1.1: Basic properties

Introduction

About this book

This book is a one semester course in basic analysis. It started its life as my lecture notes for teaching Math 444 at the University of Illinois at Urbana-Champaign (UIUC) in Fall semester 2009. Later I added the metric space chapter to teach Math 521 at University of Wisconsin–Madison (UW). A prerequisite for this course is a basic proof course, using for example , , or .

It should be possible to use the book for both a basic course for students who do not necessarily wish to go to graduate school (such as UIUC 444), but also as a more advanced one-semester course that also covers topics such as metric spaces (such as UW 521). Here are my suggestions for what to cover in a semester course. For a slower course such as UIUC 444:

§0.3, §1.1–§1.4, §2.1–§2.5, §3.1–§3.4, §4.1–§4.2, §5.1–§5.3, §6.1–§6.3

For a more rigorous course covering metric spaces that runs quite a bit faster (such as UW 521):

§0.3, §1.1–§1.4, §2.1–§2.5, §3.1–§3.4, §4.1–§4.2, §5.1–§5.3, §6.1–§6.2, §7.1–§7.6

It should also be possible to run a faster course without metric spaces covering all sections of chapters 0 through 6. The approximate number of lectures given in the section notes through chapter 6 are a very rough estimate and were designed for
the slower course. The first few chapters of the book can be used in an introductory proofs course as is for example done at Iowa State University Math 201, where this book is used in conjunction with Hammack’s Book of Proof.

The book normally used for the class at UIUC is Bartle and Sherbert, *Introduction to Real Analysis* third edition. The structure of the beginning of the book somewhat follows the standard syllabus of UIUC Math 444 and therefore has some similarities with. A major difference is that we define the Riemann integral using Darboux sums and not tagged partitions. The Darboux approach is far more appropriate for a course of this level.

Our approach allows us to fit a course such as UIUC 444 within a semester and still spend some extra time on the interchange of limits and end with Picard’s theorem on the existence and uniqueness of solutions of ordinary differential equations. This theorem is a wonderful example that uses many results proved in the book. For more advanced students, material may be covered faster so that we arrive at metric spaces and prove Picard’s theorem using the fixed point theorem as is usual.

Other excellent books exist. My favorite is Rudin’s excellent *Principles of Mathematical Analysis* or as it is commonly and lovingly called *baby Rudin* (to distinguish it from his other great analysis textbook). I took a lot of inspiration and ideas from Rudin. However, Rudin is a bit more advanced and ambitious than this present course. For those that wish to continue mathematics, Rudin is a fine investment. An inexpensive and somewhat simpler alternative to Rudin is Rosenlicht’s *Introduction to Analysis*. There is also the freely downloadable *Introduction to Real Analysis* by William Trench.

A note about the style of some of the proofs: Many proofs traditionally done by contradiction, I prefer to do by a direct proof or by contrapositive. While the book does include proofs by contradiction, I only do so when the contrapositive statement seemed too awkward, or when contradiction follows rather quickly. In my opinion, contradiction is more likely to get beginning students into trouble, as we are talking about objects that do not exist.

I try to avoid unnecessary formalism where it is unhelpful. Furthermore, the proofs and the language get slightly less formal as we progress through the book, as more and more details are left out to avoid clutter.

As a general rule, I use \(\equiv\) instead of \(=\) to define an object rather than to simply show equality. I use this symbol rather more liberally than is usual for emphasis. I use it even when the context is “local,” that is, I may simply define a function \(f(x) := x^2\) for a single exercise or example.

Finally, I would like to acknowledge Jana Maříková, Glen Pugh, Paul Vojta, Frank Beatrous, Sönmez Şahutoğlu, Jim Brandt, Kenji Kozai, and Arthur Busch, for teaching with the book and giving me lots of useful feedback. Frank Beatrous wrote the University of Pittsburgh version extensions, which served as inspiration for many of the recent additions. I would also like to thank Dan Stoneham, Jeremy Sutter, Eliya Gwetta, Daniel Pimentel-Alarcón, Steve Hoerning, Yi Zhang, Nicole Caviris, Kristopher Lee, Baoyue Bi, Hannah Lund, Trevor Mannella, Mitchel Meyer, Gregory Beauregard, Chase Meadors, Andreas Giannopoulos, an anonymous reader, and in general all the students in my classes for suggestions and finding errors and typos.
About analysis

Analysis is the branch of mathematics that deals with inequalities and limits. The present course deals with the most basic concepts in analysis. The goal of the course is to acquaint the reader with rigorous proofs in analysis and also to set a firm foundation for calculus of one variable.

Calculus has prepared you, the student, for using mathematics without telling you why what you learned is true. To use, or teach, mathematics effectively, you cannot simply know what is true, you must know why it is true. This course shows you why calculus is true. It is here to give you a good understanding of the concept of a limit, the derivative, and the integral.

Let us use an analogy. An auto mechanic that has learned to change the oil, fix broken headlights, and charge the battery, will only be able to do those simple tasks. He will be unable to work independently to diagnose and fix problems. A high school teacher that does not understand the definition of the Riemann integral or the derivative may not be able to properly answer all the students’ questions. To this day I remember several nonsensical statements I heard from my calculus teacher in high school, who simply did not understand the concept of the limit, though he could “do” all problems in calculus.

We start with a discussion of the real number system, most importantly its completeness property, which is the basis for all that comes after. We then discuss the simplest form of a limit, the limit of a sequence. Afterwards, we study functions of one variable, continuity, and the derivative. Next, we define the Riemann integral and prove the fundamental theorem of calculus. We discuss sequences of functions and the interchange of limits. Finally, we give an introduction to metric spaces.

Let us give the most important difference between analysis and algebra. In algebra, we prove equalities directly; we prove that an object, a number perhaps, is equal to another object. In analysis, we usually prove inequalities. To illustrate the point, consider the following statement.

Let \( x \) be a real number. If \( 0 \leq x < \epsilon \) for all real numbers \( \epsilon > 0 \), then \( x = 0 \).

This statement is the general idea of what we do in analysis. If we wish to show that \( x = 0 \), we show that \( 0 \leq x < \epsilon \) for all positive \( \epsilon \).

The term real analysis is a little bit of a misnomer. I prefer to use simply analysis. The other type of analysis, complex analysis, really builds up on the present material, rather than being distinct. Furthermore, a more advanced course on real analysis would talk about complex numbers often. I suspect the nomenclature is historical baggage.

Let us get on with the show…

Basic set theory

Note: 1–3 lectures (some material can be skipped or covered lightly)
Before we start talking about analysis we need to fix some language. Modern¹ analysis uses the language of sets, and therefore that is where we start. We talk about sets in a rather informal way, using the so-called “naïve set theory.” Do not worry, that is what majority of mathematicians use, and it is hard to get into trouble.

We assume the reader has seen basic set theory and has had a course in basic proof writing. This section should be thought of as a refresher.

**Sets**

A set is a collection of objects called elements or members. A set with no objects is called the empty set and is denoted by \(\emptyset\) (or sometimes by \(\{\}\)).

Think of a set as a club with a certain membership. For example, the students who play chess are members of the chess club. However, do not take the analogy too far. A set is only defined by the members that form the set; two sets that have the same members are the same set.

Most of the time we will consider sets of numbers. For example, the set \(S := \{0, 1, 2\}\) is the set containing the three elements 0, 1, and 2. We write \(1 \in S\) to denote that the number 1 belongs to the set \(S\). That is, 1 is a member of \(S\).

Similarly we write \(7 \notin S\) to denote that the number 7 is not in \(S\). That is, 7 is not a member of \(S\). The elements of all sets under consideration come from some set we call the universe. For simplicity, we often consider the universe to be the set that contains only the elements we are interested in. The universe is generally understood from context and is not explicitly mentioned. In this course, our universe will most often be the set of real numbers.

While the elements of a set are often numbers, other objects, such as other sets, can be elements of a set. A set may also contain some of the same elements as another set. For example, \(T := \{0, 2\}\) contains the numbers 0 and 2. In this case all elements of \(T\) also belong to \(S\). We write \(T \subset S\). More formally we make the following definition.

i. A set \(A\) is a subset of a set \(B\) if \(x \in A\) implies \(x \in B\), and we write \(A \subset B\). That is, all members of \(A\) are also members of \(B\).

ii. Two sets \(A\) and \(B\) are equal if \(A \subset B\) and \(B \subset A\). We write \(A = B\). That is, \(A\) and \(B\) contain exactly the same elements. If it is not true that \(A\) and \(B\) are equal, then we write \(A \not= B\).

iii. A set \(A\) is a proper subset of \(B\) if \(A \subset B\) and \(A \not= B\). We write \(A \subsetneq B\).

For example, for \(S\) and \(T\) defined above \(T \subset S\), but \(T \not= S\). So \(T\) is a proper subset of \(S\). If \(A = B\), then \(A\) and \(B\) are simply two names for the same exact set. Let us mention the set building notation, \(\{x \in A : P(x)\}\). This notation refers to a subset of the set \(A\) containing all elements of \(A\) that satisfy the property \(P(x)\). The notation is sometimes abbreviated, \(\{A\}\) is not mentioned when understood from context. Furthermore, \(\{x \in A\}\) is sometimes replaced with a formula to make the notation easier to read.

The following are sets including the standard notations.

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
i. The set of natural numbers, \(\{\mathbb{N}\} := \{1, 2, 3, \ldots\}\).

ii. The set of integers, \(\{\mathbb{Z}\} := \{0, -1, 1, -2, 2, \ldots\}\).

iii. The set of rational numbers, \(\{\mathbb{Q}\} := \{\frac{m}{n} : m, n \in \mathbb{Z} \text{ and } n \not= 0\}\).

iv. The set of even natural numbers, \(\{2m : m \in \mathbb{N}\}\).

v. The set of real numbers, \(\{\mathbb{R}\}\).

Note that \(\{\mathbb{N}\} \subset \{\mathbb{Z}\} \subset \{\mathbb{Q}\} \subset \{\mathbb{R}\}\).

There are many operations we want to do with sets.

i. A union of two sets \(A\) and \(B\) is defined as \(\{A \cup B := \{x : x \in A \text{ or } x \in B\}\}\).

ii. An intersection of two sets \(A\) and \(B\) is defined as \(\{A \cap B := \{x : x \in A \text{ and } x \in B\}\}\).

iii. A complement of \(B\) relative to \(A\) (or set-theoretic difference of \(A\) and \(B\)) is defined as \(\{A \setminus B := \{x : x \in A \text{ and } x \notin B\}\}\).

iv. We say complement of \(B\) and write \(B^c\) instead of \(A \setminus B\) if the set \(A\) is either the entire universe or is the obvious set containing \(B\), and is understood from context.

v. We say sets \(A\) and \(B\) are disjoint if \(\{A \cap B = \emptyset\}\).

The notation \(B^c\) may be a little vague at this point. If the set \(\{B\}\) is a subset of the real numbers \(\{\mathbb{R}\}\), then \(B^c\) means \(\{\mathbb{R}\setminus B\}\). If \(\{B\}\) is naturally a subset of the natural numbers, then \(B^c\) is \(\{\mathbb{N}\setminus B\}\). If ambiguity would ever arise, we will use the set difference notation \(A \setminus B\).

We illustrate the operations on the Venn diagrams in . Let us now establish one of most basic theorems about sets and logic.

Let \(\{A, B, C\}\) be sets. Then \(\{\begin{align*} &\{B \cup C\}^c = B^c \cap C^c, \\
&\{B \cap C\}^c = B^c \cup C^c\end{align*}\}\) or, more generally, \(\{\begin{align*} &A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C), \\
&A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)\end{align*}\}\) The first statement is proved by the second statement if we assume the set \(\{A\}\) is our “universe.”

Let us prove \(\{A \setminus \text{setminus } (B \cup C) = (A \setminus \text{setminus } B) \setminus \text{cap } (A \setminus \text{setminus } C)\}. Remember the definition of equality of sets. First, we must show that if \(\{x \in A \setminus \text{setminus } (B \cup C)\},\) then \(\{x \in (A \setminus \text{setminus } B) \setminus \text{cap } (A \setminus \text{setminus } C)\}. Second, we must also show that if \(\{x \in (A \setminus \text{setminus } B) \setminus \text{cap } (A \setminus \text{setminus } C)\},\) then \(\{x \in A \setminus \text{setminus } (B \cup C)\}.

So let us assume \(\{x \in A \setminus \text{setminus } (B \cup C)\}. Then \(\{x\}\) is in \(\{A\}\), but not in \(\{B\}\) nor \(\{C\}\). Hence \(\{x\}\) is in \(\{A\}\) and not in \(\{B\}\), that is, \(\{x \in A \setminus \text{setminus } B\}\). Similarly \(\{x \in A \setminus \text{setminus } C\}. Thus \(\{x \in (A \setminus \text{setminus } B) \setminus \text{cap } (A \setminus \text{setminus } C)\}.

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
On the other hand suppose \((x \in (A \setminus B) \cap (A \setminus C))\). In particular \((x \in (A \setminus B))\), so \((x \in A)\) and \((x \notin B)\). Also as \((x \in (A \setminus C))\), then \((x \notin C)\). Hence \((x \in A \setminus (B \cup C))\).

The proof of the other equality is left as an exercise.

We will also need to intersect or union several sets at once. If there are only finitely many, then we simply apply the union or intersection operation several times. However, suppose we have an infinite collection of sets (a set of sets) \(\{ A_0, A_1, A_2, A_3, \ldots \}\). We define \[
\bigcup_{n=1}^{\infty} A_n := \{ x : x \in A_n \text{ for some } n \in \mathbb{N} \}, \\
\bigcap_{n=1}^{\infty} A_n := \{ x : x \in A_n \text{ for all } n \in \mathbb{N} \}.
\]

We can also have sets indexed by two integers. For example, we can have the set of sets \(\{ A_{1,1}, A_{1,2}, A_{2,1}, A_{2,2}, A_{1,3}, A_{3,1}, \ldots \}\). Then we write \[
\bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} A_{n,m} = \bigcup_{n=1}^{\infty} \left( \bigcup_{m=1}^{\infty} A_{n,m} \right) .
\]

It is not hard to see that we can take the unions in any order. However, switching the order of unions and intersections is not generally permitted without proof. For example, \[
\bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \{ k \in \mathbb{N} : mk < n \} = \bigcup_{n=1}^{\infty} \emptyset = \emptyset .
\]

Sometimes, the index set is not the natural numbers. In this case we need a more general notation. Suppose \(\{ I \} \) is some set and for each \(\iota \in I\), we have a set \((A_{\iota})\). Then we define \[
\bigcup_{\iota \in I} A_{\iota} := \{ x : x \in A_{\iota} \text{ for some } \iota \in I \}, \\
\bigcap_{\iota \in I} A_{\iota} := \{ x : x \in A_{\iota} \text{ for all } \iota \in I \}.
\]

### Induction

When a statement includes an arbitrary natural number, a common method of proof is the principle of induction. We start with the set of natural numbers \(\{ \mathbb{N} \} = \{ 1, 2, 3, \ldots \}\), and we give them their natural ordering, that is, \((1 < 2 < 3 < 4 < \ldots)\). By \((S \subseteq \mathbb{N})\), we mean that there exists an \((x \in S)\), such that for every \((y \in S)\), we have \((x \leq y)\).

The natural numbers \(\{ \mathbb{N} \}\) ordered in the natural way possess the so-called well ordering property. We take this property as an axiom; we simply assume it is true.

Every nonempty subset of \(\{ \mathbb{N} \}\) has a least (smallest) element.

The principle of induction is the following theorem, which is equivalent to the well ordering property of the natural numbers.

**[induction:thm] Let \((P(n))\) be a statement depending on a natural number \((n)\). Suppose that**

i. **(basis statement)** \((P(1))\) is true,

ii. **(induction step)** if \((P(n))\) is true, then \((P(n+1))\) is true.
Then \((P(n))\) is true for all \(n \in \{ \text{mathbb}\{N\}\}\).

Suppose \(\neg(S)\) is the set of natural numbers \(m\) for which \(P(m)\) is not true. Suppose \(\neg(S)\) is nonempty. Then \(\neg(S)\) has a least element by the well ordering property. Let us call \(m\) the least element of \(\neg(S)\). We know \(\neg(1 \ \text{not in } S)\) by assumption. Therefore \(m > 1\) and \(m - 1\) is a natural number as well. Since \(m\) was the least element of \(\neg(S)\), we know that \(P(m-1)\) is true. But by the induction step we see that \(P(m-1+1) = P(m)\) is true, contradicting the statement that \(m \in S\). Therefore \(\neg(S)\) is empty and \(P(n)\) is true for all \(n \in \{ \text{mathbb}\{N\}\}\).

Sometimes it is convenient to start at a different number than 1, but all that changes is the labeling. The assumption that \(\neg(P(n))\) is true in “if \(P(n)\) is true, then \(P(n+1)\) is true” is usually called the *induction hypothesis*.

Let us prove that for all \(n \in \{ \text{mathbb}\{N\}\}\), \(\lfloor 2^{\lfloor n-1 \rfloor} \leq n! \rfloor\). We let \(P(n)\) be the statement that \(2^{\lfloor n-1 \rfloor} \leq n!\) is true. By plugging in \(n=1\), we see that \(P(1)\) is true.

Suppose \(P(n)\) is true. That is, suppose \(2^{\lfloor n-1 \rfloor} \leq n!\) holds. Multiply both sides by 2 to obtain \(2^{\lfloor n \rfloor} \leq 2(n!\rfloor\). As \(2\lfloor n+1 \rfloor\) when \(n \in \{ \text{mathbb}\{N\}\}\), we have \(2(n!) \leq (n+1)!\). That is, \(2^{\lfloor n \rfloor} \leq 2(n!) \leq (n+1)!\) and hence \(P(n+1)\) is true. By the principle of induction, we see that \(P(n)\) is true for all \(n\), and hence \(2^{\lfloor n-1 \rfloor} \leq n!\) is true for all \(n \in \{ \text{mathbb}\{N\}\}\).

We claim that for all \(c \neq 1\), \(1 + c + c^2 + \cdots + c^n = \frac{1-c^{n+1}}{1-c}\). Let \(P(n)\) be the statement that \(1 + c + c^2 + \cdots + c^n = \frac{1-c^{n+1}}{1-c}\) is true.

Proof: It is easy to check that the equation holds with \(n=1\). Suppose it is true for \(P(n)\). Then \(\lfloor 1 + c + c^2 + \cdots + c^n + c^{n+1} \rfloor = \lfloor (1 + c + c^2 + \cdots + c^n) + c^{n+1} \rfloor = \frac{1-c^{n+1}}{1-c} + c^{n+1} = \frac{1-c^{n+2}}{1-c}\).

There is an equivalent principle called strong induction. The proof that strong induction is equivalent to induction is left as an exercise.

Let \(P(n)\) be a statement depending on a natural number \(n\). Suppose that

1. *(basis statement)* \(P(1)\) is true,
2. *(induction step)* if \(P(k)\) is true for all \(k = 1, 2, \ldots, n\), then \(P(n+1)\) is true.

Then \(P(n)\) is true for all \(n \in \{ \text{mathbb}\{N\}\}\).

**Functions**

Informally, a *set-theoretic function* \(\varphi\) taking a set \(\{A\}\) to a set \(\{B\}\) is a mapping that to each \(x \in \{A\}\) assigns a unique \(y \in \{B\}\). We write \(\varphi: \{A\} \to \{B\}\). For example, we define a function \(\varphi: \{1, 2\} \to \{T, S\}\) taking \(\{1, 2\}\) to \(\{T = 1, S = 2\}\) by assigning \(\varphi(0) := 2\) and \(\varphi(1) := 1\). That is, a function \(\varphi: \{A\} \to \{B\}\) is a black box, into which we stick an element of \(\{A\}\) and the function spits out an element of \(\{B\}\). Sometimes \(\varphi\) is called a *mapping* and we say \(\varphi\) maps \(\{A\}\) to \(\{B\}\).
Often, functions are defined by some sort of formula, however, you should really think of a function as just a very big table of values. The subtle issue here is that a single function can have several different formulas, all giving the same function. Also, for many functions, there is no formula that expresses its values.

To define a function rigorously first let us define the Cartesian product.

