set $\{f(x) \mid f \in A\}$ is bounded in $\mathbb{R}$.

Arzela-Ascoli Theorem. Let $X$ be compact and $A \subset C(X)$. Then $\overline{A}$ is compact in $C(X)$ if and only if $A$ is equicontinuous and pointwise bounded.

Proof. First assume $\overline{A}$ is compact. For equicontinuity, let $\epsilon > 0$. By compactness there exist $f_1, \ldots, f_k \in \overline{A}$ such that

$$\overline{A} \subset \bigcup_{1}^{k} B_{\epsilon/3}(f_n).$$

Since each $f_n$ is continuous on the compact space $X$, it’s uniformly continuous, and so, taking the minimum of finitely many positive numbers, we can find $\delta > 0$ such that $|f_n(x) - f_n(y)| < \epsilon/3$ whenever $n = 1, \ldots, k, d(x, y) < \delta$.

If $f \in A$ and $d(x, y) < \delta$, there exists $n = 1, \ldots, k$ such that $\|f - f_n\| < \epsilon/3$, and then

$$|f(x) - f(y)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| < \epsilon.$$

For pointwise boundedness, note that $\overline{A}$ is bounded in $C(X)$ by compactness, so there exists $M \in \mathbb{R}$ such that $\|f\| \leq M$ for all $f \in \overline{A}$. In particular, for each $x \in X$,

$$|f(x)| \leq M \quad \text{for all } f \in A.$$
Conversely, assume $A$ is equicontinuous and pointwise bounded. Claim: $\overline{A}$ is also equicontinuous and pointwise bounded. For equicontinuity, let $\epsilon > 0$. Choose $\delta > 0$ such that

$$|f(x) - f(y)| < \frac{\epsilon}{2} \quad \text{whenever } f \in A, d(x, y) < \delta.$$ 

Let $f \in \overline{A}$, and choose a sequence $(f_n)$ in $A$ converging to $f$. Then $d(x, y) < \delta$ implies

$$|f(x) - f(y)| = \lim_{n \to \infty} |f_n(x) - f_n(y)| \leq \frac{\epsilon}{2} < \epsilon.$$ 

Similarly, for pointwise boundedness, let $x \in X$. Choose $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $f \in A$.

Let $f \in \overline{A}$, and choose a sequence $(f_n)$ in $A$ converging to $f$. Then

$$|f(x)| = \lim_{n \to \infty} |f_n(x)| \leq M.$$ 

Thus, without loss of generality $A$ is closed. To show that $A$ is compact, it suffices to show that every sequence $(f_n)$ in $A$ has a uniformly convergent subsequence (because this subsequence will then converge in $C(X)$, and since $A$ is closed the limit will be in $A$).

We use a trick: since $X$ is compact, it has a countable dense subset $\{x_i \mid i \in \mathbb{N}\}$. Claim: $(f_n)$ has a subsequence which converges pointwise on $D$. We use the Cantor Diagonalization Argument. Since the sequence $(f_n(x_1))_{n=1}^{\infty}$ is bounded in $\mathbb{R}$, there is a subsequence $(f_{1n})$ of $(f_n)$ such that $(f_{1n}(x_1))$ converges in $\mathbb{R}$. Similarly, there is a subsequence $(f_{2n})$ of $(f_{1n})$ such that $(f_{2n}(x_2))$ converges in $\mathbb{R}$. Continue inductively, so that for each $i \in \mathbb{N}$ we get a sequence $(f_{in})_{n=1}^{\infty}$ such that $(f_{i+1,n})_{n=1}^{\infty}$ is a subsequence of $(f_{in})_{n=1}^{\infty}$ and $(f_{in}(x_i))_{n=1}^{\infty}$ converges in $\mathbb{R}$. Now put $g_n = f_{mn}$. This is a subsequence of $(f_n)$. For each $i \in \mathbb{N}$, the sequence $(g_n(x_i))_{n=1}^{\infty}$ converges in $\mathbb{R}$ since the tail $(g_n(x_i))_{n=i}^{\infty}$ is a subsequence of $(f_{in}(x_i))_{n=1}^{\infty}$. This proves the claim.

It now suffices to show that the subsequence $(g_n)$ of $(f_n)$ converges uniformly on $X$. Let $\epsilon > 0$. By equicontinuity there exists $\delta > 0$ such that

$$|g_n(x) - g_n(y)| < \frac{\epsilon}{3} \quad \text{whenever } n \in \mathbb{N}, d(x, y) < \delta.$$ 

By compactness of $X$ and density of $D$ there exists a finite subset $S$ of $\mathbb{N}$ such that

$$X = \bigcup_{i \in S} B_\delta(x_i).$$
By pointwise convergence, we can take the maximum of finitely many natural numbers to find $k \in \mathbb{N}$ such that
\[ |g_n(x_i) - g_j(x_i)| < \frac{\epsilon}{3} \quad \text{for all } n, j \geq k, i \in S. \]
If $n, j \geq k$ and $x \in X$, there exists $i \in S$ such that $d(x, x_i) < \delta$, and then
\[ |g_n(x) - g_j(x)| \leq |g_n(x) - g_n(x_i)| + |g_n(x_i) - g_j(x_i)| + |g_j(x_i) - g_j(x)| < \epsilon. \]
Thus $(g_n)$ is a uniformly convergent subsequence of $(f_n)$, and we are done. \qed

In the proof of the converse direction in the Arzela-Ascoli Theorem, we showed that $\overline{A}$ is equicontinuous and pointwise bounded if $A$ is. We could have avoided this using the following general fact:

**Exercise 19.19.** Let $X$ be a metric space and $A \subset X$. Prove that $\overline{A}$ is compact if and only if every sequence in $A$ has a subsequence which converges in $X$.

**Exercise 19.20.** Put
\[ A = \{ f \in C[0, 1] \mid f(0) = 3 \text{ and } |f'(x)| \leq 5 \text{ for all } x \in (0, 1) \}. \]
Prove that every sequence $(f_n)$ in $A$ has a uniformly convergent subsequence.

**Exercise 19.21.** Is the set $\{ x^n \mid n \in \mathbb{N} \}$ of functions in $C[0, 1]$ equicontinuous?

**Exercise 19.22.** Let $(f_n)$ be a sequence of Riemann integrable functions on $[a, b]$, and suppose that there exists $M \in \mathbb{R}$ such that $|f_n(x)| \leq M$ for all $n \in \mathbb{N}$ and $x \in [a, b]$. Put $g_n(x) = \int_a^x f_n$. Prove that $(g_n)$ has a subsequence which converges uniformly on $[a, b]$.

The following theorem is also very deep. We’ll see later that many functions can be expressed as power series, which are certain kinds of limits of polynomials. But the class of functions we get this way is fairly restrictive: they are all infinitely differentiable, for example. Polynomials are so nice, it’s useful to be able to use them to approximate other functions. Of course, if a function is not differentiable, we can’t hope to approximate in the special manner of power series. But now a miracle occurs: if all we want to do is approximate uniformly, we only need continuity:

**Weierstrass Approximation Theorem.** The polynomials are dense in $C[a, b]$. 
Proof. Let \( f \in C[a,b] \). Without loss of generality \([a,b] = [0,1]\) and \( f(0) = f(1) = 0 \). Extend \( f \) to be 0 on \([0,1]^c\). Then \( f \) is bounded and uniformly continuous on \( \mathbb{R} \).

Define polynomials \( q_n \) by
\[
q_n(x) = c_n(1 - x^2)^n,
\]
where \( c_n \) is chosen so that \( \int_{-1}^{1} q_n = 1 \) for all \( n \in \mathbb{N} \). Next, define \( p_n : [0,1] \to \mathbb{R} \) by
\[
p_n(x) = \int_{-1}^{1} f(x+t)q_n(t) \, dt.
\]

For \( x \in [0,1] \),
\[
p_n(x) = \int_{x-1}^{x+1} f(t)q_n(t-x) \, dt = \int_{0}^{1} f(t)q_n(t-x) \, dt,
\]
so \( p_n \) is a polynomial.

Let \( \epsilon > 0 \). Choose \( \delta > 0 \) such that
\[
|f(x) - f(y)| < \frac{\epsilon}{2}
\]
whenever \( |x-y| < \delta \), and without loss of generality \( \delta \leq 1 \). By continuity, if \( |x-y| \leq \delta \) then \( |f(x) - f(y)| \leq \epsilon/2 \). Put \( M = \sup |f| \). Then for all \( x \in [0,1] \),
\[
|p_n(x) - f(x)| = \left| \int_{-1}^{1} (f(x+t) - f(x))q_n(t) \, dt \right|
\leq \int_{-1}^{1} |f(x+t) - f(x)|q_n(t) \, dt
\leq 2M \int_{-1}^{-\delta} q_n + \frac{\epsilon}{2} \int_{-\delta}^{\delta} q_n + 2M \int_{\delta}^{1} q_n
\leq 4M \int_{\delta}^{1} q_n + \frac{\epsilon}{2}.
\]
Thus it suffices to show \( q_n \to 0 \) uniformly on \([\delta,1]\). For \( x \in [\delta,1] \),
\[
|q_n(x)| = q_n(x) \leq c_n(1 - \delta^2)^n.
\]
Now,
\[
\int_{-1}^{1} (1 - x^2)^n \, dx \geq \int_{0}^{1} (1 - x^2)^n \, dx
\geq \int_{0}^{1} (1 - x)^n \, dx
= \frac{1}{n+1},
\]
so we have

\[ c_n \leq n + 1, \]

hence \( c_n(1 - \delta^2)^n \to 0. \) Therefore \( q_n \to 0 \) uniformly on \([\delta, 1]\), as desired.