Let \( \langle A \rangle \) and \( \langle B \rangle \) be sets. The **Cartesian product** is the set of tuples defined as

\[
\langle A \rangle \times \langle B \rangle := \{ (x,y) : x \in A, y \in B \} .
\]

For example, the set \( \langle [0,1] \rangle \times \langle [0,1] \rangle \) is a set in the plane bounded by a square with vertices \( \langle (0,0) \rangle, \langle (0,1) \rangle, \langle (1,0) \rangle, \) and \( \langle (1,1) \rangle \). When \( \langle A \rangle \rangle \) and \( \langle B \rangle \rangle \) are the same set we sometimes use a superscript 2 to denote such a product. For example \( \langle [0,1] \rangle^2 = \langle [0,1] \rangle \times \langle [0,1] \rangle \rangle \), or \( \langle \mathbb{R} \rangle^2 = \langle \mathbb{R} \rangle \times \langle \mathbb{R} \rangle \rangle \) (the Cartesian plane).

A function \( \langle f \rangle \colon \langle A \rangle \to \langle B \rangle \rangle \) is a subset \( \langle f \rangle \rangle \) of \( \langle A \rangle \times \langle B \rangle \rangle \) such that for each \( \langle x \in A \rangle \rangle \), there is a unique \( \langle (x,y) \in f \rangle \rangle \). We then write \( \langle f(x) = y \rangle \rangle \). Sometimes the set \( \langle f \rangle \rangle \) is called the **graph** of the function rather than the function itself.

The set \( \langle A \rangle \rangle \) is called the **domain** of \( \langle f \rangle \rangle \) (and sometimes confusingly denoted \( \langle D(f) \rangle \rangle \)). The set \( \langle R(f) := \{ y \in B : \text{there exists an } x \in A \text{ such that } (x,y) \in f \} \rangle \rangle \) is called the **range** of \( \langle f \rangle \rangle \). Note that \( \langle R(f) \rangle \rangle \) can possibly be a proper subset of \( \langle B \rangle \rangle \), while the domain of \( \langle f \rangle \rangle \) is always equal to \( \langle A \rangle \rangle \). We usually assume that the domain of \( \langle f \rangle \rangle \) is nonempty.

From calculus, you are most familiar with functions taking real numbers to real numbers. However, you saw some other types of functions as well. For example, the derivative is a function mapping the set of differentiable functions to the set of all functions. Another example is the Laplace transform, which also takes functions to functions. Yet another example is the function that takes a continuous function \( \langle g \rangle \rangle \) defined on the interval \( \langle [0,1] \rangle \rangle \) and returns the number \( \langle \int_0^1 g(x) \, dx \rangle \rangle \).

Let \( \langle f \rangle \colon \langle A \rangle \to \langle B \rangle \) be a function, and \( \langle C \subseteq A \rangle \rangle \). Define the **image** (or **direct image**) of \( \langle C \rangle \rangle \) as \( \langle f(C) := \{ f(x) : x \in C \} \rangle \rangle \). Let \( \langle D \subseteq B \rangle \rangle \). Define the **inverse image** as \( \langle f^{-1}(D) := \{ x : f(x) \in D \} \rangle \rangle \).

Define the function \( \langle f \rangle \colon \langle \mathbb{R} \rangle \to \langle \mathbb{R} \rangle \rangle \) by \( \langle f(x) := \sin(\pi x) \rangle \rangle \). Then \( \langle f([0,1]) \rangle = \{ 0,1 \} \rangle \), \( \langle f^{-1}([0,1]) \rangle \rangle \), etc.

[st:propinv] Let \( \langle f \rangle \colon \langle A \rangle \to \langle B \rangle \rangle \). Let \( \langle C, D \rangle \rangle \) be subsets of \( \langle B \rangle \rangle \). Then \( \langle \begin{aligned} \langle f \rangle^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D) , \\ \langle f \rangle^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D) , \\ \langle f \rangle^{-1}(C^c) = (f^{-1}(C))^c \end{aligned} \rangle \rangle \). Read the last line as \( \langle f \rangle^{-1}(B \setminus C) = A \setminus f^{-1}(C) \rangle \rangle \).

Let us start with the union. Suppose \( \langle x \in f^{-1}(C \cup D) \rangle \rangle \). That means \( \langle x \rangle \) maps to \( \langle C \rangle \rangle \) or \( \langle D \rangle \rangle \). Thus \( \langle f^{-1}(C \cup D) \rangle \subseteq f^{-1}(C) \cup f^{-1}(D) \rangle \rangle \). Conversely if \( \langle x \in f^{-1}(C) \rangle \rangle \), then \( \langle x \in f^{-1}(C \cup D) \rangle \rangle \). Similarly for \( \langle x \in f^{-1}(D) \rangle \rangle \). Hence \( \langle f \rangle^{-1}(C \cup D) \subseteq f^{-1}(C) \cup f^{-1}(D) \rangle \rangle \) and we have equality.

The rest of the proof is left as an exercise.
The proposition does not hold for direct images. We do have the following weaker result.

\[
\text{Let } f : A \to B. \text{ Let } C, D \text{ be subsets of } \{A\}. \text{ Then } \begin{aligned}
& f(C \cup D) = f(C) \cup f(D), \\
& f(C \cap D) \subset f(C) \cap f(D). \end{aligned}
\]

The proof is left as an exercise.

A function \(f\) is said to be **injective** or one-to-one if \((f(x_1) = f(x_2)) \implies (x_1 = x_2)\). In other words, for all \(y \in B\) the set \((f^{-1}\{y\})\) is empty or consists of a single element. We call such an \(f\) an injection.

The function \(f\) is said to be **surjective** or onto if \((f(A) = B)\). We call such an \(f\) a surjection.

A function \(f\) that is both an injection and a surjection is said to be **bijective**, and we say \(f\) is a bijection.

When \((f : A \to B)\) is a bijection, then \((f^{-1}\{y\})\) is always a unique element of \(\{A\}\), and we can consider \((f^{-1}\{y\})\) as a function \((f^{-1} : B \to A)\). In this case, we call \((f^{-1}\{y\})\) the inverse function of \(f\). For example, for the bijection \((f : \mathbb{R} \to \mathbb{R})\) defined by \(f(x) := x^3\) we have \((f^{-1}\{y\})(x) = \sqrt[3]{y}\).

A final piece of notation for functions that we need is the composition of functions.

Let \((f : A \to B), (g : B \to C)\). The function \((g \circ f : A \to C)\) is defined as \((g \circ f)(x) := g(f(x))\).

**Cardinality**

A subtle issue in set theory and one generating a considerable amount of confusion among students is that of cardinality, or “size” of sets. The concept of cardinality is important in modern mathematics in general and in analysis in particular. In this section, we will see the first really unexpected theorem.

Let \(\{A\}\) and \(\{B\}\) be sets. We say \(\{A\}\) and \(\{B\}\) have the same cardinality when there exists a bijection \((f : A \to B)\). We denote by \(|\{A\}|\) the equivalence class of all sets with the same cardinality as \(\{A\}\) and we simply call \(|\{A\}|\) the cardinality of \(\{A\}\).

Note that \(|\{A\}|\) has the same cardinality as the empty set if and only if \(|\{A\}|\) itself is the empty set. We then write \(|\{A\}| := 0\).

Suppose \(|\{A\}|\) has the same cardinality as \(|\{1,2,3,\ldots,n\}|\) for some \(|\{n\}|\). We then write \(|\{A\}| := n\), and we say \(|\{A\}|\) is finite. When \(|\{A\}|\) is the empty set, we also call \(|\{A\}|\) finite.

We say \(|\{A\}|\) is **infinite** or “of infinite cardinality” if \(|\{A\}|\) is not finite.

That the notation \(|\{A\}| := n\) is justified we leave as an exercise. That is, for each nonempty finite set \(|\{A\}|\), there exists a unique natural number \(|n|\) such that there exists a bijection from \(|\{A\}|\) to \(|\{1,2,3,\ldots,n\}|\).
We can order sets by size.

[def:comparecards] We write \( |A| \leq |B| \) if there exists an injection from \( A \) to \( B \). We write \( |A| = |B| \) if \( A \) and \( B \) have the same cardinality. We write \( |A| < |B| \) if \( |A| \leq |B| \), but \( A \) and \( B \) do not have the same cardinality.

We state without proof that \( |A| = |B| \) have the same cardinality if and only if \( |A| \leq |B| \) and \( |B| \leq |A| \). This is the so-called Cantor-Bernstein-Schroeder theorem. Furthermore, if \( |A| \) and \( |B| \) are any two sets, we can always write \( |A| \leq |B| \) or \( |B| \leq |A| \). The issues surrounding this last statement are very subtle. As we do not require either of these two statements, we omit proofs.

The truly interesting cases of cardinality are infinite sets. We start with the following definition.

If \( |A| = |B| \) \( \Rightarrow \) \( A \) is said to be **countably infinite**. If \( |A| \) is finite or countably infinite, then we say \( |A| \) is **countable**. If \( |A| \) is not countable, then \( |A| \) is said to be **uncountable**.

The cardinality of \( \mathbb{N} \) is usually denoted as \( \aleph_0 \) (read as aleph-naught) \(^2\).

The set of even natural numbers has the same cardinality as \( \{ \mathbb{N} \} \). Proof: Given an even natural number, write it as \( 2n \) for some \( n \in \mathbb{N} \). Then create a bijection taking \( 2n \) to \( n \).

In fact, let us mention without proof the following characterization of infinite sets: A set is infinite if and only if it is in one-to-one correspondence with a proper subset of itself.

\( \{ \mathbb{N} \} \times \{ \mathbb{N} \} \) is a countably infinite set. Proof: Arrange the elements of \( \{ \mathbb{N} \} \times \{ \mathbb{N} \} \) as follows \( (1,1), (1,2), (2,1), (1,3), (2,2), (3,1), \ldots \). That is, always write down first all the elements whose two entries sum to \( k \), then write down all the elements whose entries sum to \( k+1 \) and so on. Then define a bijection with \( \{ \mathbb{N} \} \) by letting 1 go to \( (1,1) \), 2 go to \( (1,2) \) and so on.

The set of rational numbers is countable. Proof: (informal) Follow the same procedure as in the previous example, writing \( \{ \frac{1}{1} \}, \{ \frac{1}{2} \}, \{ \frac{2}{1} \}, \{ \frac{1}{3} \}, \{ \frac{2}{2} \}, \{ \frac{2}{3} \}, \ldots \). However, leave out any fraction (such as \( \{ \frac{2}{2} \} \)) that has already appeared.
For completeness we mention the following statement. If \((A \subset B)\) and \((B)\) is countable, then \((A)\) is countable. Similarly if \((A)\) is uncountable, then \((B)\) is uncountable. As we will not need this statement in the sequel, and as the proof requires the Cantor-Bernstein-Schroeder theorem mentioned above, we will not give it here.

We give the first truly striking result. First, we need a notation for the set of all subsets of a set.

If \(\{A\}\) is a set, we define the power set of \(\{A\}\), denoted by \(\{\text{mathcal\{P\}\} (A)\}\), to be the set of all subsets of \(\{A\}\).

For example, if \(\{A := \{ 1,2\}\}\), then \(\{\text{mathcal\{P\}\} (A) = \{ \emptyset, \{ 1 \}, \{ 2 \}, \{ 1, 2 \} \}\}\). For a finite set \(\{A\}\) of cardinality \(\langle n\rangle\), the cardinality of \(\{\text{mathcal\{P\}\} (A)\}\) is \(2^{\langle n\rangle}\). This fact is left as an exercise. Hence, for finite sets the cardinality of \(\{\text{mathcal\{P\}\} (A)\}\) is strictly larger than the cardinality of \(\langle A\rangle\). What is an unexpected and striking fact is that this statement is still true for infinite sets.

\[\left\lvert \{A\} \right\rvert < \left\lvert \{\text{mathcal\{P\}\} (A)\}\right\rvert\]

There exists an injection \(\{f\colon A \to \{\text{mathcal\{P\}\} (A)\}\}\). For any \(\langle x \in A\rangle\), define \(\langle f(x) := \{ x \}\rangle\). Therefore \(\left\lvert \{A\} \right\rvert \leq \left\lvert \{\text{mathcal\{P\}\} (A)\}\right\rvert\).

To finish the proof, we must show that no function \(\{f\colon A \to \{\text{mathcal\{P\}\} (A)\}\}\) is a surjection. Suppose \(\{f\colon A \to \{\text{mathcal\{P\}\} (A)\}\}\) is a function. So for \(\langle x \in A\rangle\), \(\langle f(x)\rangle\) is a subset of \(\langle A\rangle\). Define the set \(\{B := \{ x \in A : x \not\in f(x) \}\}\). We claim that \(\langle B\rangle\) is not in the range of \(\langle f\rangle\) and hence \(\langle f\rangle\) is not a surjection. Suppose there exists an \(\langle x_0 \rangle\) such that \(\langle f(x_0) = B\rangle\). Either \(\langle x_0 \in B\rangle\) or \(\langle x_0 \not\in B\rangle\). If \(\langle x_0 \in B\rangle\), then \(\langle x_0 \not\in f(x_0) = B\rangle\), which is a contradiction. If \(\langle x_0 \not\in B\rangle\), then \(\langle x_0 \in f(x_0) = B\rangle\), which is again a contradiction. Thus such an \(\langle x_0 \rangle\) does not exist. Therefore, \(\langle B\rangle\) is not in the range of \(\langle f\rangle\), and \(\langle f\rangle\) is not a surjection. As \(\langle f\rangle\) was an arbitrary function, no surjection exists.

One particular consequence of this theorem is that there do exist uncountable sets, as \(\{\text{mathcal\{P\}\} (\{\text{mathbb\{N\}\})\}\}\) must be uncountable. A related fact is that the set of real numbers (which we study in the next chapter) is uncountable. The existence of uncountable sets may seem unintuitive, and the theorem caused quite a controversy at the time it was announced. The theorem not only says that uncountable sets exist, but that there in fact exist progressively larger and larger infinite sets \(\{\text{mathbb\{N\}\}}\), \(\{\text{mathcal\{P\}\} (\{\text{mathbb\{N\}\})))\}, \{\text{mathcal\{P\}\} (\{\text{mathcal\{P\}\} (\{\text{mathbb\{N\}\})\})))\}, \{\text{mathcal\{P\}\} (\{\text{mathcal\{P\}\} (\{\text{mathcal\{P\}\} (\{\text{mathbb\{N\}\})\}))\})\}, etc….

**Exercises**

Show \(\{A \setminus \text{setminus (B \cap C)} = (A \setminus \text{setminus B}) \cup (A \setminus \text{setminus C})\}\).

Prove that the principle of strong induction is equivalent to the standard induction.
Finish the proof of .

a. Prove .

b. Find an example for which equality of sets in \( f( C \cap D) \subset f( C) \cap f( D) \) fails. That is, find an \( f(\cdot) \), \( (A\cdot) \), \( (B\cdot) \), \( (C\cdot) \), and \( (D\cdot) \) such that \( f( C \cap D) \) is a proper subset of \( f( C) \cap f( D) \).

Prove that if \( (A\cdot) \) is finite, then there exists a unique number \( (n\cdot) \) such that there exists a bijection between \( (A\cdot) \) and \( (\{ 1, 2, 3, \ldots, n \}) \). In other words, the notation \( (\left\lfloor \text{left}\{ A \text{ \right\rceil} \vdash \text{right} := n \}) \) is justified. Hint: Show that if \( (n > m\cdot) \), then there is no injection from \( (\{ 1, 2, 3, \ldots, n \}) \) to \( (\{ 1, 2, 3, \ldots, m \}) \).

Prove

a. \( (A \cap (B \cup C) = (A \cap B) \cup (A \cap C)) \)

b. \( (A \cup (B \cap C) = (A \cup B) \cap (A \cup C)) \)

Let \( (A \Delta B) \) denote the symmetric difference, that is, the set of all elements that belong to either \( (A\cdot) \) or \( (B\cdot) \), but not to both \( (A\cdot) \) and \( (B\cdot) \).

a. Draw a Venn diagram for \( (A \Delta B) \).

b. Show \( (A \Delta B = (A \setminus \text{setminus } B) \cup (B \setminus \text{setminus } A)) \).

c. Show \( (A \Delta B = (A \cup B) \setminus \text{setminus } (A \cap B)) \).

For each \( (n \in \mathbb{N}) \), let \( (A_n := \{ (n+1)k : k \in \mathbb{N} \}) \).

a. Find \( (A_1 \setminus \text{cap } A_2) \).

b. Find \( (\bigcup_{n=1}^{\infty} n=1} \A_n)\).

c. Find \( (\bigcap_{n=1}^{\infty} n=1} \A_n)\).

Determine \( (\{\mathcal{P}\{\}}(S)) \) (the power set) for each of the following:

a. \( (S = \emptyset) \),

b. \( (S = \{1\}) \),

c. \( (S = \{1,2\}) \),

d. \( (S = \{1,2,3,4\}) \).

Let \( (f \colon A \to B) \) and \( (g \colon B \to C) \) be functions.
a. Prove that if \((g \circ f)\) is injective, then \((f)\) is injective.

b. Prove that if \((g \circ f)\) is surjective, then \((g')\) is surjective.

c. Find an explicit example where \((g \circ f)\) is bijective, but neither \((f)\) nor \((g')\) are bijective.

Prove that \(n < 2^n\) by induction.

Show that for a finite set \(\mathcal{A}\) of cardinality \(n\), the cardinality of \(\mathcal{P}(\mathcal{A})\) is \(2^n\).

Prove \(\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n(n+1)} = \frac{n}{n+1}\) for all \(n \in \mathbb{N}\).

Prove \(1^3 + 2^3 + \cdots + n^3 = (\frac{n(n+1)}{2})^2\) for all \(n \in \mathbb{N}\).

Prove that \(n^3 + 5n\) is divisible by \(6\) for all \(n \in \mathbb{N}\).

Find the smallest \(n \in \mathbb{N}\) such that \(2\{n+5\}^2 < n^3\) and call it \(n_0\). Show that \(2\{n+5\}^2 < n^3\) for all \(n \geq n_0\).

Find all \(n \in \mathbb{N}\) such that \(n^2 < 2^n\).

Finish the proof that the is equivalent to the well ordering property of \(\mathbb{N}\). That is, prove the well ordering property for \(\mathbb{N}\) using the principle of induction.

Give an example of a countable collection of finite sets \(A_1, A_2, \ldots\), whose union is not a finite set.

Give an example of a countable collection of infinite sets \(A_1, A_2, \ldots\), with \(A_j \cap A_k\) being infinite for all \(j\) and \(k\), such that \(\bigcap_{j=1}^{\infty} A_j\) is nonempty and finite.

---

**Real Numbers**

### Basic properties

*Note: 1.5 lectures*

The main object we work with in analysis is the set of real numbers. As this set is so fundamental, often much time is spent on formally constructing the set of real numbers. However, we take an easier approach here and just assume that a set with the correct properties exists. We need to start with the definitions of those properties.

An ordered set is a set \(\mathcal{S}\), together with a relation \((<)\) such that

i. For any \((x, y \in \mathcal{S})\), exactly one of \((x < y)\), \((x=y)\), or \((y < x)\) holds.
ii. If \((x < y)\) and \((y < z)\), then \((x < z)\).

We write \((x \leq y)\) if \((x < y)\) or \((x = y)\). We define \((>)\) and \((\geq)\) in the obvious way.

For example, the set of rational numbers \((\{\text{mathbb}\{Q\}\})\) is an ordered set by letting \((x < y)\) if and only if \((y-x)\) is a positive rational number, that is if \((y-x = \nicefrac{p}{q})\) where \((p, q \in \{\text{mathbb}\{N\}\})\). Similarly, \((\{\text{mathbb}\{N\}\})\) and \((\{\text{mathbb}\{Z\}\})\) are also ordered sets.

There are other ordered sets than sets of numbers. For example, the set of countries can be ordered by landmass, so for example India \((>)\) Lichtenstein. Any time you sort a set in some way, you are making an ordered set. A typical ordered set that you have used since primary school is the dictionary. It is the ordered set of words where the order is the so-called lexicographic ordering. Such ordered sets appear often for example in computer science. In this class we will mostly be interested in ordered set of numbers however.