The Weierstrass Approximation Theorem tells us a lot about the metric space \( C[a, b] \). For example:

**Exercise 19.23.** Use the Weierstrass Approximation Theorem to prove that \( C[a, b] \) is separable.

So, we can approximate any continuous function uniformly on any bounded interval. What about unbounded intervals? Nope:

**Exercise 19.24.** Let \( f : \mathbb{R} \to \mathbb{R} \), and suppose \((f_n)\) is a sequence of polynomials converging uniformly to \( f \) on \( \mathbb{R} \). Prove that \( f \) is a polynomial.
Standing Hypothesis. Throughout this section all functions will be real-valued, unless otherwise specified.

Notation and Terminology. As before, the letter $X$ will refer to a metric space, unless otherwise specified.

Definition. Given a sequence $(f_n)$ of real-valued functions on $X$, the series $\sum_{n=1}^{\infty} f_n$ is the sequence $(\sum_{n=1}^{k} f_n)_{k=1}^{\infty}$ of partial sums.

Thus, a series of functions is the same as the sequence of partial sums, so we can apply all terminology and results concerning sequences of functions. In particular, we can talk about the series $\sum f_n$ converging pointwise or uniformly. Also, the sum is the pointwise limit of the sequence of partial sums.

Also, we can apply terminology and results concerning real series to series of functions by evaluating the terms at points of $X$, for example:

Definition. A series $\sum f_n$ of real-valued functions on $X$ converges absolutely if for every $x \in X$ the real series $\sum f_n(x)$ does.

Uniform Cauchy Criterion for Series. $\sum f_n$ converges uniformly on $X$ if and only if for all $\epsilon > 0$ there exists $l \in \mathbb{N}$ such that

$$\left| \sum_{j}^{k} f_n(x) \right| < \epsilon \quad \text{for all } k \geq j \geq l, \ x \in X.$$  

Proof. This follows immediately from the Uniform Cauchy Criterion for sequences of functions.

Weierstrass $M$-Test. Assume there exists a nonnegative sequence $(M_n)$ in $\mathbb{R}$ such that both

(i) $\sum M_n$ converges, and

(ii) $|f_n(x)| \leq M_n$ for all $n \in \mathbb{N}, x \in X$.

Then $\sum f_n$ converges uniformly and absolutely on $X$.

Proof. Let $\epsilon > 0$. Choose $l \in \mathbb{N}$ such that $\sum_{j}^{k} M_n < \epsilon$ for all $k \geq j \geq l$. Then

$$\left| \sum_{j}^{k} f_n(x) \right| \leq \sum_{j}^{k} |f_n(x)| \leq \sum_{j}^{k} M_n < \epsilon \quad \text{for all } x \in X, k \geq j \geq l.$$

Exercise 20.1. Prove that $\sum x^n$ converges uniformly on every interval of the form $[-s, s]$ with $0 < s < 1$. 
Exercise 20.2. Find an example of a series of functions which converges uniformly on \( \mathbb{R} \) but does not satisfy the hypotheses of the Weierstrass M-Test. Hint: try constant functions.

**Theorem 20.3.** 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 19.4.

**Theorem 20.4.** 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 19.7.

**Theorem 20.5.** 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 19.11.

**Exercise 20.6.** Prove:

(a) \( \sum_{n=1}^{\infty} \frac{1}{(x+n^2)^2} \) converges uniformly on \((0, 1)\).

(b) \( \sum_{n=1}^{\infty} \frac{1}{x+n^2} \) converges on \((0, 1)\) to a differentiable function, and

\[
\frac{d}{dx} \sum_{n=1}^{\infty} \frac{1}{x+n^2} = -\sum_{n=1}^{\infty} \frac{1}{(x+n^2)^2} \quad \text{for all } x \in (0, 1).
\]

Hint: show that the series converges for \( x = 1/2 \).

**Exercise 20.7.** Define \( f : [0, 1] \to \mathbb{R} \) as follows: first of all, \( f(x) = 1 \) if \( x = 0 \) and \( f(x) = 0 \) if \( x \notin \mathbb{Q} \). For every nonzero rational \( x \in [0, 1] \), there exist unique \( n, k \in \mathbb{N} \) such that \( x = n/k \) in lowest terms, that is, \( n \) and \( k \) have no common prime divisors; in this case define \( f(x) = 1/k \).

Prove:

(a) The discontinuities of \( f \) are precisely the rationals in \([0, 1]\).

(b) \( f \) is integrable.
21. Power series

Definition. 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 the power series with coefficients \(c_n\) and center \(a\).

Cauchy-Hadamard Theorem. Let \(\sum_{n=0}^{\infty} c_n(x-a)^n\) be a power series, and put

\[ r = \lim \left| \frac{c_n}{|x-a|^{1/n}} \right| = \begin{cases} 0 & \text{if the lim sup is } \infty, \\ \infty & \text{if the lim sup is } 0. \end{cases} \]

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

\[ \lim |c_n(x-a)^n|^{1/n} = |x-a| \lim |c_n|^{1/n} = \frac{|x-a|}{r}, \]

so the first part follows from the Root Test.

For the other part, let \(0 < s < r\), and pick \(t \in (s, r)\). Then

\[ \frac{1}{t} > \frac{1}{r} = \lim |c_n|^{1/n}, \]

so there exists \(k \in \mathbb{N}\) such that

\[ |c_n|^{1/n} < \frac{1}{t} \quad \text{for all } n \geq k. \]

Thus

\[ |c_n(x-a)^n| \leq \left( \frac{s}{t} \right)^n \quad \text{whenever } n \geq k, |x-a| < s. \]

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. With the above notation, \(r\) is the radius of convergence of the power series \(\sum c_n(x-a)^n\).

The power series converges at least in the open interval of convergence \((a-r, a+r)\), and possibly also at one or both of the endpoints \(a \pm r\) (depending upon the particular power series). The set of all values of \(x\) where the power series converges is also sometimes called the interval of convergence.

Exercise 21.1. Prove:
(a) The power series $\sum n^n x^n$ has radius of convergence 0.
(b) The power series $\sum x^n/n^n$ has radius of convergence $\infty$.

**Exercise 21.2.** Find all values of $x$ for which the series
\[ \sum_{n=1}^{\infty} \frac{x^n}{n^2} \]
converges.

**Exercise 21.3.** Consider the power series $\sum x^n/n^p$, where $p$ is a real number.

(a) Prove that the radius of convergence is 1 for every $p$.
(b) For what values of $p$ does the series converge at $x = 1$?
(c) For what values of $p$ does the series converge at $x = -1$?

**Exercise 21.4.** Let $\sum_{n=0}^{\infty} c_n x^n$ be a power series centered at 0. Let $r > 0$, and suppose the sequence $(c_n r^n)$ is bounded. Prove that the radius of convergence of the power series $\sum_{n=0}^{\infty} c_n x^n$ is at least $r$.

**Exercise 21.5.** Let $\sum c_n (x-a)^n$ be a power series, and suppose that the sequence $|c_n/c_{n+1}|$ converges to a real number $r$. Prove that $r$ is the radius of convergence of the power series (be sure to properly handle the special case $r = 0$).

On the other hand, prove that if the above sequence diverges to $\infty$, then the radius of convergence of the power series is $\infty$.

**Exercise 21.6.** Prove that the power series $\sum_{n=0}^{\infty} x^n/n!$ has radius of convergence $\infty$. You may use the result of the preceding exercise.

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 21.7.** If the power series $\sum c_n(x-a)^n$ has radius of convergence $r$, then both power series
\[ \sum nc_n(x-a)^{n-1} \quad \text{and} \quad \sum \frac{c_n}{n+1}(x-a)^{n+1} \]
also have radius of convergence $r$.

Moreover, let $r > 0$, and define $f : (a-r,a+r) \to \mathbb{R}$ by $f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n$. Then for all $x \in (a-r,a+r)$,
\[ f'(x) = \sum_{n=1}^{\infty} nc_n(x-a)^{n-1} \quad \text{and} \quad \int_a^x f = \sum_{n=0}^{\infty} \frac{c_n}{n+1}(x-a)^{n+1}. \]
Proof. First of all, multiplying by \(x - a\) term-by-term does not change the radius of convergence, so \(\sum nc_n(x - a)^{n-1}\) has the same radius of convergence as \(\sum nc_n(x - a)^n\). Since \(n^{1/n} \to 1\), the Cauchy-Hadamard Theorem implies that \(\sum nc_n(x - a)^n\) has the same radius of convergence as \(\sum c_n(x - a)^n\). Thus the power series \(\sum nc_n(x - a)^{n-1}\) 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 \(\sum nc_n(x - a)^n\) 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 20.5 (on differentiating series) tells us \(f\) is differentiable and

\[
f'(x) = \sum_{0}^{\infty} \frac{d}{dx} c_n(x - a)^n = \sum_{1}^{\infty} nc_n(x - a)^{n-1}.
\]

Similarly, since \(\sum c_n(t - a)^n\) converges uniformly for \(t\) in the closed interval with endpoints \(a\) and \(x\), Theorem 20.4 (on integrating series) tells us

\[
\int_{a}^{x} f(t) \, dt = \sum_{0}^{\infty} \int_{a}^{x} c_n(t - a)^n \, dt = \sum_{0}^{\infty} \frac{c_n}{n+1} (x - a)^{n+1}.
\]

Thus, the open interval of convergence \((a - r, a + r)\) of a power series is preserved by differentiating or integrating term-by-term. However, convergence at the endpoints is delicate, and can be created or destroyed. For example:

Example. The geometric series \(\sum x^n\) diverges at both endpoints \(\pm 1\), but the integrated series \(\sum x^{n+1}/(n+1)\) converges at \(-1\).