Let \((E \subset S)\), where \((S)\) is an ordered set.

i. If there exists a \((b \in S)\) such that \((x \leq b)\) for all \((x \in E)\), then we say \((E)\) is bounded above and \((b)\) is an upper bound of \((E)\).

ii. If there exists a \((b \in S)\) such that \((x \geq b)\) for all \((x \in E)\), then we say \((E)\) is bounded below and \((b)\) is a lower bound of \((E)\).

iii. If there exists an upper bound \((b_0)\) of \((E)\) such that whenever \((b)\) is any upper bound for \((E)\) we have \((b_0 \leq b)\), then \((b_0)\) is called the least upper bound or the supremum of \((E)\). We write \([\sup E := b_0].\)

iv. Similarly, if there exists a lower bound \((b_0)\) of \((E)\) such that whenever \((b)\) is any lower bound for \((E)\) we have \((b_0 \geq b)\), then \((b_0)\) is called the greatest lower bound or the infimum of \((E)\). We write \([\inf E := b_0].\)

When a set \((E)\) is both bounded above and bounded below, we say simply that \((E)\) is bounded.

A supremum or infimum for \((E)\) (even if they exist) need not be in \((E)\). For example, the set \((E := \{ x \in \{\text{mathbb}\{Q\}\} : x < 1 \})\) has a least upper bound of 1, but 1 is not in the set \((E)\) itself. On the other hand, if we take \((G := \{ x \in \{\text{mathbb}\{Q\}\} : x \leq 1 \})\), then the least upper bound of \((G)\) is clearly also 1, and in this case \((1 \in G)\). On the other hand, the set \((P := \{ x \in \{\text{mathbb}\{Q\}\} : x \geq 0 \})\) has no upper bound (why?) and therefore it can have no least upper bound. On the other hand 0 is the greatest lower bound of \((P)\).

[defn:lub] An ordered set \((S)\) has the least-upper-bound property if every nonempty subset \((E \subset S)\) that is bounded above has a least upper bound, that is \((\sup E)\) exists in \((S)\).

The least-upper-bound property is sometimes called the completeness property or the Dedekind completeness property.

The set \((\{\text{mathbb}\{Q\}\})\) of rational numbers does not have the least-upper-bound property. The subset \((\{ x \in \{\text{mathbb}\{Q\}\} : x^2 < 2 \})\) does not have a supremum in \((\{\text{mathbb}\{Q\}\})\). The obvious supremum \((\sqrt{2})\) is not rational. Suppose \((x)\) in \((\{\text{mathbb}\{Q\}\})\) such that \((x^2 = 2)\). Write \((x = \nicefrac{m}{n})\) in lowest terms. So
\((\{\nicefrac{m}{n}\} \cup 2^2 = 2)\) or \((m^2 = 2n^2)\). Hence \((m^2 = 2n^2)\) is divisible by 2 and so \((m)\) is divisible by 2. Write \((m = 2k)\) and so \((\{(2k)\}^2 = 2n^2)\). Divide by 2 and note that \((2k^2 = n^2)\), and hence \((n)\) is divisible by 2. But that is a contradiction as \((\{\nicefrac{m}{n}\})\) was in lowest terms.

That \((\{\text{mathbb Q}\})\) does not have the least-upper-bound property is one of the most important reasons why we work with \((\{\text{mathbb R}\})\) in analysis. The set \((\{\text{mathbb Q}\})\) is just fine for algebraists. But analysts require the least-upper-bound property to do any work. We also require our real numbers to have many algebraic properties. In particular, we require that they are a field.

A set \((F)\) is called a field if it has two operations defined on it, addition \((x+y)\) and multiplication \((xy)\), and if it satisfies the following axioms.

1. If \((x \in F)\) and \((y \in F)\), then \((x+y \in F)\).
2. (Commutativity of addition) \((x+y = y+x)\) for all \((x,y \in F)\).
3. (Associativity of addition) \((x+y+z = x+(y+z))\) for all \((x,y,z \in F)\).
4. There exists an element \((0 \in F)\) such that \((0+x = x)\) for all \((x \in F)\).
5. For every element \((x \in F)\) there exists an element \((-x \in F)\) such that \((x + (-x) = 0)\).
6. \((xy)z = x(yz)\) for all \((x,y,z \in F)\).
7. \((1 \not= 0)\) such that \((1x = x)\) for all \((x \in F)\).
8. \((x \not= 0)\) there exists an element \((\nicefrac{1}{x} \in F)\) such that \((x(\nicefrac{1}{x}) = 1)\).
9. \((x+y+z = xy+xz)\) for all \((x,y,z \in F)\).

The set \((\{\text{mathbb Q}\})\) of rational numbers is a field. On the other hand \((\{\text{mathbb Z}\})\) is not a field, as it does not contain multiplicative inverses. For example, there is no \((x \in \{\text{mathbb Z}\})\) such that \((2x = 1)\), so (M5) is not satisfied. You can check that (M5) is the only property that fails. 

We will assume the basic facts about fields that are easily proved from the axioms. For example, \((0x = 0)\) is easily proved by noting that \((xx = (0+x)x = 0x+xx)\), using (A4), (D), and (M2). Then using (A5) on \((xx)\), along with (A2), (A3), and (A4), we obtain \((0 = 0x)\).

A field \((F)\) is said to be an ordered field if \((F)\) is also an ordered set such that:
For \(x,y,z \in F\), \(x < y\) implies \(x+z < y+z\).

ii. For \(x,y \in F\), \(x > 0\) and \(y > 0\) implies \(xy > 0\).

If \(x > 0\), we say \(x\) is positive. If \(x < 0\), we say \(x\) is negative. We also say \(x\) is nonnegative if \(x \geq 0\), and \(x\) is nonpositive if \(x \leq 0\).

For example, it can be checked that the rational numbers \(\mathbb{Q}\) with the standard ordering is an ordered field.

Let \(F\) be an ordered field and \(x,y,z \in F\). Then:

i. If \(x > 0\), then \(-x < 0\) (and vice-versa).

ii. If \(x > 0\) and \(y < z\), then \(xy < xz\).

iii. If \(x < 0\) and \(y < z\), then \(xy > xz\).

iv. If \(x \not= 0\), then \(x^2 > 0\).

v. If \(0 < x < y\), then \(0 < \nicefrac{1}{y} < \nicefrac{1}{x}\).

Note that \(x^2 > 0\) implies in particular that \(1 > 0\).

Let us prove part \(i\). The inequality \(x > 0\) implies by item \(i\) of definition of ordered field that \(x + (-x) > 0 + (-x))\). Now apply the algebraic properties of fields to obtain \(0 > -x\). The “vice-versa” follows by similar calculation.

For part \(ii\), first notice that \(y < z\) implies \(0 < z - y\) by applying item \(i\) of the definition of ordered fields. Now apply item \(ii\) of the definition of ordered fields to obtain \(0 < x(z-y))\). By algebraic properties we get \(0 < xz - xy\), and again applying item \(i\) of the definition we obtain \(xy < xz\).

Part \(iii\) is left as an exercise.

To prove part \(iv\) first suppose \(x > 0\). Then by item \(ii\) of the definition of ordered fields we obtain that \(x^2 > 0\) (use \(y=x\)). If \(x < 0\), we use part \(iii\) of this proposition. Plug in \(y=x\) and \(z=0\).

Finally to prove part \(v\), notice that \(\nicefrac{1}{x}\) cannot be equal to zero (why?). Suppose \(\nicefrac{1}{x} < 0\), then \(\nicefrac{1}{x} > 0\) by \(i\). Then apply part \(i\) (as \(x > 0\)) to obtain \(\nicefrac{1}{x} > \nicefrac{1}{x}\) or \(\nicefrac{1}{x} > 0\), which contradicts \(1 > 0\) by using part \(i\) again. Hence \(\nicefrac{1}{x} > 0\). Similarly \(\nicefrac{1}{y} > 0\). Thus \(\nicefrac{1}{x} \nicefrac{1}{y} > 0\) by definition of ordered field and by part \(ii\) \(\nicefrac{1}{x} \nicefrac{1}{y} x < \nicefrac{1}{x} \nicefrac{1}{y}\) By algebraic properties we get \(\nicefrac{1}{x} < \nicefrac{1}{y}\).

Product of two positive numbers (elements of an ordered field) is positive. However, it is not true that if the product is positive, then each of the two factors must be positive.
Let \((x, y \in F)\) where \((F)\) is an ordered field. Suppose \((xy > 0)\). Then either both \((x)\) and \((y)\) are positive, or both are negative.

Clearly both of the conclusions can happen. If either \((x)\) and \((y)\) are zero, then \((xy)\) is zero and hence not positive. Hence we assume that \((x)\) and \((y)\) are nonzero, and we simply need to show that if they have opposite signs, then \((xy < 0)\).
Without loss of generality suppose \((x > 0)\) and \((y < 0)\). Multiply \((y < 0)\) by \((x)\) to get \((xy < 0x = 0)\). The result follows by contrapositive.

The reader may also know about the complex numbers, usually denoted by \((\mathbb{C})\). That is, \((\mathbb{C})\) is the set of numbers of the form \((x + iy)\), where \((x)\) and \((y)\) are real numbers, and \((i)\) is the imaginary number, a number such that \((i^2 = -1)\). The reader may remember from algebra that \((\mathbb{C})\) is also a field, however, it is not an ordered field. While one can make \((\mathbb{C})\) into an ordered set in some way, it can be proved that it is not possible to put an order on \((\mathbb{C})\) that will make it an ordered field.

**Exercises**


[exercise:finitesethasminmax] Let \((S)\) be an ordered set. Let \((A \subset S)\) be a nonempty finite subset. Then \((A)\) is bounded. Furthermore, \((\inf, A)\) exists and is in \((A)\) and \((\sup, A)\) exists and is in \((A)\). Hint: Use .

Let \((x, y \in F)\), where \((F)\) is an ordered field. Suppose \((0 < x < y)\). Show that \((x^2 < y^2)\).

Let \((S)\) be an ordered set. Let \((B \subset S)\) be bounded (above and below). Let \((A \subset B)\) be a nonempty subset. Suppose all the \((\inf)\)'s and \((\sup)\)'s exist. Show that \((\inf, B \leq \inf, A \land \inf \leq \sup, A \land \inf \leq \sup, B .)\]

Let \((S)\) be an ordered set. Let \((A \subset S)\) and suppose \((b)\) is an upper bound for \((A)\). Suppose \((b \in A)\). Show that \((b = \sup, A)\).

Let \((S)\) be an ordered set. Let \((A \subset S)\) be a nonempty subset that is bounded above. Suppose \((\sup, A)\) exists and \((\sup, A \not\in A)\).

Find a (nonstandard) ordering of the set of natural numbers \((\mathbb{N})\) such that there exists a nonempty proper subset \((A \subsetneq \mathbb{N})\) such that \((\sup, A)\) exists in \((\mathbb{N})\), but \((\sup, A \not\in A)\).

Let \((F = \{0, 1, 2\})\). a) Prove that there is exactly one way to define addition and multiplication so that \((F)\) is a field if \((0)\) and \((1)\) have their usual meaning of (A4) and (M4). b) Show that \((F)\) cannot be an ordered field.

[exercise:dominatingb] Let \((S)\) be an ordered set and \((A)\) is a nonempty subset such that \((\sup, A)\) exists. Suppose there is a \((B \subset A)\) such that whenever \((x \in A)\) there is a \((y \in B)\) such that \((x \leq y)\). Show that \((\sup, B)\) exists and \((\sup, B = \sup, A)\).

Let \((D)\) be the ordered set of all possible words (not just English words, all strings of letters of arbitrary length) using the Latin alphabet using only lower case letters. The order is the lexicographic order as in a dictionary (e.g. \(aaa < dog <\.\))
Let \((A)\) be the subset of \((D)\) containing the words whose first letter is ‘a’ (e.g. \(a \in A\), \(abcd \in A\)). Show that \((A)\) has a supremum and find what it is.

The set of real numbers

Note: 2 lectures, the extended real numbers are optional

The set of real numbers

We finally get to the real number system. To simplify matters, instead of constructing the real number set from the rational numbers, we simply state their existence as a theorem without proof. Notice that \((\mathbb{Q})\) is an ordered field.

There exists a unique ordered field \((\mathbb{R})\) with the such that \((\mathbb{Q})\subset \mathbb{R}\).

Note that also \((\mathbb{N})\subset \mathbb{Q}\). We saw that \((1 > 0)\). By (exercise) we can prove that \((n > 0)\) for all \((n \in \mathbb{N})\). Similarly we can easily verify all the statements we know about rational numbers and their natural ordering.

Let us prove one of the most basic but useful results about the real numbers. The following proposition is essentially how an analyst proves that a number is zero.

If \((x \in \mathbb{R})\) is such that \((x \geq 0)\) and \((x \leq \epsilon)\) for all \((\epsilon > 0)\), then \((x = 0)\).

If \((x > 0)\), then \((0 < \nicefrac{x}{2})\) (why?). Taking \((\epsilon = \nicefrac{x}{2})\) obtains a contradiction. Thus \((x=0)\).

A related simple fact is that any time we have two real numbers \((a < b)\), then there is another real number \((c)\) such that \((a < c < b)\). Just take for example \((c = \frac{a+b}{2})\) (why?). In fact, there are infinitely many real numbers between \((a)\) and \((b)\).

The most useful property of \((\mathbb{R})\) for analysts is not just that it is an ordered field, but that it has the . Essentially we want \((\mathbb{Q})\), but we also want to take suprema (and infima) willy-nilly. So what we do is to throw in enough numbers to obtain \((\mathbb{R})\).

We mentioned already that \((\mathbb{R})\) must contain elements that are not in \((\mathbb{Q})\) because of the . We saw there is no rational square root of two. The set \((\{x \in \mathbb{Q} : x^2 < 2 \}\}) implies the existence of the real number \((\sqrt{2})\), although this fact requires a bit of work.

[example:sqrt2] Claim: There exists a unique positive real number \((r)\) such that \((r^2 = 2)\). We denote \((r)\) by \((\sqrt{2})\).
Take the set \( A := \{ x \in \mathbb{R} : x^2 < 2 \} \). First if \( x^2 < 2 \), then \( x < 2 \). To see this fact, note that \( x^2 < 2 \) implies \( x < \sqrt{2} \) (use \( \sqrt{2} \approx 1.414 \)). Hence, any number \( x \) such that \( x^2 < 2 \) is not in \( A \). Thus \( A \) is bounded above. On the other hand, \( (1 \in A) \), so \( A \) is nonempty.

Let us define \( r := \sup A \). We will show that \( (r^2 = 2) \) by showing that \( (r^2 \leq 2) \) and \( (r^2 \geq 2) \). This is the way analysts show equality, by showing two inequalities. We already know that \( (r \geq 1) \).

In the following, it may seem we are pulling certain expressions out of a hat. When writing a proof such as this we would, of course, come up with the expressions only after playing around with what we wish to prove. The order in which we write the proof is not necessarily the order in which we come up with the proof.

Let us first show that \( (r^2 \leq 2) \). Take a positive number \( s \) such that \( (s^2 < 2) \). We wish to find an \( h > 0 \) such that \( ((s+h)^2 < 2) \). As \( (2-s^2 > 0) \), we have \( (\frac{2-s^2}{2s+1} > 0) \). We choose an \( h \in \mathbb{R} \) such that \( 0 < h < \frac{2-s^2}{2s+1} \). Furthermore, we assume \( h < 1 \).

\[
\begin{aligned}
(s+h)^2 - s^2 & = h(2s + h) \\
& < h(2s+1) & & \quad \bigl(\text{since } h < 1\bigr) \\
& < 2-s^2 & & \left(\text{since } h < \frac{2-s^2}{2s+1} \right).
\end{aligned}
\]

Therefore, \( (s+h)^2 < 2 \). Hence \( (s+h \in A) \), but as \( (h > 0) \) we have \( (s+h > s) \). So \( (s < r = \sup A) \). As \( (s) \) was an arbitrary positive number such that \( (s^2 < 2) \), it follows that \( (r^2 \leq 2) \).

Now take a positive number \( s \) such that \( (s^2 > 2) \). We wish to find an \( h > 0 \) such that \( ((s+h)^2 > 2) \). As \( (s^2 > 2) \) we have \( (\frac{s^2-2}{2s} > 0) \). We choose an \( h \in \mathbb{R} \) such that \( 0 < h < \frac{s^2-2}{2s} \) and \( h < s \).

\[
\begin{aligned}
s^2 - (s-h)^2 & = 2sh - h^2 \\
& < 2sh \\
& < s^2-2 \quad \left(\text{since } h < \frac{s^2-2}{2s} \right).
\end{aligned}
\]

By subtracting \( (s^2) \) from both sides and multiplying by \( (-1) \), we find \( (s-h)^2 > 2 \). Therefore \( (s-h \notin A) \).

Furthermore, if \( (x \geq \text{sgn} s-h) \), then \( (x^2 \geq ((s-h)^2 > 2)) \). As \( (x > 0) \) and \( (s-h > 0) \) and so \( (x \notin A) \). Thus \( (s-h) \) is an upper bound for \( (A) \). However, \( (s-h < s) \), or in other words \( (s > r = \sup A) \). Thus \( (r^2 \geq 2) \).

Together, \( (r^2 \geq 2) \) and \( (r^2 \leq 2) \) imply \( (r^2 = 2) \). The existence part is finished. We still need to handle uniqueness. Suppose \( (s \in \mathbb{R} \setminus \mathbb{Q}) \) such that \( (s^2 = 2) \) and \( (s > 0) \). Thus \( (s^2 = r^2) \). However, if \( (0 < s < r) \), then \( (s^2 < r^2) \). Similarly \( (0 < r < s) \) implies \( (r^2 < s^2) \). Hence \( (s = r) \).

The number \( \sqrt{2} \) is an irrational number. The set \( \{ \mathbb{R} \} \setminus \{ \mathbb{Q} \} \) is called the set of irrational numbers. We just saw that \( \{ \mathbb{R} \} \setminus \{ \mathbb{Q} \} \) is nonempty. Not only is it nonempty, we will see later that it is very large indeed.

Using the same technique as above, we can show that a positive real number \( (x^\frac{1}{n}) \) exists for all \( (n \in \mathbb{N}) \) and all \( (x > 0) \). That is, for each \( (x > 0) \), there exists a unique positive real number \( (r) \) such that \( (r^n = x) \). The proof is left as an exercise.
Archimedean property

As we have seen, there are plenty of real numbers in any interval. But there are also infinitely many rational numbers in any interval. The following is one of the fundamental facts about the real numbers. The two parts of the next theorem are actually equivalent, even though it may not seem like that at first sight.

\[\text{Archimedean property}\]

i. \(\text{(Archimedean property)}\) If \((x, y \in \mathbb{Q})\) and \((x > 0)\), then there exists an \((n \in \mathbb{N})\) such that \(nx > y\).

ii. \(\text{(Q is dense in R)}\) If \((x, y \in \mathbb{R})\) and \((x < y)\), then there exists an \((r \in \mathbb{Q})\) such that \((x < r < y)\).

Let us prove \(\text{Archimedean property}\). We divide through by \((x)\) and then \(\text{(Archimedean property)}\) says that for any real number \((t := \frac{y}{x})\), we can find natural number \((n)\) such that \((n > t)\). In other words, \(\text{(Archimedean property)}\) says that \(\mathbb{N} \subset \mathbb{R}\) is not bounded above. Suppose for contradiction that \(\mathbb{N}\) is bounded above. Let \((b := \sup \mathbb{N})\). The number \((b-1)\) cannot possibly be an upper bound for \(\mathbb{N}\) as it is strictly less than \((b)\) (the supremum). Thus there exists an \((m \in \mathbb{N})\) such that \((m > b-1)\). We add one to obtain \((m+1 > b)\), which contradicts \((b)\) being an upper bound.