The above results show that functions which can be expressed as power series have lots of nice properties. Indeed, they have lots more. For example, their zeros are isolated:

Exercise 21.8. Let \(\sum_{n=0}^{\infty} c_n(x - a)^n\) be a power series with positive radius of convergence \(r\), and assume at least one \(c_n\) is nonzero. Show that there exists \(\delta \in (0, r)\) such that \(\sum_{n=0}^{\infty} c_n(x - a)^n \neq 0\) if \(0 < |x - a| < \delta\).

The above property (isolated zeros) is pretty special — it’s certainly not shared by all differentiable functions, for example:

Example. The function defined as \(x^2 \sin(1/x)\) for \(x \neq 0\) and 0 at 0 is differentiable on \(\mathbb{R}\) and has 0 as a cluster point of its zeros. In fact, this
example could be made as differentiable as we want just by replacing
$x^2$ by $x^n$ for any $n \geq 2$.

**Corollary 21.9.** 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

$$ f^{(n)}(x) = \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} c_k(x-a)^{k-n}. $$

Thus

$$ f^{(n)}(a) = n!c_n. \quad \square $$

**Definition.** If $f$ is infinitely differentiable at $a$, the power series

$$ \sum_{n=0}^{\infty} \frac{f^{(n)}(a)(x-a)^n}{n!} $$

is the Taylor series of $f$ at $a$. The special case $a = 0$ occurs so often
it’s given its own name: the MacLaurin series of $f$.

**Example.** Since $e^x$ is its own derivative, the MacLaurin series is $\sum_{n=0}^{\infty} \frac{x^n}{n!}$.

It’s often inconvenient to compute a Taylor series directly, because
the derivatives might be complicated. By the above corollary, if we can
somehow find a power series representation of a function, that must be
the Taylor series.

**Exercise 21.10.** We know that the MacLaurin series of $1/(1 - x)$ is
the geometric series $\sum_{n=0}^{\infty} x^n$.

(a) Find the MacLaurin series of $1/(1 + x)$ by substitution.

(b) Use this to prove that

$$ \log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots \quad \text{for all } x \in (-1, 1). $$

**Exercise 21.11.** Find the MacLaurin series of $1/(1 - x)^2$.

**Exercise 21.12.** (a) Find the MacLaurin series of $1/(1 + x^2)$.

(b) Use this to prove that

$$ \tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots \quad \text{for all } x \in (-1, 1). $$

---

And by this I mean either write the series in summation notation with a formula
for the $n$th term or show at least the first three terms of the series in a simplified
form which clearly indicates what the pattern is (if there is a pattern to the terms).
Note that the partial sums of a power series are polynomials. The partial sum \( \sum_{k=0}^{n} \frac{f^{(k)}(a)(x-a)^{k}}{k!} \) of the Taylor series is sometimes called the \( n^{th} \) degree Taylor polynomial, although it might have degree smaller than \( n \) because the \( n^{th} \) derivative of \( f \) at \( a \) could be 0.

It is natural to ask how well the Taylor polynomials approximate the function \( f \). The following result, a generalization of the Mean Value Theorem to higher order derivatives, is the main tool used to answer this question:

**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 \( g: I \rightarrow \mathbb{R} \) by

\[
g(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)(x-a)^{k}}{k!} + M(t-a)^{n},
\]

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. \( \square \)

The difference between the function and its \( n^{th} \) degree Taylor polynomial is sometimes called the \( n^{th} \) remainder term, and the above result says that the remainder looks similar to the next term in the
Taylor series. At a given point \( x \), the Taylor series of \( f \) converges to \( f(x) \) if and only if the remainder terms go to 0 as \( n \to \infty \). Taylor’s Theorem can also tell us how the Taylor series converges to \( f \); for example:

**Example.** Suppose \( |f^{(n+1)}| \) is bounded above by \( M \) on an open interval \( I \) containing \( a \), and let \( p \) be the \( n \)th degree Taylor polynomial of \( f \) at \( a \). Then

\[
|f(x) - p(x)| \leq M \frac{|x-a|^{n+1}}{(n+1)!} \quad \text{for all } x \in I,
\]

since for each \( x \in I \) there exists \( c \) between \( x \) and \( a \) such that

\[
f(x) - p(x) = \frac{f^{(n+1)}(c)(x-a)^{n+1}}{(n+1)!}.
\]

This shows that the approximation \( f(x) \approx p(x) \) tends to be better for \( x \) close to \( a \).