Let us tackle \(\text{Q is dense in R}\). First assume \((x \geq 0)\). Note that \((y-x > 0)\). By \(\text{(Archimedean property)}\), there exists an \((n \in \mathbb{N})\) such that \((n(y-x) > 1)\). Also by \(\text{(Archimedean property)}\) the set \(A := \{k \in \mathbb{N} : k > nx\}\) is nonempty. By the well ordering property of \(\mathbb{N}\), \(A\) has a least element \((m)\). As \((m \in A)\), then \((m > nx)\). We divide through by \((n)\) to get \((x < \frac{m}{n})\). As \((m)\) is the least element of \(A\), \((m-1 \not \in A)\). If \((m > 1)\), then \((m-1 \in \mathbb{N})\), but \((m-1 \not \in A)\) and so \((m-1 \leq nx)\). If \((m=1)\), then \((m-1 = 0)\), and \((m-1 \leq nx)\) still holds as \((x \geq 0)\). In other words, \((m-1 \leq nx \leq m)\). On the other hand from \((n(y-x) > 1)\) we obtain \((ny > 1+nx)\). Hence \((ny > 1+nx \geq m)\), and therefore \((y > \frac{m}{n})\). Putting everything together we obtain \((x < \frac{m}{n} < y)\). So let \((r = \frac{m}{n})\).

Now assume \((x < 0)\). If \((y > 0)\), then we just take \((r=0)\). If \((y \leq 0)\), then \((0 \leq y < -x)\), and we find a rational \((q)\) such that \((-y < q < -x)\). Then take \((r = -q)\).

Let us state and prove a simple but useful corollary of the .

\((\inf \{\frac{1}{n} : n \in \mathbb{N}\} = 0)\).

Let \((A := \{\frac{1}{n} : n \in \mathbb{N}\})\). Obviously \((A)\) is not empty. Furthermore, \((\{\frac{1}{n} : n \in \mathbb{N}\}) > 0)\) and so 0 is a lower bound, and \((b := \inf A)\) exists. As 0 is a lower bound, then \((b \geq 0)\). Now take an arbitrary \((a > 0)\).

By the there exists an \((m)\) such that \((na > 1)\), or in other words \((a > \frac{1}{n})\). Therefore \((a)\) cannot be a lower bound for \((A)\). Hence \((b=0)\).
Using supremum and infimum

We want to make sure that suprema and infima are compatible with algebraic operations. For a set \( A \subset \mathbb{R} \) and a number \( x \in \mathbb{R} \) define
\[
\begin{aligned}
x + A & := \{ x+y \in \mathbb{R} : y \in A \} , \\
x A & := \{ xy \in \mathbb{R} : y \in A \} .
\end{aligned}
\]

Let \( A \subset \mathbb{R} \) be nonempty.

1. If \( (x \in \mathbb{R}) \) and \( (A) \) is bounded above, then \( \sup (x+A) = x + \sup A \).

2. If \( (x \in \mathbb{R}) \) and \( (A) \) is bounded below, then \( \inf (x+A) = x + \inf A \).

3. If \( (x > 0) \) and \( (A) \) is bounded above, then \( \sup (xA) = x \left( \sup A \right) \).

4. If \( (x > 0) \) and \( (A) \) is bounded below, then \( \inf (xA) = x \left( \inf A \right) \).

5. If \( (x < 0) \) and \( (A) \) is bounded above, then \( \sup (xA) = x \left( \inf A \right) \).

6. If \( (x < 0) \) and \( (A) \) is bounded below, then \( \inf (xA) = x \left( \sup A \right) \).

Do note that multiplying a set by a negative number switches supremum for an infimum and vice-versa. Also, as the proposition implies that supremum (resp. infimum) of \( (x+A) \) or \( (xA) \) exists, it also implies that \( (x+A) \) or \( (xA) \) is nonempty and bounded from above (resp. from below).

Let us only prove the first statement. The rest are left as exercises.

Suppose \( (b) \) is an upper bound for \( (A) \). That is, \( (y \leq b) \) for all \( (y \in A) \). Then \( (x+y \leq x+b) \) for all \( (y \in A) \), and so \( (x+b) \) is an upper bound for \( (x+A) \). In particular, if \( (b = \sup A) \), then \( (\sup (x+A) \leq x+b = x+ \sup A) \).

The other direction is similar. If \( (b) \) is an upper bound for \( (x+A) \), then \( (x+y \leq b) \) for all \( (y \in A) \) and so \( (y \leq b-x) \) for all \( (y \in A) \). So \( (b-x) \) is an upper bound for \( (A) \). If \( (b = \sup (x+A)) \), then \( (\sup A \leq b-x = \sup (x+A) - x) \). And the result follows.

Sometimes we need to apply supremum or infimum twice. Here is an example.

\[\text{[infsupineq:prop] Let} \quad (A, B \subset \mathbb{R}) \text{ be nonempty sets such that } (x \leq y) \text{ whenever } (x \in A) \text{ and } (y \in B). \quad \text{Then } (A) \text{ is bounded above, } (B) \text{ is bounded below, and } (\sup A \leq \inf B).\]

Any \( (x \in A) \) is a lower bound for \( (B) \). Therefore \( (x \leq \inf B) \) for all \( (x \in A) \), so \( (\inf B) \) is an upper bound for \( (A) \). Hence, \( (\sup A \leq \inf B) \).

We must be careful about strict inequalities and taking suprema and infima. Note that \( (x < y) \) whenever \( (x \in A) \) and \( (y \in B) \) still only implies \( (\sup A \leq \inf B) \), and not a strict inequality. This is an important subtle point that comes up often. For example, take \( \{A := \{ 0 \}\} \) and take \( \{B := \{ \frac{1}{n} : n \in \mathbb{N} \}\} \). Then \( 0 < \frac{1}{n} \) for all \( n \in \mathbb{N} \). However, \( (\sup A = 0) \) and \( (\inf B = 0) \).
The proof of the following often used elementary fact is left to the reader. A similar statement holds for infima.

[prop:existsxepsfromsup] If \((S \subset \{\text{mathbb \{R\}\})}\) is a nonempty set, bounded from above, then for every \((\epsilon > 0)\) there exists \((x \in S)\) such that \((\sup S - \epsilon < x \leq \sup S)\).

To make using suprema and infima even easier, we may want to write \((\sup A)\) and \((\inf A)\) without worrying about \((A)\) being bounded and nonempty. We make the following natural definitions.

Let \((A \subset \{\text{mathbb \{R\}\})\) be a set.

i. If \((A)\) is empty, then \((\sup A := \infty)\).

ii. If \((A)\) is not bounded above, then \((\sup A := \infty)\).

iii. If \((A)\) is empty, then \((\inf A := \infty)\).

iv. If \((A)\) is not bounded below, then \((\inf A := -\infty)\).

For convenience, \((\infty)\) and \((-\infty)\) are sometimes treated as if they were numbers, except we do not allow arbitrary arithmetic with them. We make \((\{\text{mathbb \{R\}\})\}^{\ast\ast} := \{\text{mathbb \{R\}\})\}^{\cup \{\infty, \inf \}}\) into an ordered set by letting \([-\infty < \infty \quad \text{and} \quad x \in \{\text{mathbb \{R\}\})\}^{\} \quad \text{for all } x \in \{\text{mathbb \{R\}\})\}^{\}}\). The set \((\{\text{mathbb \{R\}\})\}^{\ast\ast})\) is called the set of extended real numbers. It is possible to define some arithmetic on \((\{\text{mathbb \{R\}\})\}^{\ast\ast})\). Most operations are extended in an obvious way, but we must leave \((\infty - \infty)\), \((0 \cdot \infty)\), \((-\infty)\) undefined. We refrain from using this arithmetic, it leads to easy mistakes as \((\{\text{mathbb \{R\}\})\}^{\ast\ast})\) is not a field. Now we can take suprema and infima without fear of emptiness or unboundedness. In this book we mostly avoid using \((\{\text{mathbb \{R\}\})\}^{\ast\ast})\) outside of exercises, and leave such generalizations to the interested reader.

**Maxima and minima**

By we know a finite set of numbers always has a supremum or an infimum that is contained in the set itself. In this case we usually do not use the words supremum or infimum.

When a set \((A)\) of real numbers is bounded above, such that \((\sup A \in A)\), then we can use the word maximum and the notation \((\max A)\) to denote the supremum. Similarly for infimum; when a set \((A)\) is bounded below and \((\inf A \in A)\), then we can use the word minimum and the notation \((\min A)\). For example, \(\begin{aligned} & \max \{1,2,4,\pi,100\} = 100, \quad \min \{1,2,4,\pi,100\} = 1. \end{aligned}\) While writing \((\sup A)\) and \((\inf A)\) may be technically correct in this situation, \((\max A)\) and \((\min A)\) are generally used to emphasize that the supremum or infimum is in the set itself.

**Exercises**

Prove that if \((t > 0)\) \((t \in \{\text{mathbb \{R\}\})\})\), then there exists an \((n \in \{\text{mathbb \{N\}\})\})\) such that \((\frac{n^2}{1} < t)\).

Prove that if \((t \geq 0)\) \((t \in \{\text{mathbb \{R\}\})\})\), then there exists an \((n \in \{\text{mathbb \{N\}\})\})\) such that \((n-1 \leq t < n)\).
Finish the proof of .

Let \( (x, y) \in \mathbb{R} \). Suppose \( x^2 + y^2 = 0 \). Prove that \( x = 0 \) and \( y = 0 \).

Show that \( \sqrt{3} \) is irrational.

Let \( n \in \mathbb{N} \). Show that either \( \sqrt{n} \) is either an integer or it is irrational.

Prove the arithmetic-geometric mean inequality. That is, for two positive real numbers \( x, y \) we have \( \sqrt{xy} \leq \frac{x+y}{2} \). Furthermore, equality occurs if and only if \( x = y \).

Show that for any two real numbers \( x \) and \( y \) such that \( x < y \), there exists an irrational number \( s \) such that \( x < s < y \). Hint: Apply the density of \( \mathbb{Q} \) to \( \frac{x}{\sqrt{2}} \) and \( \frac{y}{\sqrt{2}} \).

[exercise:supofsum] Let \( A \) and \( B \) be two nonempty bounded sets of real numbers. Let \( C := \{ a+b : a \in A, b \in B \} \). Show that \( C \) is a bounded set and that \( \sup C = \sup A + \sup B \).

Let \( A \) and \( B \) be two nonempty bounded sets of nonnegative real numbers. Define the set \( C := \{ ab : a \in A, b \in B \} \). Show that \( C \) is a bounded set and that \( \sup C = \sup A \cdot \sup B \).

Given \( x > 0 \) and \( n \in \mathbb{N} \), show that there exists a unique positive real number \( r \) such that \( x = r^n \). Usually \( r \) is denoted by \( x^{1/n} \).

Prove .

[exercise:bernoulliineq] Prove the so-called Bernoulli’s inequality: If \( 1+x > 0 \) then for all \( n \in \mathbb{N} \) we have \( (1+x)^n \geq 1+nx \).

---

**Absolute value**

*Note: 0.5–1 lecture*

A concept we will encounter over and over is the concept of absolute value. You want to think of the absolute value as the “size” of a real number. Let us give a formal definition. \[ |x| := \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases} \]

Let us give the main features of the absolute value as a proposition.

[prop:absbas]

i. \( |x| \geq 0 \), and \( |x| = 0 \) if and only if \( x = 0 \).

ii. \( |-x| = |x| \) for all \( x \in \mathbb{R} \) (with \( \mathbb{R} \) denoting the set of real numbers).

---

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
iii. \[ \forall x, y \in \mathbb{R}, |xy| = |x||y| \]

iv. \[ \forall x \in \mathbb{R}, |x|^2 = x^2 \]

v. \[ \forall x \in \mathbb{R}, |x| \leq y \iff -y \leq x \leq y \]

vi. \[ \forall x \in \mathbb{R}, -|x| \leq x \leq |x| \]

[prop:absbas:i]: This statement is not difficult to see from the definition.

[prop:absbas:ii]: Suppose \( x > 0 \), then \( |-x| = -(-x) = x = |x| \). Similarly when \( x < 0 \), or \( x = 0 \).

[prop:absbas:iii]: If \( x \) or \( y \) is zero, then the result is obvious. When \( x \) and \( y \) are both positive, then \( |x| = x \) and \( |y| = y \). If \( x \) and \( y \) are both negative then \( (xy) \) is still positive and \( (xy = |x||y|) \). Next assume \( (x > 0) \) and \( (y > 0) \). Then \( |x||y| = (xy) = x(y) = xy \). Now \( (xy) \) is negative and hence \( xy = -(xy) \). Similarly if \( (x < 0) \) and \( (y < 0) \).

[prop:absbas:iv]: Obvious if \( x \geq 0 \). If \( x < 0 \), then \( |x|^2 = (-x)^2 = x^2 \).

[prop:absbas:v]: Suppose \( x, y \in \mathbb{R} \). Obviously \( y \geq 0 \) and hence \( x \leq y \). If \( x < 0 \), then \( (x \geq y) \) means \( x \leq y \). Negating both sides we get \( x < y \). Again \( y \geq 0 \) and so \( y > x \). Hence, \( y > x \).

On the other hand, suppose \( y \geq x \) is true. If \( y \geq 0 \), then \( x < y \). If \( x < 0 \), then \( y \geq x \) implies \( (-x) \geq y \), which is equivalent to \( |x| \geq y \).

[prop:absbas:vi]: Apply [prop:absbas:v] with \( y = |x| \).

A property used frequently enough to give it a name is the so-called triangle inequality.

\[ \forall x, y \in \mathbb{R}, |x+y| \leq |x| + |y| \]

From we have \( |x+y| \leq |x| + |y| \) and \( |x+y| \leq |x| + |y| \). We add these two inequalities to obtain \[ |x+y| \leq |x| + |y| \]. Again by we have \( |x+y| \leq |x| + |y| \).

There are other often applied versions of the triangle inequality.

Let \( (x, y \in \mathbb{R}) \)
\[ i. \text{ (reverse triangle inequality)} \quad \forall x, y \in \mathbb{R}, \quad |x - y| \leq |x| - |y|. \]

\[ ii. \quad |x - y| \leq |x| + |y|. \]

Let us plug in \((x=a-b)\) and \((y=b)\) into the standard triangle inequality to obtain \(|a| = |a-b+b| \leq |a-b| + |b|\), or \(|a| - |b| \leq |a-b|\). Switching the roles of \((a)\) and \((b)\) we obtain \(|b| - |a| \leq |b-a| = |a-b|\). Now applying again we obtain the reverse triangle inequality.

The second version of the triangle inequality is obtained from the standard one by just replacing \((y)\) with \((-y)\) and noting again that \(|-y| = |y|\).

Let \(x_1, x_2, \ldots, x_n \in \mathbb{R}\). Then \(|x_1 + x_2 + \cdots + x_n| \leq |x_1| + |x_2| + \cdots + |x_n|\). We proceed by induction. The conclusion holds trivially for \(n=1\), and for \(n=2\) it is the standard triangle inequality. Suppose the corollary holds for \(n\). Take \(n+1\) numbers \((x_1, x_2, \ldots, x_{n+1})\) and first use the standard triangle inequality, then the induction hypothesis \(|x_1 + x_2 + \cdots + x_n| \leq |x_1| + |x_2| + \cdots + |x_n|\) to get \(|x_1 + x_2 + \cdots + x_{n+1}| \leq |x_1 + x_2 + \cdots + x_n| + |x_{n+1}|\).

Let us see an example of the use of the triangle inequality.

Find a number \(M\) such that \(|x^2-9x+1| \leq M\) for all \(-1 \leq x \leq 5\).

Using the triangle inequality, write \(|x^2-9x+1| \leq |x^2| + |-9x| + |1| = |x|^2 + 9|x| + 1\). It is obvious that \(|x|^2 + 9|x| + 1\) is largest when \(|x|\) is largest. In the interval provided, \(|x|^2 + 9|x| + 1\) is largest when \(|x|=5\) and so \(|5|^2 + 9|5| + 1 = 71\). One possibility for \(M\) is \(M = 71\). There are, of course, other \(M\) that work.

The last example leads us to the concept of bounded functions.

Suppose \(f : D \to \mathbb{R}\) is a function. We say \(f\) is bounded if there exists a number \(M\) such that \(|f(x)| \leq M\) for all \(x \in D\).

In the example we proved \(|x^2-9x+1|\) is bounded when considered as a function on \((-1 \leq x \leq 5)\). On the other hand, if we consider the same polynomial as a function on the whole real line \((-\infty \leq x \leq \infty)\), then it is not bounded.

For a function \(f : D \to \mathbb{R}\) we write \(\sup_{x \in D} f(x)\) and \(|f| := \sup_{x \in D} |f(x)|\). For example if, as
before, \( f(x) = x^2 - 9x + 1 \), for \(-1 \leq x \leq 5\), a little bit of calculus shows 
\[
\sup_{x \in D} f(x) = \sup_{-1 \leq x \leq 5} (x^2 -9x+1) = 11, \quad \inf_{x \in D} f(x) = \inf_{-1 \leq x \leq 5} (x^2 -9x+1) = \text{nicefrac{-77}{4}}.
\]

If \( f \colon D \to \mathbb{R} \) and \( g \colon D \to \mathbb{R} \) (\( D \) nonempty) are bounded functions and \( f(x) \leq g(x) \quad \text{for all } x \in D \), then \( \sup_{x \in D} f(x) \leq \sup_{x \in D} g(x) \) and \( \inf_{x \in D} f(x) \leq \inf_{x \in D} g(x) \).

You should be careful with the variables. The \( x \) on the left side of the inequality in \[\text{(prop:funcsupinf:eq)}\] is different from the \( x \) on the right. You should really think of the first inequality as \( \sup_{x \in D} f(x) \leq \sup_{y \in D} g(y) \). Let us prove this inequality. If \( b \) is an upper bound for \( g(D) \), then \( f(x) \leq g(x) \leq b \) for all \( x \in D \), and hence \( b \) is an upper bound for \( f(D) \). Taking the least upper bound we get that for all \( x \in D \) \( f(x) \leq \sup_{y \in D} g(y) \). Therefore \( \sup_{y \in D} g(y) \) is an upper bound for \( f(D) \) and thus greater than or equal to the least upper bound of \( f(D) \). \( \sup_{x \in D} f(x) \leq \sup_{y \in D} g(y) \). The second inequality (the statement about the inf) is left as an exercise.

A common mistake is to conclude \( \sup_{x \in D} f(x) \leq \inf_{y \in D} g(y) \) is not true given the hypothesis of the claim above. For this stronger inequality we need the stronger hypothesis \( f(x) \leq g(y) \quad \text{for all } x \in D \text{ and } y \in D \). The proof as well as a counterexample is left as an exercise.

**Exercises**

Show that \( \left| x-y \right| < \epsilon \) if and only if \( x-\epsilon < y < x+\epsilon \).

Show that

a. \( \max \{x,y\} = \frac{x+y+\left| x-y \right|}{2} \)

b. \( \min \{x,y\} = \frac{x+y-\left| x-y \right|}{2} \)

Find a number \( M \) such that \( \left| x^3-x^2+8x \right| \leq M \) for all \(-2 \leq x \leq 10\).

Finish the proof of . That is, prove that given any set \( D \), and two bounded functions \( f \colon D \to \mathbb{R} \) and \( g \colon D \to \mathbb{R} \) such that \( f(x) \leq g(x) \) for all \( x \in D \), then \( \inf_{x \in D} f(x) \leq \inf_{x \in D} g(x) \).

Let \( f \colon D \to \mathbb{R} \) and \( g \colon D \to \mathbb{R} \) be functions \( (D \) nonempty).

a. Suppose \( f(x) \leq g(y) \) for all \( x \in D \) and \( y \in D \). Show that \( \sup_{x \in D} f(x) \leq \inf_{x \in D} g(x) \).

b. Find a specific \( D \), \( f \), and \( g \), such that \( f(x) \leq g(x) \) for all \( x \in D \), but \( \sup_{x \in D} f(x) > \inf_{x \in D} g(x) \).

Prove without the assumption that the functions are bounded. Hint: You need to use the extended real numbers.
Let \( D \) be a nonempty set. Suppose \( f : D \to \mathbb{R} \) and \( g : D \to \mathbb{R} \) are bounded functions. a) Show
\[
\sup_{x \in D} (f(x) + g(x)) \leq \sup_{x \in D} f(x) + \sup_{x \in D} g(x)
\]
\[
\inf_{x \in D} (f(x) + g(x)) \geq \inf_{x \in D} f(x) + \inf_{x \in D} g(x)
\]
b) Find examples where we obtain strict inequalities.