**Exercise 21.13.** Prove that for all \( x \in \mathbb{R} \),

\[
|\sin x - x + \frac{x^3}{3!}| \leq \frac{|x|^5}{5!}.
\]

The main use of Taylor’s Theorem is to show that the Taylor series converges to the function.

**Example.** We’ve already see that the Maclaurin series of \( e^x \) is \( \sum_{n=0}^{\infty} \frac{x^n}{n!} \). The Ratio Test shows that the radius of convergence of this series is \( \infty \). Surprisingly, this is one of the easiest ways to show that

\[
\lim_{n \to \infty} \frac{x^n}{n!} = 0 \quad \text{for all } x \in \mathbb{R}.
\]

However, to show that the series converges to \( e^x \) requires us to analyze the remainder term, which you’ll do in the following exercise.

**Exercise 21.14.** Use Taylor’s Theorem to prove that

\[
e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \quad \text{for all } x \in \mathbb{R}.
\]

The preceding exercise is not quite a special case of:

**Exercise 21.15.** Let \( I \) be an open interval containing \( a \), and suppose there exists \( M \in \mathbb{R} \) such that

\[
|f^{(n)}(x)| \leq M^n \quad \text{for all } n \in \mathbb{N}, x \in I.
\]

Prove that the Taylor series centered at \( a \) converges to \( f \) at every \( x \) in \( I \).
We shouldn’t let ourselves be too complacent, however — somehow irritating, even if the Taylor series of \( f \) converges at \( x \), it might not converge to \( f(x) \). In fact, the Taylor series of \( f \) may have a positive radius of convergence but only converge to \( f \) at the center \( a \). For example:

**Exercise 21.16.** Let

\[
f(x) = \begin{cases} 
e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}
\]

Prove:

(a) For each \( n = 0, 1, \ldots \) there exists a rational function \( g \) such that

\[
f^{(n)}(x) = \begin{cases} g(x)e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}
\]

Hint: use induction and l’Hôpital’s Rule.

(b) The MacLaurin series of \( f \) has radius of convergence \( \infty \).

(c) The MacLaurin series of \( f \) converges to \( f(x) \) only at \( x = 0 \).

**Exercise 21.17.** Show that

\[
\sum_{n=0}^{\infty} \int_0^1 \frac{x^{2n}}{n!} e^{-x^2} \, dx = 1.
\]

Remember to give the reasons justifying your steps.

Suppose we know that the Taylor series converges to \( f(x) \) whenever \( |x - a| < r \), where \( r \) is the radius of convergence. And suppose the Taylor series happens to converge at an endpoint. Does it converge to \( f \) there? None of our general theory so far tells us anything about this. Fortunately, a miracle occurs: if \( f \) is continuous at the endpoint, then the Taylor series converges to \( f \) there. This can be proven using the following special case:

**Exercise 21.18.** In this exercise you’ll prove **Abel’s Theorem**: Let \( \sum_0^{\infty} c_n x^n \) be a MacLaurin series with radius of convergence 1, and suppose \( \sum_0^{\infty} c_n \) converges. Then \( \sum_0^{\infty} c_n x^n \) converges uniformly on \([0, 1]\).

All you have to do is:

(a) Prove that it’s enough to show uniform convergence on \([0, 1]\).

(b) For each \( k \) let \( b_k = \sum_k^{\infty} c_n \). Prove that \( \sum_0^{\infty} b_n x^n \) converges absolutely for all \( x \in [0, 1] \) (hint: show that the coefficients \( b_n \) are bounded).
(c) Prove that
\[ \sum_{k} c_n x^n = b_k x^k + x^k (x - 1) \sum_{n=0}^{\infty} b_{n+k+1} x^n \quad \text{for all } k \in \mathbb{N}, x \in [0, 1). \]

Don’t forget to say why your manipulations with series are valid.

(d) Prove that for all \( \epsilon > 0 \) there exists \( l \in \mathbb{N} \) such that
\[ \left| \sum_{k}^{\infty} c_n x^n \right| < \epsilon \quad \text{for all } k \geq l, x \in [0, 1). \]

Here’s everyone’s favorite application of Abel’s Theorem:

Example. We’ve seen that \( \log(1 + x) = x - x^2/2 + x^3/3 - \cdots \) for all \( x \in (-1, 1) \). The series diverges at \(-1\), but at 1 it’s the alternating harmonic series, which converges. Since \( \log(1 + x) \) is continuous at \( x = 1 \), Abel’s Theorem tells us the sum of the alternating harmonic series is \( \log 2 \).
22. Trigonometry

We’ve been using the trig functions, especially sin and cos, but we never rigorously established their properties. We’ll see now that this can be done using power series.