**Intervals and the size of \( \mathbb{R} \)**

*Note: 0.5–1 lecture (proof of uncountability of \( \mathbb{R} \) can be optional)*

You surely saw the notation for intervals before, but let us give a formal definition here. For \( a, b \in \mathbb{R} \) such that \( a < b \) we define
\[
\begin{aligned}
\[a,b\] &= \{ x \in \mathbb{R} : a \leq x \leq b \}, \\
(a,b) &= \{ x \in \mathbb{R} : a < x < b \}, \\
(a,b] &= \{ x \in \mathbb{R} : a < x \leq b \}, \\
[a,b) &= \{ x \in \mathbb{R} : a \leq x < b \}.
\end{aligned}
\]

The interval \([a,b]\) is called a **closed interval** and \((a,b)\) is called an **open interval**. The intervals of the form \((a,b]\) and \([a,b)\) are called **half-open intervals**.

The above intervals were all **bounded intervals**, since both \((a)\) and \((b)\) were real numbers. We define **unbounded intervals**, where \(a \in \mathbb{R}\) and \(b \in \mathbb{R}\), such that \(a < b\) and \(a = -\infty\) or \(b = \infty\)
\[
\begin{aligned}
\[a,\infty) &= \{ x \in \mathbb{R} : a \leq x \}, \\
(a,\infty) &= \{ x \in \mathbb{R} : a < x \}, \\
(-\infty,b] &= \{ x \in \mathbb{R} : x \leq b \}, \\
(-\infty,b) &= \{ x \in \mathbb{R} : x < b \}.
\end{aligned}
\]

For completeness we define \((-\infty,\infty) = \mathbb{R}\).

In short, an interval is a set \( I \subseteq \mathbb{R} \) with at least 2 elements, such that if \( a < b < c \) and \( a, c \in I \), then \( b \in I \). See .

We have already seen that any open interval \((a,b)\) (where \(a < b\) of course) must be nonempty. For example, it contains the number \( \frac{a+b}{2} \). An unexpected fact is that from a set-theoretic perspective, all intervals have the same “size,” that is, they all have the same cardinality. For example the map \( f(x) = 2x \) takes the interval \([0,1]\) bijectively to the interval \([0,2]\). Maybe more interestingly, the function \( f(x) = \tan(x) \) is a bijective map from \((-\frac{\pi}{2},\frac{\pi}{2}) \) to \( \mathbb{R} \). Hence the bounded interval \((-\frac{\pi}{2},\frac{\pi}{2}) \) has the same cardinality as \( \mathbb{R} \). It is not completely straightforward to construct a bijective map from \([0,1]\) to \((0,1)\), but it is possible.

And do not worry, there does exist a way to measure the “size” of subsets of real numbers that “sees” the difference between \([0,1]\) and \((0,1)\). However, its proper definition requires much more machinery than we have right now.

Let us say more about the cardinality of intervals and hence about the cardinality of \( \mathbb{R} \). We have seen that there exist irrational numbers, that is \( \mathbb{R} \setminus \mathbb{Q} \) is nonempty. The question is: How many irrational numbers are there? It turns out there are a lot more irrational numbers than rational numbers. We have seen that \( \mathbb{Q} \) is countable, and we will show that \( \mathbb{R} \) is uncountable. In fact, the cardinality of \( \mathbb{R} \) is the same as the cardinality of \( \mathcal{P}(\mathbb{N}) \), although we will not prove this claim here.
We give a modified version of Cantor's original proof from 1874 as this proof requires the least setup. Normally this proof is stated as a contradiction proof, but a proof by contrapositive is easier to understand.

Let \((X \subseteq \mathbb{R})\) be a countably infinite subset such that for any two real numbers \((a < b)\), there is an \((x \in X)\) such that \((a < x < b)\). Were \((\mathbb{R})\) countable, then we could take \((X = \mathbb{R})\). If we show that \((X)\) is necessarily a proper subset, then \(X = \mathbb{R}\) cannot equal \((\mathbb{R})\), and \((\mathbb{R})\) must be uncountable.

As \((X)\) is countably infinite, there is a bijection from \((\mathbb{N})\) to \((X)\). Consequently, we write \((X)\) as a sequence of real numbers \((x_1, x_2, x_3, \ldots)\), such that each number in \((X)\) is given by \((x_j)\) for some \(j \in \mathbb{N}\). Let us inductively construct two sequences of real numbers \((a_1, a_2, a_3, \ldots)\) and \((b_1, b_2, b_3, \ldots)\). Let \(a_1 := x_1\) and \(b_1 := x_1 + 1\). Note that \((a_1 < b_1)\) and \((x_1 \notin (a_1, b_1))\). For \((k > 1)\), suppose \((a_{k-1})\) and \((b_{k-1})\) has been defined. Let us also suppose \((x_{j \in \{a_{k-1}, b_{k-1}\}})\) does not contain any \((x_j)\) for any \(j = 1, \ldots, k-1\).

(i) Define \((a_k := x_j)\), where \(j\) is the smallest \(j \in \mathbb{N}\) such that \((x_j \in (a_{k-1}, b_{k-1}))\). Such an \((x_j)\) exists by our assumption on \((X)\).

(ii) Next, define \((b_k := x_j)\) where \(j\) is the smallest \(j \in \mathbb{N}\) such that \((x_j \in (a_k, b_{k-1}))\).

Notice that \((a_k < b_k)\) and \((a_k < a_{k+1} < b_k < b_{k+1})\). Also notice that \((x_k \notin (a_k, b_k))\) does not contain \((x_k)\) and hence does not contain any \((x_j)\) for \(j = 1, \ldots, k\).

Claim: \((a_j < b_k)\) for all \((j)\) and \((k)\) in \((\mathbb{N})\). Let us first assume \((j < k)\). Then \((a_j < a_{j+1} < \cdots < a_{k-1} < a_k < b_k)\). Similarly for \((j > k)\). The claim follows.

Let \((A = \{ a_j : j \in \mathbb{N}\})\) and \((B = \{ b_j : j \in \mathbb{N}\})\). By and the claim above we have \((\sup A, A \notin \mathbb{N})\). Define \((y := \sup A)\). The number \((y)\) cannot be a member of \((A)\). If \((y = a_j)\) for some \((j)\), then \((y < a_{j+1})\), which is impossible. Similarly \((y)\) cannot be a member of \((B)\). Therefore, \((a_j < y)\) for all \((j)\) in \((\mathbb{N})\) and \((y < b_j)\) for all \((j)\) in \((\mathbb{N})\). In other words \((y \notin (a_j, b_j))\) for all \((j)\) in \((\mathbb{N})\).

Finally we must show that \((y \notin \mathbb{N})\). If we do so, then we will have constructed a real number not in \((X)\) showing that \((X)\) must have been a proper subset. Take any \((x_k \in X)\). By the above construction \((x_k \notin (a_k, b_k))\), so \((x_k \neq y)\) as \((y \notin (a_k, b_k))\).

Therefore, the sequence \((x_1, x_2, \ldots)\) cannot contain all elements of \((\mathbb{R})\) and thus \((\mathbb{R})\) is uncountable.

**Exercises**

For \((a < b)\), construct an explicit bijection from \((\{a,b\})\) to \((\{0,1\})\).

Suppose \((f : [0,1] \to (0,1))\) is a bijection. Using \((f)\), construct a bijection from \(([-1,1])\) to \((\mathbb{R})\).
Suppose $I \subset \mathbb{R}$ is a subset with at least 2 elements such that if $a < b < c$ and $(a, c) \in I$, then it is one of the nine types of intervals explicitly given in this section. Furthermore, prove that the intervals given in this section all satisfy this property.

Construct an explicit bijection from $\langle(0,1)\rangle$ to $\langle(0,1)\rangle$. Hint: One approach is as follows: First map $\langle(\nicefrac{1}{2},1]\rangle$ to $\langle(0,\nicefrac{1}{2}]\rangle$, then map $\langle(\nicefrac{1}{4},\nicefrac{1}{2}]\rangle$ to $\langle(\nicefrac{1}{2},\nicefrac{3}{4}]\rangle$, etc…. Write down the map explicitly, that is, write down an algorithm that tells you exactly what number goes where. Then prove that the map is a bijection.

Construct an explicit bijection from $\langle[0,1]\rangle$ to $\langle(0,1)\rangle$.

a) Show that every closed interval $\langle[a,b]\rangle$ is the intersection of countably many open intervals. b) Show that every open interval $\langle(a,b)\rangle$ is a countable union of closed intervals. c) Show that an intersection of a possibly infinite family of closed intervals is either empty, a single point, or a closed interval.

Suppose $\langle S \rangle$ is a set of disjoint open intervals in $\langle \{\text{mathbb} \{R\} \} \rangle$. That is, if $\langle(a,b) \in S \rangle$ and $\langle(c,d) \in S \rangle$, then either $\langle(a,b) = (c,d)\rangle$ or $\langle(a,b) \cap (c,d) = \emptyset\rangle$. Prove $\langle S \rangle$ is a countable set.

Prove that the cardinality of $\langle[0,1]\rangle$ is the same as the cardinality of $\langle(0,1)\rangle$ by showing that $\langle\left\lfloor 0,1 \right\rfloor \rangle$ is a countable set and $\langle\left\lceil 0,1 \right\rceil \rangle$ is a countable set. See . Note that this requires the Cantor-Bernstein-Schroeder theorem we stated without proof. Also note that this proof does not give you an explicit bijection.

A number $\langle x \rangle$ is algebraic if $\langle x \rangle$ is a root of a polynomial with integer coefficients, in other words, $\langle a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 = 0 \rangle$ where all $\langle a_n \in \{\text{mathbb} \{Z\} \} \rangle$. a) Show that there are only countably many algebraic numbers. b) Show that there exist non-algebraic numbers (follow in the footsteps of Cantor, use uncountability of $\langle \{\text{mathbb} \{R\} \} \rangle$). Hint: Feel free to use the fact that a polynomial of degree $\langle n \rangle$ has at most $\langle n \rangle$ real roots.

---

**Decimal representation of the reals**

*Note: 1 lecture (optional)*

We often think of real numbers as their *decimal representation*. For a positive integer $\langle n \rangle$, we find the digits $\langle d_K, d_{K-1}, d_2, d_1, d_0 \rangle$ for some $\langle K \rangle$, where each $\langle d_j \rangle$ is an integer between $\langle 0 \rangle$ and $\langle 9 \rangle$, then $\langle n = d_K 10^K + d_{K-1} 10^{K-1} + \ldots + d_1 10 + d_0 \rangle$. We often assume $\langle d_K \not= 0 \rangle$. To represent $\langle n \rangle$ we write the sequence of digits: $\langle n = d_K d_{K-1} \ldots d_1 d_0 \rangle$. By a (decimal) *digit*, we mean an integer between $\langle 0 \rangle$ and $\langle 9 \rangle$.

Similarly we represent some rational numbers. That is, for certain numbers $\langle x \rangle$, we can find negative integer $\langle -M \rangle$, a positive integer $\langle K \rangle$, and digits $\langle d_K, d_{K-1}, d_2, d_1, d_0, d_{-1}, \ldots, d_{-M} \rangle$, such that $\langle x = d_K 10^K + d_{K-1} 10^{K-1} + \ldots + d_{-1} 10^{-1} + d_{-2} 10^{-2} + \ldots + d_{-M} 10^{-M} \rangle$. We write $\langle x = \ldots \rangle$.
Not every real number has such a representation, even the simple rational number \(\frac{1}{3}\) does not. The irrational number \(\sqrt{2}\) does not have such a representation either. To get a representation for all real numbers we must allow infinitely many digits.

Let us from now on consider only real numbers in the interval \((0,1]\). If we find a representation for these, we simply add integers to them to obtain a representation for all real numbers. Suppose we take an infinite sequence of decimal digits: \\
\(0.d_1d_2d_3\ldots\). That is, we have a digit \(d_j\) for every \(j \in \mathbb{N}\). We have renumbered the digits to avoid the negative signs. We say this sequence of digits represents a real number \(x\) if \\
\[x = \sup_{n \in \mathbb{N}} \left( \frac{d_1}{10} + \frac{d_2}{10^2} + \frac{d_3}{10^3} + \cdots + \frac{d_n}{10^n} \right) .\]

We call \\
\[D_n := \frac{d_1}{10} + \frac{d_2}{10^2} + \frac{d_3}{10^3} + \cdots + \frac{d_n}{10^n}\]

the truncation of \(x\) to \(n\) decimal digits.

[prop:decimalprop]

i. Every infinite sequence of digits \(0.d_1d_2d_3\ldots\) represents a unique real number \(x \in [0,1]\).

ii. For every \(x \in (0,1]\) there exists an infinite sequence of digits \(0.d_1d_2d_3\ldots\) that represents \(x\). There exists a unique representation such that \\
\[D_n < x \leq D_n + \frac{1}{10^n} \quad \text{for all } n \in \mathbb{N} .\]

Let us start with the first item. Suppose there is an infinite sequence of digits \(0.d_1d_2d_3\ldots\). We use the geometric sum formula to write \\
\[D_n = \frac{d_1}{10} + \frac{d_2}{10^2} + \frac{d_3}{10^3} + \cdots + \frac{d_n}{10^n} = 1 - \left( 1 - \frac{1}{10} \right)^n < 1 .\]

In particular, \(D_n < 1\) for all \(n\). As \(D_n \geq 0\) is obvious, we obtain \\
\[0 \leq \sup_{n \in \mathbb{N}} D_n \leq 1 ,\]

and therefore \(0.d_1d_2d_3\ldots\) represents a unique number \(x \in [0,1]\).

We move on to the second item. Take \(x \in (0,1]\). First let us tackle the existence. By convention define \(D_0 := 0\), then automatically we obtain \(D_0 < x \leq D_0 + \{10\}^{-0}\). Suppose for induction that we defined all the digits \(\{d_1,d_2,d_3\ldots\}\), and that \(D_n < x \leq D_n + \{10\}^{-n}\). We need to define \(\{d_{n+1}\}\).

By the of the real numbers we find an integer \(j\) such that \\
\(\{x-D_n \leq j \{10\}^{-1}\}\). We take the least such \(j\) and obtain \\
\[\{\text{eq:theDnjineq}\} \quad \{D_{n+1}\} < x-D_n \leq j \{10\}^{-1}\].\]

Let \(d_{n+1} := j-1\). As \(D_n < x\), then \(d_{n+1} = j-1 \geq 0\). On the other hand since \(\{x-D_n \leq \{10\}^{-n}\}\) we have that \(j\) is at most 10, and therefore \(d_{n+1} \leq 9\). So \(d_{n+1}\) is a decimal digit. Since \(\{D_{n+1}\} = D_n + d_{n+1} \{10\}^{-1}\) we add \(D_{n+1}\) to the inequality above: \\
\[\{D_{n+1}\} = D_n + (j-1) \{10\}^{-1} < x \leq D_n + j \{10\}^{-1} = D_n + (j-1) \{10\}^{-1} + (10^{-1} = D_n + (j-1) \{10\}^{-1} + (10^{-1} + (10\}^{-1} + (10\}^{-1}\) holds. We have inductively defined an infinite sequence of digits \(0.d_1d_2d_3\ldots\). As \(D_n < x\) for all \(n\), then \(\sup
\[
D_n : n \in \{\text{mathbb}\{N\}\} \ \text{and} \ \sup \{D_m : m \in \{\text{mathbb}\{N\}\}\} = x).
\]

As \(x-{10}^{-n} \leq D_n\), then \(x-{10}^{-n} \leq \sup \{D_n : n \in \{\text{mathbb}\{N\}\}\}\) for all \(n\). The two inequalities together imply \(\sup \{D_n : n \in \{\text{mathbb}\{N\}\}\} = x\).

What is left to show is the uniqueness. Suppose \(0.e_1e_2e_3\ldots\) is another representation of \(x\). Let \((E_n)\) be the \(n\)-digit truncation of \(0.e_1e_2e_3\ldots\), and suppose \(E_n < x \leq E_n + {10}^{-n}\) for all \(n\). Suppose for some \(K \in \{\text{mathbb}\{N\}\}\), \(e_n = d_n\) for all \(n < K\), so \(D_{K-1} = E_{K-1}\). Then \(E_K = D_{K-1} + e_K{10}^{-K} < x \leq E_K + {10}^{-K}\). Subtracting \(D_{K-1}\) and multiplying by \(10\) we get \(e_K < (x - D_{K-1}){10}^K\). Similarly we obtain \(d_K < (x - D_{K-1}){10}^K\). Hence, both \(e_K\) and \(d_K\) are the largest integer \(j\) such that \(j < (x - D_{K-1}){10}^K\), and therefore \(e_K = d_K\). That is, the representation is unique.

The representation is not unique if we do not require the extra condition in the proposition. For example, for the number \(\nicefrac{1}{2}\) the method in the proof obtains the representation \(0.49999\ldots\). However, we also have the representation \(0.5000\ldots\). The key requirement that makes the representation unique is \(D_n < x\) for all \(n\). The inequality \(x \leq D_n + {10}^{-n}\) is true for every representation by the computation in the beginning of the proof.

The only numbers that have nonunique representations are ones that end either in an infinite sequence of \(0\)'s or \(9\)'s, because the only representation for which \(D_n = x\) is one where all digits past \(n\)th one are zero. In this case there are exactly two representations of \(x\) (see the exercises).

Let us give another proof of the uncountability of the reals using decimal representations. This is Cantor’s second proof, and is probably more well known. While this proof may seem shorter, it is because we have already done the hard part above and we are left with a slick trick to prove that \(\{\text{mathbb}\{R\}\}\) is uncountable. This trick is called Cantor diagonalization and finds use in other proofs as well.

The set \(\{0,1\}\) is uncountable.

Let \(X := \{\ x_1,x_2,x_3,\ldots \}\) be any countable subset of real numbers in \(\{0,1\}\). We will construct a real number not in \(X\). Let \(x_n = 0.d_1^nd_2^nd_3^n\ldots\) be the unique representation from the proposition, that is \(d_j^n\) is the \(j\)th digit of the \(n\)th number. Let \(e_n := 1\) if \(d_n^n \not= 1\), and let \(e_n := 2\) if \(d_n^n = 1\). Let \(E_n\) be the \(n\)-digit truncation of \(y = 0.e_1e_2e_3\ldots\). Because all the digits are nonzero we get that \(E_n < y\). Therefore \(E_n < y \leq E_n + 1\) for all \(n\), and the representation is the unique one for \(y\) from the proposition.

But for every \(n\), the \(n\)th digit of \(y\) is different from the \(n\)th digit of \(x_n\), so \(y \not= x_n\). Therefore \(y\) was an arbitrary countable subset, \(\{0,1\}\) must be uncountable.

Using decimal digits we can also find lots of numbers that are not rational. The following proposition is true for every rational number, but we give it only for \(x \in (0,1]\) for simplicity.

If \(x \in (0,1]\) is a rational number and \(x = 0.d_1d_2d_3\ldots\), then the decimal digits eventually start repeating. That is, there are positive integers \(N\) and \(P\), such that for all \(n \geq N\), \(d_n = d_{n+P}\).

Let \(x = \nicefrac{p}{q}\) for positive integers \(p\) and \(q\). Let us suppose \(x\) is a number with a unique representation, as otherwise we have seen above that both its representations are repeating.
To compute the first digit we take \(10p\) and divide by \(q\). The quotient is the first digit \(d_1\) and the remainder \(r\) is some integer between 0 and \(q-1\). That is, \(d_1\) is the largest integer such that \(d_1 q \leq 10p\) and then \(r = 10p - d_1q\).

The next digit is computed by dividing \(10r\) by \(q\), and so on. We notice that at each step there are at most \(q\) possible remainders and hence at some point the process must start repeating. In fact we see that \(P\) is at most \(q\).

The converse of the proposition is also true and is left as an exercise.

The number \(x = 0.101001000100001000001\ldots\) is irrational. That is, the digits are \(n\) zeros, then a one, then \(n+1\) zeros, then a one, and so on and so forth. The fact that \(x\) is irrational follows from the proposition; the digits never start repeating. For every \(P\), if we go far enough, we find a 1 that is followed by at least \(P+1\) zeros.

**Exercises**

What is the decimal representation of \(1\) guaranteed by ? Make sure to show that it does satisfy the condition.

Prove the converse of , that is, if the digits in the decimal representation of \(x\) are eventually repeating, then \(x\) must be rational.

Show that real numbers \(x \in (0,1)\) with nonunique decimal representation are exactly the rational numbers that can be written as \(\frac{m}{10^n}\) for some integers \(m\) and \(n\). In this case show that there exist exactly two representations of \(x\).

Let \(b \geq 2\) be an integer. Define a representation of a real number in \([0,1]\) in terms of base \(b\) rather than base 10 and prove for base \(b\).

Using the previous exercise with \(b=2\) (binary), show that cardinality of \(\{\text{mathbb \{R\}}\}\) is the same as the cardinality of \(\{\text{mathcal \{P\}}\{\text{mathbb \{N\}}\}\}\), obtaining yet another (though related) proof that \(\{\text{mathbb \{R\}}\}\) is uncountable. Hint: Construct two injections, one from \([0,1]\) to \(\{\text{mathcal \{P\}}\{\text{mathbb \{N\}}\}\}\) and one from \(\{\text{mathcal \{P\}}\{\text{mathbb \{N\}}\}\}\) to \([0,1]\). Hint 2: Given a set \(A \subset \text{mathbb \{N\}}\), let the \(n\)th binary digit of \(x\) be 1 if \(n \in A\).

Construct a bijection between \([0,1]\) and \([0,1] \times [0,1]\). Hint: consider even and odd digits, and be careful about the uniqueness of representation.

**Sequences and Series**

**Sequences and limits**

*Note: 2.5 lectures*
Analysis is essentially about taking limits. The most basic type of a limit is a limit of a sequence of real numbers. We have already seen sequences used informally. Let us give the formal definition.

A sequence (of real numbers) is a function \( x : \mathbb{N} \to \mathbb{R} \). Instead of \( x(n) \) we usually denote the \( n \)th element in the sequence by \( x_n \). We use the notation \( \{ x_n \} \), or more precisely \( \{ x_n \}_{n=1}^{\infty} \), to denote a sequence.

A sequence \( \{ x_n \} \) is bounded if there exists a \( B \in \mathbb{R} \) such that \( |x_n| \leq B \) for all \( n \in \mathbb{N} \). In other words, the sequence \( \{ x_n \} \) is bounded whenever the set \( \{ x_n : n \in \mathbb{N} \} \) is bounded.

When we need to give a concrete sequence we often give each term as a formula in terms of \( n \). For example, \( \{ \frac{1}{n} \}_{n=1}^{\infty} \), or simply \( \{ \frac{1}{n} \} \), stands for the sequence \( 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots \). The sequence \( \{ \frac{1}{n} \} \) is a bounded sequence (\( B=1 \) will suffice). On the other hand the sequence \( \{ n \} \) stands for \( 1,2,3,4,\ldots \), and this sequence is not bounded (why?).

While the notation for a sequence is similar to that of a set, the notions are distinct. For example, \( \{ (-1)^n \}_{n=1}^{\infty} \) is the sequence \( -1,1,-1,1,\ldots \), whereas the set of values, the range of the sequence, is just the set \( \{ -1, 1 \} \).

We can write this set as \( \{ (-1)^n : n \in \mathbb{N} \} \). When ambiguity can arise, we use the words sequence or set to distinguish the two concepts.

Another example of a sequence is the so-called constant sequence. That is a sequence \( \{ c \} = c,c,c,c,\ldots \) consisting of a single constant \( c \in \mathbb{R} \) repeating indefinitely.

We now get to the idea of a limit of a sequence. We will see in that the notation below is well defined. That is, if a limit exists, then it is unique. So it makes sense to talk about the limit of a sequence.

A sequence \( \{ x_n \} \) is said to converge to a number \( x \) if for every \( \epsilon > 0 \), there exists an \( N \in \mathbb{N} \) such that \( |x_n - x| < \epsilon \) for all \( n \geq N \). The number \( x \) is said to be the limit of \( \{ x_n \} \). We write \( \lim_{n \to \infty} x_n = x \).

A sequence that converges is said to be convergent. Otherwise, the sequence is said to be divergent.

It is good to know intuitively what a limit means. It means that eventually every number in the sequence is close to the number \( x \). More precisely, we can get arbitrarily close to the limit, provided we go far enough in the sequence. It does not mean we ever reach the limit. It is possible, and quite common, that there is no \( x_n \) in the sequence that equals the limit \( x \). We illustrate the concept in . In the figure we first think of the sequence as a graph, as it is a function of \( \{ \text{mathbb\{N\}\} \} \). Seconded we also plot it as a sequence of labeled points on the real line.

When we write \( \lim x_n = x \) for some real number \( x \), we are saying two things. First, that \( \{ x_n \} \) is convergent, and second that the limit is \( x \).
The above definition is one of the most important definitions in analysis, and it is necessary to understand it perfectly. The key point in the definition is that given \( \epsilon > 0 \), we can find an \( M \). The \( M \) can depend on \( \epsilon \), so we only pick an \( M \) once we know \( \epsilon \). Let us illustrate this concept on a few examples.

The constant sequence \( 1, 1, 1, \ldots \) is convergent and the limit is 1. For every \( \epsilon > 0 \), we pick \( M = 1 \).

Claim: The sequence \( \{ \frac{1}{n} \} \) is convergent and \( \lim_{n \to \infty} \frac{1}{n} = 0 \). Proof: Given an \( \epsilon > 0 \), we find an \( M \) such that \( 0 < \frac{1}{M} < \epsilon \). Then for all \( n \geq M \) we have

\[
\left| \frac{1}{n} \right| = \frac{1}{n} \leq \frac{1}{M} < \epsilon.
\]

The sequence \( \{ (-1)^n \} \) is divergent. Proof: If there were a limit \( x \), then for \( \epsilon = \frac{1}{2} \) we expect an \( M \) that satisfies the definition. Suppose such an \( M \) exists, then for an even \( n \geq M \) we compute

\[
\frac{1}{2} > \left| (-1)^n - x \right| = \left| 1 - x \right| \quad \text{and} \quad \frac{1}{2} > \left| (-1)^{n+1} - x \right| = \left| -1 - x \right|.
\]

But

\[
2 = \left| (1 - x) - (-1 -x) \right| < \frac{1}{2} + \frac{1}{2} = 1,
\]

and that is a contradiction.

[prop:limisunique] A convergent sequence has a unique limit. The proof of this proposition exhibits a useful technique in analysis. Many proofs follow the same general scheme. We want to show a certain quantity is zero. We write the quantity using the triangle inequality as two quantities, and we estimate each one by arbitrarily small numbers.

Suppose the sequence \( \{ x_n \} \) has the limit \( x \) and the limit \( y \). Take an arbitrary \( \epsilon > 0 \). From the definition find an \( M_1 \) such that for all \( n \geq M_1 \), \( \left| x_n - x \right| < \frac{\epsilon}{2} \). Similarly find an \( M_2 \) such that for all \( n \geq M_2 \) we have \( \left| x_n - y \right| < \frac{\epsilon}{2} \). Take \( M := \max \{ M_1, M_2 \} \). For \( n \geq M \) we have

\[
\begin{split}
\left| y - x \right| &= \left| x_n - x + (x_n - x) + (y - x) \right| \\
&\leq \left| x_n - x \right| + \left| x_n - y \right| + \left| y - x \right| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{split}
\]

As \( \left| y - x \right| < \epsilon \) for all \( \epsilon > 0 \), then \( \left| y - x \right| = 0 \) and \( y = x \). Hence the limit (if it exists) is unique.

A convergent sequence \( \{ x_n \} \) is bounded. Suppose \( \{ x_n \} \) converges to \( x \). Thus there exists an \( M \) such that for all \( n \geq M \) we have \( \left| x_n - x \right| < \epsilon \). Let \( B_1 := \max \{ \left| x_n - x \right| \} \) and note that for \( n \geq M \) we have

\[
\begin{split}
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon.
\end{split}
\]

The set \( \{ x_n \} \) is bounded. Suppose \( \{ x_n \} \) converges to \( x \). Thus there exists an \( M \) such that for all \( n \geq M \) we have \( \left| x_n - x \right| < \epsilon \). Let \( B_1 := \max \{ \left| x_n - x \right| \} \) and note that for \( n \geq M \) we have

\[
\begin{split}
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon.
\end{split}
\]

The set \( \{ x_n \} \) is bounded. Suppose \( \{ x_n \} \) converges to \( x \). Thus there exists an \( M \) such that for all \( n \geq M \) we have \( \left| x_n - x \right| < \epsilon \). Let \( B_1 := \max \{ \left| x_n - x \right| \} \) and note that for \( n \geq M \) we have

\[
\begin{split}
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon.
\end{split}
\]

The set \( \{ x_n \} \) is bounded. Suppose \( \{ x_n \} \) converges to \( x \). Thus there exists an \( M \) such that for all \( n \geq M \) we have \( \left| x_n - x \right| < \epsilon \). Let \( B_1 := \max \{ \left| x_n - x \right| \} \) and note that for \( n \geq M \) we have

\[
\begin{split}
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon.
\end{split}
\]

The set \( \{ x_n \} \) is bounded. Suppose \( \{ x_n \} \) converges to \( x \). Thus there exists an \( M \) such that for all \( n \geq M \) we have \( \left| x_n - x \right| < \epsilon \). Let \( B_1 := \max \{ \left| x_n - x \right| \} \) and note that for \( n \geq M \) we have

\[
\begin{split}
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon.
\end{split}
\]

The set \( \{ x_n \} \) is bounded. Suppose \( \{ x_n \} \) converges to \( x \). Thus there exists an \( M \) such that for all \( n \geq M \) we have \( \left| x_n - x \right| < \epsilon \). Let \( B_1 := \max \{ \left| x_n - x \right| \} \) and note that for \( n \geq M \) we have

\[
\begin{split}
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon.
\end{split}
\]

The set \( \{ x_n \} \) is bounded. Suppose \( \{ x_n \} \) converges to \( x \). Thus there exists an \( M \) such that for all \( n \geq M \) we have \( \left| x_n - x \right| < \epsilon \). Let \( B_1 := \max \{ \left| x_n - x \right| \} \) and note that for \( n \geq M \) we have

\[
\begin{split}
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon.
\end{split}
\]

The set \( \{ x_n \} \) is bounded. Suppose \( \{ x_n \} \) converges to \( x \). Thus there exists an \( M \) such that for all \( n \geq M \) we have \( \left| x_n - x \right| < \epsilon \). Let \( B_1 := \max \{ \left| x_n - x \right| \} \) and note that for \( n \geq M \) we have

\[
\begin{split}
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon.
\end{split}
\]

The set \( \{ x_n \} \) is bounded. Suppose \( \{ x_n \} \) converges to \( x \). Thus there exists an \( M \) such that for all \( n \geq M \) we have \( \left| x_n - x \right| < \epsilon \). Let \( B_1 := \max \{ \left| x_n - x \right| \} \) and note that for \( n \geq M \) we have

\[
\begin{split}
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon.
\end{split}
\]

The set \( \{ x_n \} \) is bounded. Suppose \( \{ x_n \} \) converges to \( x \). Thus there exists an \( M \) such that for all \( n \geq M \) we have \( \left| x_n - x \right| < \epsilon \). Let \( B_1 := \max \{ \left| x_n - x \right| \} \) and note that for \( n \geq M \) we have

\[
\begin{split}
\left| x_n - x \right| &< \epsilon \\
\left| x_n - x \right| &< \epsilon.
\end{split}
\]
Let us show the sequence \(\left\{ \frac{n^2+1}{n^2+n} \right\}\) converges and \(\lim_{n\to\infty} \frac{n^2+1}{n^2+n} = 1\).

Given \(\epsilon > 0\), find \(M \in \mathbb{N}\) such that \(\frac{n^2+1}{n^2+n} < \epsilon\). Then for any \(n \geq M\) we have \(\abs{\frac{n^2+1}{n^2+n} - 1} = \frac{n-1}{n^2+n} \leq \frac{1}{n+1} = \frac{1}{M+1} < \epsilon\). Therefore, \(\lim \frac{n^2+1}{n^2+n} = 1\).

**Monotone sequences**

The simplest type of a sequence is a monotone sequence. Checking that a monotone sequence converges is as easy as checking that it is bounded. It is also easy to find the limit for a convergent monotone sequence, provided we can find the supremum or infimum of a countable set of numbers.

A sequence \(\left\{ x_n \right\}\) is **monotone increasing** if \(x_n \leq x_{n+1}\) for all \(n \in \mathbb{N}\). A sequence \(\left\{ x_n \right\}\) is **monotone decreasing** if \(x_n \geq x_{n+1}\) for all \(n \in \mathbb{N}\). If a sequence is either monotone increasing or monotone decreasing, we can simply say the sequence is **monotone**. Some authors also use the word **monotonic**.

For example, \(\left\{ \frac{1}{n} \right\}\) is monotone decreasing, the constant sequence \(\left\{ 1 \right\}\) is both monotone increasing and monotone decreasing, and \(\left\{ (-1)^n \right\}\) is not monotone. First few terms of a sample monotone increasing sequence are shown in .

A sequence \(\left\{ x_n \right\}\) is **monotone increasing** if \(x_n < x_{n+1}\) for all \(n \in \mathbb{N}\). A sequence \(\left\{ x_n \right\}\) is **monotone decreasing** if \(x_n > x_{n+1}\) for all \(n \in \mathbb{N}\). If a sequence is either monotone increasing or monotone decreasing, we can simply say the sequence is **monotone**. Some authors also use the word **monotonic**.

For example, \(\left\{ \frac{1}{n} \right\}\) is monotone decreasing, the constant sequence \(\left\{ 1 \right\}\) is both monotone increasing and monotone decreasing, and \(\left\{ (-1)^n \right\}\) is not monotone. First few terms of a sample monotone increasing sequence are shown in .

A monotone sequence \(\left\{ x_n \right\}\) is bounded if and only if it is convergent. Furthermore, if \(\left\{ x_n \right\}\) is monotone increasing and bounded, then \(\lim_{n\to\infty} x_n = \sup \{ x_n : n \in \mathbb{N} \}\). If \(\left\{ x_n \right\}\) is monotone decreasing and bounded, then \(\lim_{n\to\infty} x_n = \inf \{ x_n : n \in \mathbb{N} \}\).

Let us suppose the sequence is monotone increasing. Suppose the sequence is bounded, so there exists a \(B\) such that \(\forall n, x_n \leq B\) for all \(n\), that is the set \(\{ x_n : n \in \mathbb{N} \}\) is bounded from above. Let \(x = \sup \{ x_n : n \in \mathbb{N} \}\). Let \(\epsilon > 0\) be arbitrary. As \(x\) is the supremum, then there must be at least one \(M \in \mathbb{N}\) such that \(x_{M} > x - \epsilon\) (because \(x\) is the supremum). As \(\left\{ x_n \right\}\) is monotone increasing, then it is easy to see (by ) that \(x_{M} \geq x_{M+1}\) for all \(n \geq M\). Hence \(\abs{x_{n} - x_{M}} = x_{n} - x_{M} \leq \epsilon\). Therefore the sequence converges to \(x\). We already know that a convergent sequence is bounded, which completes the other direction of the implication.

The proof for monotone decreasing sequences is left as an exercise.

Take the sequence \(\left\{ \frac{1}{\sqrt{n}} \right\}\). First \(\frac{1}{\sqrt{n}} > 0\) for all \(n \in \mathbb{N}\), and hence the sequence is bounded from below. Let us show that it is monotone decreasing. We start with \(\frac{1}{\sqrt{n+1}} > \frac{1}{\sqrt{n}}\) (why is that true?). From this inequality we obtain
\[
\frac{1}{\sqrt{n+1}} \leq \frac{1}{\sqrt{n}} .
\]
So the sequence is monotone decreasing and bounded from below (hence bounded). We apply the theorem to note that the sequence is convergent and in fact \(\lim_{n \to \infty} \frac{1}{\sqrt{n}} = \inf \{ \frac{1}{\sqrt{n}} : n \in \mathbb{N} \} \). We already know that the infimum is greater than or equal to 0, as 0 is a lower bound. Take a number \(b \geq 0\) such that \(b \leq \frac{1}{\sqrt{n}}\) for all \(n\). We square both sides to obtain \(b^2 \leq \frac{1}{n}\) for all \(n \in \mathbb{N}\). We have seen before that this implies that \(b^2 \leq 0\) (a consequence of the ). As we also have \(b^2 \geq 0\), then \(b^2 = 0\) and so \(b = 0\). Hence \(b = 0\) is the greatest lower bound, and \(\lim \frac{1}{\sqrt{n}} = 0\).

A word of caution: We must show that a monotone sequence is bounded in order to use . For example, the sequence \(\{ 1 + \nicefrac{1}{2} + \cdots + \nicefrac{1}{n} \}\) is a monotone increasing sequence that grows very slowly. We will see, once we get to series, that this sequence has no upper bound and so does not converge. It is not at all obvious that this sequence has no upper bound.

A common example of where monotone sequences arise is the following proposition. The proof is left as an exercise.

\[
\text{\begin{prop:supinfseq} Let } (S \subset \mathbb{R}) \text{ be a nonempty bounded set. Then there exist monotone sequences } (x_n) \text{ and } (y_n) \text{ such that } \sup S = \lim_{n \to \infty} x_n \text{ and } \inf S = \lim_{n \to \infty} y_n \text{.} \end{prop:supinfseq}}
\]

**Tail of a sequence**

For a sequence \(\{x_n\}\), the \((K)\)-tail (where \(K \in \mathbb{N}\)) or just the tail of the sequence is the sequence starting at \(K+1\), usually written as \(\{x_{n+K}\}_{n=1}^{\infty}\) or \(\{x_{n+K}\}_{n=K+1}^{\infty}\).

The main result about the tail of a sequence is the following proposition.

Let \(\{x_n\}_{n=1}^{\infty}\) be a sequence. Then the following statements are equivalent:

i. [prop:ktail:i] The sequence \(\{x_n\}_{n=1}^{\infty}\) converges.

ii. [prop:ktail:ii] The \((K)\)-tail \(\{x_{n+K}\}_{n=1}^{\infty}\) converges for all \(K \in \mathbb{N}\).

iii. [prop:ktail:iii] The \((K)\)-tail \(\{x_{n+K}\}_{n=1}^{\infty}\) converges for some \(K \in \mathbb{N}\).

Furthermore, if any (and hence all) of the limits exist, then for any \(K \in \mathbb{N}\), \(\lim_{n \to \infty} x_n = \lim_{n \to \infty} x_{n+K}\).

It is clear that [prop:ktail:ii] implies [prop:ktail:i]. We will therefore show first that [prop:ktail:ii] implies [prop:ktail:iii], and then we will show that [prop:ktail:iii] implies [prop:ktail:i]. In the process we will also show that the limits are equal.

Let us start with [prop:ktail:ii] implies [prop:ktail:iii]. Suppose \(\{x_n\}\) converges to some \(\lim_{n \to \mathbb{R}} x_n = x \in \mathbb{R}\). Let \(K \in \mathbb{N}\) be arbitrary. Define \(y_n := x_{n+K}\), we wish to show that \(\{y_n\}\) converges to \(x\). That is, given an \(\epsilon > 0\), there exists an \(M \in \mathbb{N}\) such that \(\left| y_n - x \right| < \epsilon\) for all
\( n \geq M \). Note that \( (n \geq M) \) implies \( (n+K \geq M) \). Therefore, it is true that for all \( (n \geq M) \) we have that \( \| x - y_n \| = \| x - x_{n+K} \| < \epsilon \). Therefore \( \{ y_n \} \) converges to \( (x) \).

Let us move to \([\text{prop:ktail:iii}]\) implies \([\text{prop:ktail:i}]\). Let \( (K \in \mathbb{N}) \) be given, define \( (y_n := x_{n+K}) \), and suppose that \( (y_n) \) converges \( (x) \). That is, given an \( \epsilon > 0 \), there exists an \( M' \in \mathbb{N} \) such that \( \| x - y_{n-K} \| < \epsilon \) for all \( n \geq M' \). Let \( M := M' + K \). Then \( (n \geq M) \) implies \( (n - K \geq M') \). Thus, whenever \( n \geq M \) we have \( \| x - x_n \| = \| x - y_{n-K} \| < \epsilon \). Therefore \( (x_n) \) converges to \( (x) \).

Essentially, the limit does not care about how the sequence begins, it only cares about the tail of the sequence. That is, the beginning of the sequence may be arbitrary.