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 \( \sum_{0}^{\infty} (|c_1| + |c_2|)|x|^n/n! \) 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
\[
\cos x = \sum_{0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!},
\]
while letting \( c_1 = 0 \) and \( c_2 = 1 \), we define \( \sin: \mathbb{R} \to \mathbb{R} \) by
\[
\sin x = \sum_{0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}.
\]

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 MacLaurin series of \( f \) 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
\[
\left| f(x) - \sum_{0}^{n-1} \frac{f^{(k)}(0)x^k}{k!} \right| = \left| \frac{f^{(n)}(c)x^n}{n!} \right| \leq \frac{M|x|^n}{n!} \xrightarrow{n \to \infty} 0,
\]
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 similarly (because the series for $\cos$ has only even powers)
\[
\cos(-x) = \cos x.
\]

Letting $x = 0$ in the Taylor series, we get
\[
\sin 0 = 0 \text{ and } \cos 0 = 1.
\]

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
\[
\frac{x^n}{n!} \geq \frac{x^{n+2}}{(n+2)!} > 0,
\]
hence the Taylor series for $\sin$ satisfies the (strict version of the) hypotheses of the Alternating Series Test. Thus
\[
\sin x > 0 \quad \text{whenever } \quad 0 < x \leq 2,
\]
so $\cos$ is strictly decreasing on $[0,2]$. Now,
\[
\cos 2 = \sum_{0}^{\infty} \frac{(-1)^n 2^{2n}}{(2n)!} = -1 + \sum_{2}^{\infty} \frac{(-1)^n 2^{2n}}{(2n)!},
\]
and by the Alternating Series Test
\[ \sum_{n=2}^{\infty} \frac{(-1)^n 2^{2n}}{(2n)!} < \frac{2^4}{4!} = \frac{2}{3}, \]
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 \( 1 = \sin^2 \frac{\pi}{2} + \cos^2 \frac{\pi}{2} = \sin^2 \frac{\pi}{2} \) and \( \sin \frac{\pi}{2} > 0 \), 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.
Abel’s Theorem, 183
absolute value, 27
algebraic, 41
alternate triangle inequality, 27, 44
alternating, 156
alternating harmonic series, 157
Alternating Series Test, 156
Approximation of Suprema, 30
Archimedean Principle, 31
Arithmetic of Continuity, 83
Arithmetic of Convergence, 63
Arithmetic of Infinite Limits, 66
Arithmetic of Limits, 75
Arzela-Ascoli Theorem, 170
Baire Category Theorem, 92
Banach space, 89
base, 94
between, 24
Bolzano-Weierstrass Theorem, 62, 90
boundary, 52
boundary point, 52
bounded, 56
bounded above, 29, 57
bounded below, 29
Cantor Diagonalization Argument, 41
Cantor Set, 100
Cartesian product, 13, 15
Cauchy, 88
series, 150
uniform, 164, 175
Cauchy’s Mean Value Theorem, 117
Cauchy-Hadamard Theorem, 177
Cauchy-Schwartz Inequality, 46
$C_b$, 108
center, 46, 177
Chain Rule, 116
Change of Variables Theorem, 140
Characterization
  of Connected Subsets of $\mathbb{R}$, 107
  of Open Subsets of $\mathbb{R}$, 108
closed
form, 150
relatively, 48
set, 48
closed ball, 47
closure, 49
cluster point, 50
coefficients, 177
compact, 98
Comparison Test, 152
Comparison Theorem for Improper Integrals, 143
complement, 10
complete, 89
Completeness Axiom of $\mathbb{R}$, 30
completeness of $\mathbb{R}^n$, 91
composition, 12
Composition of Limits, 76
connected, 106
constant function, 12
Continuity Characterization of Differentiability, 115
Continuity Characterization of Limits, 85
continuous, 82
limit characterization of, 84
open sets characterization of, 86
sequential characterization of, 82
continuous from the left, 111
continuous from the right, 111
converges, 54, 142
absolutely, 175
pointwise, 162
series, 149
uniformly, 163
converges absolutely, 152
converges conditionally, 155
convex, 109
coordinate, 13
cos, 185
countability of $\mathbb{Q}$, 38
countable, 37
cover, 98
covers, 98
Critical Point Lemma, 116
Darboux’s Intermediate Value Theorem, 122
Darboux’s Theorem, 128
De Morgan’s Laws, 17
decimal expansions, 33
decreasing, 24
degenerate, 29
dense, 52
Density of Irrationals, 34
Density of Rationals, 33
derivative, 113
ter higher order, 123
difference, 15
differentiable, 113
n-times, 123
at a point, 113
Dirichlet function, 131
disconnected, 106