For example, the sequence defined by \( (x_n := \frac{n}{n^2 + 16}) \) is decreasing if we start at \( (n = 4) \) (it is increasing before). That is, \( (x_n) = \{ 17, 11, 3, 10, 9, 8, 5, 4, 3, 2, 1 \} \). Here, \( \frac{1}{17} < \frac{1}{10} < \frac{3}{25} < \frac{1}{8} > \frac{5}{41} > \frac{3}{26} > \frac{7}{65} > \frac{1}{10} > \frac{9}{97} > \frac{5}{58} > \ldots \). Thus, the 3-tail is monotone and bounded below by zero, it is convergent, and therefore the sequence is convergent.

**Subsequences**

A very useful concept related to sequences is that of a subsequence. A subsequence of \( (x_n) \) is a sequence that contains only some of the numbers from \( (x_n) \) in the same order.

Let \( (x_n) \) be a sequence. Let \( (n_i) \) be a strictly increasing sequence of natural numbers (that is \( n_1 < n_2 < n_3 < \ldots \)). The sequence \( (x_{n_i}) \) is called a subsequence of \( (x_n) \).

For example, take the sequence \( (\frac{1}{n}) \). The sequence \( (\frac{1}{3n}) \) is a subsequence. To see how these two sequences fit in the definition, take \( n_i := 3i \). The numbers in the subsequence must come from the original sequence, so \( (1, 0, \frac{1}{3}, 0, \frac{1}{6}, \frac{1}{8}, \ldots) \) is not a subsequence of \( (\frac{1}{n}) \). Similarly order must be preserved, so the sequence \( (1, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \ldots) \) is not a subsequence of \( (\frac{1}{n}) \).

A tail of a sequence is one special type of a subsequence. For an arbitrary subsequence, we have the following proposition about convergence.

\( [\text{prop:seqtosubseq}] \) If \( (x_n) \) is a convergent sequence, then any subsequence \( (x_{n_i}) \) is also convergent and
\[
\lim_{n \to \infty} x_n = \lim_{i \to \infty} x_{n_i}.
\]

Suppose \( \lim_{n \to \infty} x_n = x \). That means that for every \( \epsilon > 0 \) we have an \( M \in \mathbb{N} \) such that for all \( n \geq M \) \( \| x_n - x \| < \epsilon \). It is not hard to prove (do it!) by that \( (n \geq M) \) \( \| x_n - x \| < \epsilon \). Hence \( i \)
\[\geq M\] implies \(n_i \geq M\). Thus, for all \((i \geq M)\) we have \[\left| x_{n_i} - x \right| < \varepsilon,\] and we are done.

Existence of a convergent subsequence does not imply convergence of the sequence itself. Take the sequence \(0,1,0,1,0,1,\ldots\). That is, \(x_n = 0\) if \(n\) is odd, and \(x_n = 1\) if \(n\) is even. The sequence \(\{x_n\}\) is divergent, however, the subsequence \(\{x_{2n}\}\) converges to 1 and the subsequence \(\{x_{2n+1}\}\) converges to 0. Compare .

**Exercises**

In the following exercises, feel free to use what you know from calculus to find the limit, if it exists. But you must prove that you found the correct limit, or prove that the series is divergent.

Is the sequence \(\{3n\}\) bounded? Prove or disprove.

Is the sequence \(\{n\}\) convergent? If so, what is the limit?

Is the sequence \(\left\lfloor \frac{(-1)^n}{2n} \right\rfloor\) convergent? If so, what is the limit?

Is the sequence \(\left\lfloor 2^{-n} \right\rfloor\) convergent? If so, what is the limit?

Is the sequence \(\left\lfloor \frac{n}{n+1} \right\rfloor\) convergent? If so, what is the limit?

Is the sequence \(\left\lfloor \frac{n}{n^2+1} \right\rfloor\) convergent? If so, what is the limit?

[exercise:absconv] Let \(\{x_n\}\) be a sequence.

a. Show that \(\lim x_n = 0\) (that is, the limit exists and is zero) if and only if \(\lim \left| x_n \right| = 0\).

b. Find an example such that \(\lim \left| x_n \right| = 0\) converges and \(\{x_n\}\) diverges.

Is the sequence \(\left\lfloor \frac{2^n}{n!} \right\rfloor\) convergent? If so, what is the limit?

Show that the sequence \(\left\lfloor \frac{1}{\sqrt[3]{n}} \right\rfloor\) is monotone, bounded, and use to find the limit.

Show that the sequence \(\left\lfloor \frac{n+1}{n} \right\rfloor\) is monotone, bounded, and use to find the limit.

Finish the proof of for monotone decreasing sequences.

Prove .

Let \(\{x_n\}\) be a convergent monotone sequence. Suppose there exists a \(k \in \mathbb{N}\) such that \(\lim_{n \to \infty} x_n = x_k\). Show that \(x_n = x_k\) for all \(n \geq k\).

Find a convergent subsequence of the sequence \(\{(n+1)^n\}\).
Let $\{x_n\}$ be a sequence defined by $x_n := \begin{cases} n & \text{if } n \text{ is odd} , \\ \frac{1}{n} & \text{if } n \text{ is even} . \end{cases}$

a. Is the sequence bounded? (prove or disprove)

b. Is there a convergent subsequence? If so, find it.

Let $\{x_n\}$ be a sequence. Suppose there are two convergent subsequences $\{x_{n_i}\}$ and $\{x_{m_i}\}$. Suppose $\lim_{i\to\infty} x_{n_i} = a$ and $\lim_{i\to\infty} x_{m_i} = b$, where $a \not= b$. Prove that $\{x_n\}$ is not convergent, without using $\epsilon$.

Find a sequence $\{x_n\}$ such that for any $y \in \mathbb{R}$, there exists a subsequence $\{x_{n_i}\}$ converging to $y$.

Let $\{x_n\}$ be a sequence. Suppose for any $\epsilon > 0$, there is an $M$ such that for all $n \geq M$, $|x_n - x| \leq \epsilon$. Show that $\lim\, x_n = x$.

Let $\{x_n\}$ be a sequence and $x \in \mathbb{R}$ such that there exists a $k \in \mathbb{N}$ such that for all $n \geq k$, $x_n = x$. Prove that $\{x_n\}$ converges to $x$.

Let $\{x_n\}$ be a sequence and define a sequence $\{y_n\}$ by $y_{2k} := x_{k^2}$ and $y_{2k-1} = x_k$ for all $k \in \mathbb{N}$. Prove that $\{x_n\}$ converges if and only if $\{y_n\}$ converges. Furthermore, prove that if they converge, then $\lim\, x_n = \lim\, y_n$.

Show that the 3-tail of the sequence defined by $x_n := \frac{n}{n^2+16}$ is monotone decreasing. Hint: Suppose $n \geq m \geq 4$ and consider the numerator of the expression $x_n - x_m$.

Suppose that $\{x_n\}$ is a sequence such that the subsequences $\{x_{2n}\}, \{x_{2n-1}\}, \{x_{3n}\}$ all converge. Show that $\{x_n\}$ is convergent.

---

**Facts about limits of sequences**

*Note: 2–2.5 lectures, recursively defined sequences can safely be skipped*

In this section we go over some basic results about the limits of sequences. We start by looking at how sequences interact with inequalities.

**Limits and inequalities**

A basic lemma about limits and inequalities is the so-called squeeze lemma. It allows us to show convergence of sequences in difficult cases if we find two other simpler convergent sequences that “squeeze” the original sequence.
Let \(\{a_n\}\), \(\{b_n\}\), and \(\{x_n\}\) be sequences such that
\[a_n \leq x_n \leq b_n \quad \text{for all } n \in \mathbb{N}.\]
Suppose \(\{a_n\}\) and \(\{b_n\}\) converge and
\[\lim_{n \to \infty} a_n = \lim_{n \to \infty} b_n.\]
Then \(\{x_n\}\) converges and
\[\lim_{n \to \infty} x_n = \lim_{n \to \infty} a_n = \lim_{n \to \infty} b_n.\]

The intuitive idea of the proof is illustrated in . If \(x\) is the limit of \(\{a_n\}\) and \(\{b_n\}\), then if they are both within
\(\frac{\epsilon}{3}\) of \(x\), then the distance between \(\{a_n\}\) and \(\{b_n\}\) is at most \(\frac{2\epsilon}{3}\). As
\(\{x_n\}\) is between \(\{a_n\}\) and \(\{b_n\}\) it is at most \(\frac{2\epsilon}{3}\) from \(\{a_n\}\). Since \(\{a_n\}\) is at most
\(\frac{2\epsilon}{3}\) away from \(\{x_n\}\), then \(\{x_n\}\) must be at most \(\epsilon\) away from \(\{x\}\). Let us follow through on
this intuition rigorously.

Let \(x := \lim\ a_n = \lim\ b_n\). Let \(\epsilon > 0\) be given.

Find an \(M_1\) such that for all \(n \geq M_1\) we have \(x - \epsilon < a_n < \epsilon\) and
an \(M_2\) such that for all \(n \geq M_2\) we have \(x - \epsilon < b_n < \epsilon\). Set \(M := \max\ \{M_1, M_2\}\). Suppose \(n \geq M\). We compute
\[|x_n - x| = |x_n - x + a_n - a_n| \leq |x_n - a_n| + |a_n - x| < \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.\]

And we are done.

One application of the is to compute limits of sequences using limits that are already known. For example, suppose we have the sequence \(\{\frac{1}{n}\}\) for all \(n \in \mathbb{N}\). We already know \(\lim_{n \to \infty} \frac{1}{n} = 0\). Hence, using the
constant sequence \(\{1\}\) and the sequence \(\{\frac{1}{n}\}\) in the squeeze lemma, we conclude
\[\lim_{n \to \infty} \frac{1}{n} = 0.\]

Limits also preserve inequalities.

Let \(x := \lim\ x_n\) and \(y := \lim\ y_n\). Let \(\epsilon > 0\) be given. Find an \(M\) such that for all \(n \geq M\) we have
\[|x_n - x| < \epsilon\] and
\[|y_n - y| < \epsilon.\]

An easy corollary is proved using constant sequences in . The proof is left as an exercise.

\[\limandineq:lemma\]
i. Let \( \{ x_n \} \) be a convergent sequence such that \( x_n \geq 0 \), then 
\[
\lim_{n \to \infty} x_n \geq 0.
\]

ii. Let \( (a, b) \in \mathbb{R} \) and let \( \{ x_n \} \) be a convergent sequence such that 
\[
[a \leq x_n \leq b] \quad \text{for all } n \in \mathbb{N}.
\]
Then 
\[
[a \leq \lim_{n \to \infty} x_n \leq b].
\]

In and we cannot simply replace all the non-strict inequalities with strict inequalities. For example, let \( x_n := \nfrac{-1}{n} \) and \( y_n := \nfrac{1}{n} \). Then \( x_n < y_n \), \( x_n < 0 \), and \( y_n > 0 \) for all \( n \). However, these inequalities are not preserved by the limit operation as we have \( \lim x_n = \lim y_n = 0 \). The moral of this example is that strict inequalities may become non-strict inequalities when limits are applied; if we know \( x_n < y_n \) for all \( n \), we may only conclude 
\[
\lim_{n \to \infty} x_n \leq \lim_{n \to \infty} y_n.
\]
This issue is a common source of errors.

**Continuity of algebraic operations**

Limits interact nicely with algebraic operations.

[prop:contalg] Let \( \{ x_n \} \) and \( \{ y_n \} \) be convergent sequences.

i. [prop:contalg:i] The sequence \( \{ z_n \} \), where \( z_n := x_n + y_n \), converges and 
\[
\lim_{n \to \infty} (x_n + y_n) = \lim_{n \to \infty} z_n = \lim_{n \to \infty} x_n + \lim_{n \to \infty} y_n.
\]

ii. [prop:contalg:ii] The sequence \( \{ z_n \} \), where \( z_n := x_n - y_n \), converges and 
\[
\lim_{n \to \infty} (x_n - y_n) = \lim_{n \to \infty} z_n = \lim_{n \to \infty} x_n - \lim_{n \to \infty} y_n.
\]

iii. [prop:contalg:iii] The sequence \( \{ z_n \} \), where \( z_n := x_n y_n \), converges and 
\[
\lim_{n \to \infty} (x_n y_n) = \lim_{n \to \infty} z_n = (\lim_{n \to \infty} x_n)(\lim_{n \to \infty} y_n).
\]

iv. [prop:contalg:iv] If \( \{x_n\} \not= 0 \) and \( \{y_n\} \not= 0 \) for all \( n \in \{\mathbb{N}\} \), then the sequence \( \{ z_n \} \), where \( z_n := \nfrac{x_n}{y_n} \), converges and 
\[
\lim_{n \to \infty} \frac{x_n}{y_n} = \lim_{n \to \infty} z_n = \frac{\lim_{n \to \infty} x_n}{\lim_{n \to \infty} y_n}.
\]

Let us start with [prop:contalg:i]. Suppose \( \{ x_n \} \) and \( \{ y_n \} \) are convergent sequences and write \( (z_n := x_n + y_n) \). Let \( x := \lim x_n, y := \lim y_n \), and \( z := x+y \).

Let \( \epsilon > 0 \) be given. Find an \( (M_1) \) such that for all \( n \geq M_1 \) we have \( \|x_n - x\| < \frac{\epsilon}{2} \). Find an \( (M_2) \) such that for all \( n \geq M_2 \) we have \( \|y_n - y\| < \frac{\epsilon}{2} \). Take \( M := \max \{ M_1, M_2 \} \). For all \( n \geq M \) we have \( \|z_n - z\| < \frac{\epsilon}{2} \). Therefore [prop:contalg:i] is proved. Proof of [prop:contalg:ii] is almost identical and is left as an exercise.

Let us tackle [prop:contalg:iii]. Suppose again that \( \{ x_n \} \) and \( \{ y_n \} \) are convergent sequences and write \( (z_n := x_n y_n) \). Let \( x := \lim x_n, y := \lim y_n \), and \( z := xy \).
Let \( \epsilon > 0 \) be given. As \( \{ \epsilon \} \) is convergent, it is bounded. Therefore, find a \( \lambda > 0 \) such that \( \forall \lambda \in \{ \epsilon \} \), \( \lambda \geq \lambda \). Find an \( M \) such that for all \( \forall \lambda \in \{ \epsilon \} \) we have \( \forall \lambda \in \{ \epsilon \} \). Find an \( M \) such that for all \( \forall \lambda \in \{ \epsilon \} \) we have \( \forall \lambda \in \{ \epsilon \} \). Take \( M = \max \{ \{ \epsilon \}, M \} \). For all \( \forall \lambda \in \{ \epsilon \} \) we have \( \forall \lambda \in \{ \epsilon \} \).

\[
\begin{align*}
\begin{split} 
\left| z_n - z \right| &= \left| (x_n y_n) - (xy) \right| \\
&= \left| x_n y_n - (x_n + x_n - x_n) y \right| \\
&= \left| x_n (y_n - y) + (x_n - x) y \right| \\
&\leq \left| x_n \right| \left| y_n - y \right| + \left| x_n - x \right| \left| y \right|
\end{split}
\end{align*}
\]

Finally let us tackle [prop:contalg:iv]. Instead of proving [prop:contalg:iv] directly, we prove the following simpler claim:

Claim: If \( \{ y_n \} \) is a convergent sequence such that \( \lim \ y_n \neq 0 \) and \( y_n \neq 0 \) for all \( n \in \mathbb{N} \), then \( \lim \ \frac{1}{y_n} = \frac{1}{\lim \ y_n} \).

Once the claim is proved, we take the sequence \( \{ \frac{1}{y_n} \} \), multiply it by the sequence \( \{ \frac{1}{y_n} \} \) and apply item [prop:contalg:iii].

Proof of claim: Let \( \{ \epsilon > 0 \} \) be given. Let \( \{ \epsilon \} = \lim \ y_n \). Find an \( M \) such that for all \( \forall \lambda \in \{ \epsilon \} \) we have \( \forall \lambda \in \{ \epsilon \} \). Notice that we can make this claim as the right hand side is positive because \( \forall \lambda \in \{ \epsilon \} \). Therefore for all \( \forall \lambda \in \{ \epsilon \} \) we have \( \forall \lambda \in \{ \epsilon \} \). So \( \forall \lambda \in \{ \epsilon \} \), \( \forall \lambda \in \{ \epsilon \} \). Subtracting \( \forall \lambda \in \{ \epsilon \} \), \( \forall \lambda \in \{ \epsilon \} \) from both sides we obtain \( \forall \lambda \in \{ \epsilon \} \), \( \forall \lambda \in \{ \epsilon \} \) or in other words, \( \forall \lambda \in \{ \epsilon \} \).

Now we finish the proof of the claim:

By plugging in constant sequences, we get several easy corollaries. If \( \forall \lambda \in \{ \mathbb{N} \} \) and \( \{ \ x_n \} \) is a convergent sequence, then for example \( \lim \ n \to \infty \ c \ x_n = c \ \lim \ n \to \infty \ x_n \) \( \forall \quad x \quad n \) \( \forall \quad c \). Similarly with constant subtraction and division.

As we can take limits past multiplication we can show (exercise) that \( \lim \ n \to \infty \ k = \{ \lim \ n \to \infty \ k \} \) for all \( k \in \{ \mathbb{N} \} \). That is, we can take limits past powers. Let us see if we can do the same with roots.

Let \( \{ \ x_n \} \) be a convergent sequence such that \( \forall \lambda \in \{ \mathbb{N} \} \). Then \( \lim \ n \to \infty \sqrt{x_n} = \sqrt{\lim \ n \to \infty \ x_n} \).
Of course to even make this statement, we need to apply to show that \( \lim x_n \geq 0 \), so that we can take the square root without worry.

Let \( \{ x_n \} \) be a convergent sequence and let \( x := \lim x_n \).

First suppose \( x=0 \). Let \( \varepsilon > 0 \) be given. Then there is an \( M \) such that for all \( n \geq M \) we have \( |x_n| < \varepsilon^2 \), or in other words \( |\sqrt{x_n}| < \varepsilon \). Hence \[ |\sqrt{x_n} - \sqrt{x}| = \sqrt{|x_n - x|} < \varepsilon. \]

Now suppose \( x > 0 \) (and hence \( |\sqrt{x}| > 0 \)). \( \begin{align*}
|\sqrt{x_n} - \sqrt{x}| &= \left| \frac{x_n - x}{\sqrt{x_n} + \sqrt{x}} \right| \\
&= \frac{1}{\sqrt{x_n} + \sqrt{x}} |x_n - x| \\
&\leq \frac{1}{\sqrt{x}} |x_n - x|.
\end{align*} \] We leave the rest of the proof to the reader.

A similar proof works for the \( k \)-th root. That is, we also obtain \( \lim x_n^{1/k} = (\lim x_n)^{1/k} \). We leave this to the reader as a challenging exercise.

We may also want to take the limit past the absolute value sign. The converse of this proposition is not true, see part b).

If \( \{ x_n \} \) is a convergent sequence, then \( \{ |x_n| \} \) is convergent and \( \lim_{n\to\infty} |x_n| = |\lim_{n\to\infty} x_n| \).

We simply note the reverse triangle inequality \[ |\sqrt{|x_n|} - \sqrt{|x|}| \leq |x_n - x|. \] Hence if \( |\left|\sqrt{x_n} - \sqrt{x}\right|\) can be made arbitrarily small, so can \( |\sqrt{|x_n|} - \sqrt{|x|}|). Details are left to the reader.

Let us see an example putting the above propositions together. Since we know that \( \lim \sqrt{\frac{1}{n}} = 0 \), then
\[ \lim_{n\to\infty} |\sqrt{1 + \frac{1}{n}} - \frac{100}{n^2}| = 1. \] That is, the limit on the left hand side exists because the right hand side exists. You really should read the above equality from right to left.

**Recursively defined sequences**

Now that we know we can interchange limits and algebraic operations, we can compute the limits of many sequences. One such class are recursively defined sequences, that is, sequences where the next number in the sequence computed using a formula from a fixed number of preceding elements in the sequence.

Let \( \{ x_n \} \) be defined by \( x_1 := 2 \) and \( x_{n+1} := x_n - \frac{x_n^2 - 2}{2x_n} \). We must first find out if this sequence is well defined; we must show we never divide by zero. Then we must find out if the sequence converges. Only then can we attempt to find the limit.