discontinuity, 82
discontinuous, 82
discrete metric, 44
disjoint, 15
Distributive Laws, 16
diverges, 54, 142
domain, 9
e, 148
elements, 5
empty set, 8
equicontinuous, 170
Euclidean norm, 46
Euler’s constant, 153
exponential function, 147
extension, 10
Extreme Value Theorem, 103
family of sets, 5
finite, 37
finite intersection property, 98
Floor Theorem, 32
function, 5
arithmetic with, 83
image, 10
pre-image, 10
Fundamental Theorem of Calculus, 138
geometric series, 150
harmonic series, 150
Heine-Borel Theorem, 100
homogeneity, 45
identity function, 12
image, 10, 37
improper integral, 142
improperly integrable, 142
increasing, 24
index set, 6
indexed family of sets, 5
inf, 29
infimum, 29
infinite, 37
limit, 65, 78
integers, 23
integrable
improperly, 142
Integral Test, 154
integrand, 126
Integration by Parts, 140
interior, 52
interior point, 52
Intermediate Value Theorem, 109
intersection, 13
intersection, 13
intersects, 15
interval, 24
bounded, 24
closed, 24
half-closed, 24
half-open, 24
open, 24
unbounded, 24
interval of convergence, 177
open, 177
inverse, 12
Inverse Function Theorem, 119
irrational numbers, 23
isolated, 85
L’Hôpital’s Rule, 120
left-hand limit, 77
lim inf, 68, 72
lim sup, 68, 71
limit, 54, 73, 79
at infinity, 79
continuity characterization of, 85
infinite, 65, 78
sequential characterization of, 73
Limit Characterization of Continuity, 82
Limit Comparison Test, 153
Lindelöf’s Theorem, 95
linear approximation, 115
logarithm function, 146
lower bound, 29
lower limit of integration, 126
lower sum, 126
MacLaurin series, 180
max, 28
maximum, 28, 103
Mean Value Theorem, 117
Cauchy, 117
Mean Value Theorem for Integrals, 135
metric, 44
metric space, 44
min, 28
minimum, 28, 103
monotone, 24
Monotone Limits, 78
N, 22
n-times differentiable, 123
natural numbers, 22
nondegenerate, 25
norm, 45
normed space, 45
Occam’s Razor, 26
one-to-one, 12
onto, 12
open, 98
relatively, 48
set, 47
open ball, 46
Open Sets Characterization of Continuity, 80
ordered pair, 15
pairwise disjoint, 15
partial sum, 149
partition, 125
point, 44
pointwise bounded, 170
positivity, 44, 45
Power Rule, 114
rational, 119
power series, 177
powers
arbitrary, 148
integer, 23
rational, 35
pre-image, 10
Principle of Mathematical Induction, 23
Product Rule, 113
proper subset, 9
Q, 23
Quotient Rule, 114
\( \mathbb{R} \), 22
radius, 46
radius of convergence, 177
range, 5
Ratio Test, 158
rational numbers, 23
real sequence, 26
real-valued, 57
Rearrangement Theorem, 157
Riemann, 158
refines, 125
relative closure, 50
relatively closed, 48
relatively open, 48
remainder term, 181
restriction, 10
Riemann integrable, 126
Riemann integral, 126
Riemann sum, 127
Riemann’s Criterion, 128
Riemann’s Rearrangement Theorem, 158
right-hand limit, 77
Rolle’s Theorem, 117
root, 34
Root Test, 159
separability of \( \mathbb{R}^n \), 94
separable, 94
sequence, 25
Sequential Characterization
closed, 58
closure, 58
cluster point, 57
of Continuity, 82
of Limits, 73
sequentially compact, 97
series, 149, 175
\( p \), 154
Taylor, 180
set, 5
closed, 48
De Morgan’s Laws, 17
Distributive Laws, 16
open, 47
proper subset, 9
sin, 185
Squeeze Theorem
Infinite Limits, 67
limits, 75
sequences, 55
step function, 137
strict, 24
strictly decreasing, 24
strictly increasing, 24
strictly monotone, 24
subcover, 98
subsequence, 58
subsequential limit, 60
subset, 9
subspace, 45
sum, 149, 175
sup, 29
superset, 9
supremum, 29
symmetry, 44
tail, 56
Taylor polynomial, 181
Taylor series, 180
Taylor’s Theorem, 181
telescoping, 150
term, 25, 149
triangle inequality, 27, 44, 45
alternate, 27, 44

table
uncountability of \( \mathbb{R} \), 38
uncountable, 37
uniform Cauchy criterion, 175
uniformly continuous, 103
union, 13, 15
unit interval, 45
upper bound, 29
upper limit of integration, 126
upper sum, 125
value, 5
weak, 24
Weierstrass M-Test, 175
Weierstrass Approximation Theorem, 172
Well-Ordering Principle, 32
\( Z \), 23