First let us prove \( \{ x_n \} \) exists and \( \{ x_n > 0 \} \) for all \( n \) (so the sequence is well defined and bounded below). Let us show this by . We know that \( (x_1 = 2 > 0) \). For the induction step, suppose \( \{ x_n > 0 \} \). Then \( x_{n+1} = x_n - \frac{x_n^2 - 2}{2x_n} \).
\[ \frac{x_n^2 - 2}{2x_n} = \frac{2x_n^2 - x_n^2 + 2}{2x_n} = \frac{x_n^2 + 2}{2x_n} .\]

If \( x_n > 0 \), then \( (x_n > 0) \) and hence \( (x_{n+1} > 0) \).

Next let us show that the sequence is monotone decreasing. If we show that \( (x_n^2 - 2 \geq 0) \) for all \( n \), then \( (x_{n+1} \leq x_n) \) for all \( n \). Obviously \( (x_1^2 - 2 = 4 - 2 = 2 > 0) \). For an arbitrary \( n \) we have \( (x_{n+1}^2 - 2 = \frac{\left( x_n^2 + 2 \right)^2}{4x_n^2} - 2 = \frac{x_n^4 + 4x_n^2 + 4 - 8x_n^2}{4x_n^2} = \frac{x_n^4 - 4x_n^2 + 4}{4x_n^2} = \frac{(x_n^2 - 2)^2}{4x_n^2} ) \). Since any number squared is nonnegative, we have that \( (x_{n+1}^2 - 2 \geq 0) \) for all \( n \). Therefore, \( (\{ x_n \}) \) is monotone decreasing and bounded \( (x_n > 0) \) for all \( n \), and the limit exists. It remains to find the limit.

Let us write \( 2x_n x_{n+1} = x_n^2 + 2 \). Since \( (\{ x_n \}) \) is the 1-tail of \( (\{ x_n \}) \), it converges to the same limit. Let us define \( x := \lim x_n \). We take the limit of both sides to obtain \( 2x^2 = x^2 + 2 \). As \( x_n > 0 \) for all \( n \) we get \( x = \sqrt{2} \), and therefore \( (x = \sqrt{2}) \).

You may have seen the above sequence before. It is the Newton’s method \(^8\) for finding the square root of 2. This method comes up very often in practice and converges very rapidly. Notice that we have used the fact that \( (x_1^2 - 2 > 0) \), although it was not strictly needed to show convergence by considering a tail of the sequence. In fact the sequence converges as long as \( (x_1 \not= 0) \), although with a negative \( (x_1) \) we would arrive at \( (x = -\sqrt{2}) \). By replacing the 2 in the numerator we obtain the square root of any positive number. These statements are left as an exercise.

You should, however, be careful. Before taking any limits, you must make sure the sequence converges. Let us see an example.

Suppose \( (x_1 := 1) \) and \( (x_{n+1} := x_n^2 + x_n) \). If we blindly assumed that the limit exists (call it \( (x) \)), then we would get the equation \( (x = x^2 + x) \), from which we might conclude \( (x = 0) \). However, it is not hard to show that \( (\{ x_n \}) \) is unbounded and therefore does not converge.

The thing to notice in this example is that the method still works, but it depends on the initial value \( (x_1) \). If we set \( (x_1 := 0) \), then the sequence converges and the limit really is 0. An entire branch of mathematics, called dynamics, deals precisely with these issues.

**Some convergence tests**

It is not always necessary to go back to the definition of convergence to prove that a sequence is convergent. We first give a simple convergence test. The main idea is that \( (\{ x_n \}) \) converges to \( (x) \) if and only if \( (\forall n \lim |x_n - x| = 0) \) converges to zero.

\[ \text{convzero:prop} \text{ Let } (\{ x_n \}) \text{ be a sequence. Suppose there is an } (x \in \{\mathbb{R}\}) \text{ and a convergent sequence } (\{ a_n \}) \text{ such that } \lim \{n/\to/\infty\} a_n = 0 \text{ and } (\forall n |x_n - x| \to |a_n|) \text{ for all } n. \text{ Then } (\{ x_n \}) \text{ converges and } (\lim, x_n = x). \]
Let \( \epsilon > 0 \) be given. Note that \( a_n \geq 0 \) for all \( n \). Find an \( M \in \mathbb{N} \) such that for all \( n \geq M \) we have \( a_n = \left\lvert {a_n - 0} \right\rvert < \epsilon \). Then, for all \( n \geq M \) we have \( \left\lvert {x_n - x} \right\rvert \leq a_n < \epsilon \). \( \qede \)

As the proposition shows, to study when a sequence has a limit is the same as studying when another sequence goes to zero. In general it may be hard to decide if a sequence converges, but for certain sequences there exist easy to apply tests that tell us if the sequence converges or not. Let us see one such test. First let us compute the limit of a very specific sequence.

Let \( c > 0 \).

i. If \( c < 1 \), then \( \lim_{n \to \infty} c^n = 0 \).

ii. If \( c > 1 \), then \( \{ c^n \} \) is unbounded.

First let us suppose \( c > 1 \). We write \( c = 1+r \) for some \( r > 0 \). By (or using the binomial theorem if you know it) we have Bernoulli’s inequality (see also): \( |c^n| = |(1+r)^n| \geq 1+nr \). By the of the real numbers, the sequence \( \{ 1+nr \} \) is unbounded (for any number \( B \), we find an \( n \in \mathbb{N} \) such that \( nr \geq B-1 \)). Therefore \( \{ c^n \} \) is unbounded.

Now let \( c < 1 \). Write \( c = \frac{1}{1+r} \), where \( r > 0 \). Then \( c^n = \frac{1}{(1+r)^n} \leq \frac{1}{1+nr} \leq \frac{1}{nr} \frac{1}{n} \). As \( \{ \frac{1}{n} \} \) converges to zero, so does \( \{ \frac{1}{1+r} \} \frac{1}{n} \). Hence, \( \{ c^n \} \) converges to zero.

If we look at the above proposition, we note that the ratio of the \( (n+1) \)th term and the \( n \)th term is \( c \). We generalize this simple result to a larger class of sequences. The following lemma will come up again once we get to series.

\[ \text{seq:ratiotest} \]

Let \( \{ x_n \} \) be a sequence such that \( x_n \neq 0 \) for all \( n \) and such that the limit \( \lim_{n \to \infty} \frac{|x_{n+1}|}{|x_n|} \) exists.

i. If \( \lim_{n \to \infty} \frac{|x_{n+1}|}{|x_n|} < 1 \), then \( \lim_{n \to \infty} x_n = 0 \).

ii. If \( \lim_{n \to \infty} \frac{|x_{n+1}|}{|x_n|} > 1 \), then \( \lim_{n \to \infty} x_n \) is unbounded (hence diverges).

If \( \lim_{n \to \infty} \frac{|x_{n+1}|}{|x_n|} \) exists, but \( \lim_{n \to \infty} x_n \) is not equal to zero, we cannot make any conclusion based on that information alone. For example, the sequence \( \{ \text{nicefrac} \{ 1 \} \{ n \} \} \) converges to zero, but \( \lim_{n \to \infty} \frac{|x_{n+1}|}{|x_n|} = 1 \). The constant sequence \( \{ 1 \} \) converges to 1, not zero, and also \( \lim_{n \to \infty} \frac{|x_{n+1}|}{|x_n|} = 1 \). The sequence \( \{ (-1)^n \} \) does not converge at all, and \( \lim_{n \to \infty} \frac{|x_{n+1}|}{|x_n|} = 1 \). Finally the sequence \( \{ \ln n \} \) is unbounded, yet again \( \lim_{n \to \infty} \frac{|x_{n+1}|}{|x_n|} = 1 \).

Suppose \( \lim_{n \to \infty} \frac{|x_{n+1}|}{|x_n|} = 1 \). As \( \lim_{n \to \infty} \frac{|x_{n+1}|}{|x_n|} = 1 \), we have that \( |x_{n+1}| \geq |x_n| \). Pick \( r \) such that \( |x_n| \leq r \). We wish to compare the sequence to the sequence \( \{ r^n \} \). The idea is that while the sequence is not going to be less than \( \{ \} \) eventually, it will eventually be less than \( \{ r^n \} \), which is still less than 1. The intuitive idea of the proof is illustrated in.
As \( r - L > 0 \), there exists an \( M \) in \( \mathbb{N} \) such that for all \( n \geq M \) we have \( \left\lvert \frac{x_{n+1}}{x_n} - L \right\rvert < r - L \). Therefore, \( \left\lvert \frac{x_{n+1}}{x_n} \right\rvert < L \). For \( n > M \) (that is for \( n \geq M + 1 \)) we write \( \frac{x_{n+1}}{x_n} = \frac{x_{M+1}}{x_M} \) \( \frac{x_{M+2}}{x_{M+1}} \cdots \frac{x_n}{x_{n-1}} < L \). The sequence \( \{ r^n \} \) converges to zero and hence \( \left\lvert x_n \right\rvert = \left\lvert x_M \right\rvert \frac{x_{M+1}}{x_M} \cdots \frac{x_n}{x_{n-1}} \frac{x_n}{x_{n-1}} \cdots \frac{x_{n-M}}{\left\lvert x_{n-M} \right\rvert} \frac{x_{n-M}}{\left\lvert x_{n-M} \right\rvert} \cdots \frac{x_n}{x_{n-M}} = \left\lvert x_M \right\rvert r^{n-M} = \left\lvert x_M \right\rvert r^{-M} r^n \). The sequence \( \left\lvert x_n \right\rvert \) converges to zero and therefore \( \left\lvert x_n \right\rvert \) converges to zero.

Now suppose \( L > 1 \). Pick \( r \) such that \( 1 < r < L \). As \( L - r > 0 \), there exists an \( M \) in \( \mathbb{N} \) such that for all \( n \geq M \) we have \( \left\lvert \frac{x_{n+1}}{x_n} - L \right\rvert < L - r \). Therefore, \( \left\lvert \frac{x_{n+1}}{x_n} \right\rvert < L - r \). Again for \( n > M \) we write \( \frac{x_{n+1}}{x_n} = \frac{x_{M+1}}{x_M} \frac{x_{M+2}}{x_{M+1}} \cdots \frac{x_{n-M}}{\left\lvert x_{n-M} \right\rvert} \frac{x_{n-M}}{\left\lvert x_{n-M} \right\rvert} \cdots \frac{x_n}{x_{n-M}} > \frac{x_{M+1}}{x_M} \frac{x_{M+2}}{x_{M+1}} \cdots \frac{x_{n-M}}{\left\lvert x_{n-M} \right\rvert} \frac{x_{n-M}}{\left\lvert x_{n-M} \right\rvert} \cdots \frac{x_n}{x_{n-M}} \). The sequence \( \left\lvert r^n \right\rvert \) is unbounded (since \( r > 1 \)), and therefore \( \left\lvert x_n \right\rvert \) cannot be bounded (if \( \frac{x_{n+1}}{x_n} \geq B \) for all \( n \), then \( r^n < \frac{x_{n+1}}{x_n} \frac{x_{n+2}}{x_{n+1}} \cdots \frac{x_n}{x_{n-1}} \frac{x_n}{x_{n-1}} \cdots \frac{x_{n-M}}{\left\lvert x_{n-M} \right\rvert} \frac{x_{n-M}}{\left\lvert x_{n-M} \right\rvert} \cdots \frac{x_n}{x_{n-M}} \)). Consequently, \( \left\lvert x_n \right\rvert \) cannot converge.

A simple application of the above lemma is to prove that \( \lim_{n \to \infty} \frac{2^n}{n!} = 0 \).

Proof: We find that \( \frac{2^{n+1}}{(n+1)!} \frac{n!}{2^n} = \frac{2}{n+1} \). It is not hard to see that \( \left\lvert \frac{2}{n+1} \right\rvert \) converges to zero. The conclusion follows by the lemma.

### Exercises

Prove . Hint: Use constant sequences and .


Prove that if \( \{ x_n \} \) is a convergent sequence, \( \forall k \in \mathbb{N} \), then \( \lim_{n \to \infty} x_n k = \left( \lim_{n \to \infty} x_n \right) k \). Hint: Use .

Suppose \( (x_1 := \frac{1}{2}) \) and \( (x_{n+1} := x_n^2) \). Show that \( \{ x_n \} \) converges and find \( \lim x_n \). Hint: You cannot divide by zero!

Let \( (x_n := \frac{n - \cos(n)}{n^2}) \). Use the to show that \( \{ x_n \} \) converges and find the limit.

Let \( (y_n := \frac{n^2}{n^2}) \) and \( (y_n := \frac{\sin(n)}{n^2}) \). Define \( (z_n := \frac{n^2}{n^2}) \) and \( (w_n := \frac{\sin(n)}{n^2}) \). Do \( \{ y_n \} \) and \( \{ w_n \} \) converge? What are the limits? Can you apply ? Why or why not?

True or false, prove or find a counterexample. If \( \{ x_n \} \) is a sequence such that \( \{ x_n^2 \} \) converges, then \( \{ x_n \} \) converges.
Show that \[
\lim_{n \to \infty} \frac{n^2}{2^n} = 0 .\]

Suppose \(\{ x_n \}\) is a sequence and suppose for some \(x \in \mathbb{R}\), the limit \(L := \lim_{n \to \infty} \left| \frac{x_{n+1} - x}{x_n - x} \right|\) exists and \(L < 1\). Show that \(\{ x_n \}\) converges to \(x\).

Let \(\{ x_n \}\) be a convergent sequence such that \(x_n \geq 0\) and \(k \in \mathbb{N}\). Then \(\lim_{n \to \infty} x_n \geq 1/k\) exists. Hint: Find an expression \(q\) such that \(\lim_{n \to \infty} x_n \geq 1/k\).

Let \((r > 0)\). Show that starting with any \(x_1 \not= 0\), the sequence defined by \(x_{n+1} := x_n - \frac{x_n^2 - r}{2x_n}\) converges to \(\sqrt{r}\) if \(x_1 > 0\) and \((-\sqrt{r})\) if \(x_1 < 0\).

a) Suppose \(\{ a_n \}\) is a bounded sequence and \(\{ b_n \}\) is a sequence converging to 0. Show that \(\{ a_n b_n \}\) converges to 0.

b) Find an example where \(\{ a_n \}\) is unbounded, \(\{ b_n \}\) converges to 0, and \(\{ a_n b_n \}\) is not convergent.

c) Find an example where \(\{ a_n \}\) is bounded, \(\{ b_n \}\) converges to some \(x \not= 0\), and \(\{ a_n b_n \}\) is not convergent.

---

**Limit superior, limit inferior, and Bolzano-Weierstrass**

*Note: 1–2 lectures, alternative proof of BW optional*

In this section we study bounded sequences and their subsequences. In particular we define the so-called limit superior and limit inferior of a bounded sequence and talk about limits of subsequences. Furthermore, we prove the Bolzano-Weierstrass theorem, which is an indispensable tool in analysis.

We have seen that every convergent sequence is bounded, although there exist many bounded divergent sequences. For example, the sequence \(\{ (-1)^n \}\) is bounded, but it is divergent. All is not lost however and we can still compute certain limits with a bounded divergent sequence.

**Upper and lower limits**

There are ways of creating monotone sequences out of any sequence, and in this fashion we get the so-called limit superior and limit inferior. These limits always exist for bounded sequences.

If a sequence \(\{ x_n \}\) is bounded, then the set \(\{ x_k : k \in \mathbb{N} \}\) is bounded. Then for every \(n\) the set \(\{ x_k : k \geq n \}\) is also bounded (as it is a subset).

**liminflimsup:ref** Let \(\{ x_n \}\) be a bounded sequence. Let \(a_n := \sup \{ x_k : k \geq n \}\) and \(b_n := \inf \{ x_k : k \leq n \}\). Define \(\limsup \{ x_n \} = \lim_{n \to \infty} a_n\) and \(\liminf \{ x_n \} = \lim_{n \to \infty} b_n\).
For a bounded sequence, liminf and limsup always exist (see below). It is possible to define liminf and limsup for unbounded sequences if we allow \(\infty\) and \(-\infty\). It is not hard to generalize the following results to include unbounded sequences, however, we first restrict our attention to bounded ones.

Let \(\{ x_n \}\) be a bounded sequence. Let \(\{a_n\}\) and \(\{b_n\}\) be as in the definition above.

i. The sequence \(\{a_n\}\) is bounded monotone decreasing and \(\{b_n\}\) is bounded monotone increasing. In particular, \(\lim\inf x_n\) and \(\lim\sup x_n\) exist.

ii. \(\lim\sup_{n \to \infty} x_n = \inf \{ a_n : n \in \mathbb{N} \}\) and \(\lim\inf_{n \to \infty} x_n = \sup \{ b_n : n \in \mathbb{N} \}\).

iii. \(\lim\inf_{n \to \infty} x_n \leq \lim\sup_{n \to \infty} x_n\).

Let \(\{ a_n \}\) be a decreasing sequence. As \(a_n\) is the least upper bound for \(\{ x_k : k \geq n \}\), it is also an upper bound for the subset \(\{ x_k : k \geq (n+1) \}\). Therefore \(a_{(n+1)}\), the least upper bound for \(\{ x_k : k \geq (n+1) \}\), has to be less than or equal to \(a_n\), that is, \(a_n \geq a_{(n+1)}\). Similarly (an exercise), \(b_n\) is an increasing sequence. It is left as an exercise to show that if \(\{x_n\}\) is bounded, then \(\{a_n\}\) and \(\{b_n\}\) must be bounded.

The second item in the proposition follows as the sequences \(\{a_n\}\) and \(\{b_n\}\) are monotone.

For the third item, we note that \(\{b_n\}\) is the set of upper bounds for \(\{a_n\}\). However, we know that \(\{a_n\}\) and \(\{b_n\}\) converge to the limsup and liminf (respectively). We apply to obtain \(\lim\inf_{n \to \infty} b_n \leq \lim\inf_{n \to \infty} a_n\). We leave it to the reader to show that the sequence \(\{a_n\}\) is not a convergent sequence.

We associate with \(\lim\sup\) and \(\lim\inf\) certain subsequences.

\[\text{subseqlimsupinf:thm}\]
If \(\{x_n\}\) is a bounded sequence, then there exists a subsequence \(\{x_{n_k}\}\) such that \(\lim_{k \to \infty} x_{n_k} = \lim\sup_{n \to \infty} x_n\). Similarly, there exists a (perhaps different) subsequence \(\{x_{m_k}\}\) such that \(\lim_{k \to \infty} x_{m_k} = \lim\inf_{n \to \infty} x_n\).

Define \(a_n := \sup \{ x_k : k \geq n \}\). Write \(x := \lim\sup x_n = \lim a_n\). Define the subsequence as follows. Pick \(n_1 := 1\) and work inductively. Suppose we have defined the subsequence until \(n_k\) for some \(k\). Now pick some \(m > n_k\) such that \(a_m > x_{m+1}\). We can do this as \(a_m\) is a supremum of the set \(\{x_n : n > m\}\) and hence there are elements of the sequence arbitrarily close (or even possibly equal) to the supremum.
Set \( n_{k+1} := m \). The subsequence \( \{ x_{n_k} \} \) is defined. Next we need to prove that it converges and has the right limit.

Note that \( a_{(n_{k-1}+1)} \geq a_{n_k} \) (why?) and that \( a_{(n_{k-1}+1)} \geq x_{n_k} \). Therefore, for every \( k > 1 \) we have
\[
\left| a_{n_k} - x_{n_k} \right| = a_{n_k} - x_{n_k} \leq a_{(n_{k-1}+1)} - x_{n_k} < \frac{1}{k}.
\]

Let us show that \( \{ x_{n_k} \} \) converges to \( x \). Note that the subsequence need not be monotone. Let \( \epsilon > 0 \) be given. As \( \{ a_n \} \) converges to \( x \), then the subsequence \( \{ a_{n_k} \} \) converges to \( x \). Thus there exists an \( M_1 \) such that for all \( n \geq M_1 \) we have
\[
\left| a_{n_k} - x \right| < \frac{\epsilon}{2}.
\]

Find an \( M_2 \) such that \( \frac{1}{M_2} \leq \frac{\epsilon}{2} \). Take \( M := \max \{ M_1 , M_2 \} \) and compute. For all \( k \geq M \) we have
\[
\left| x - x_{n_k} \right| = \left| a_{n_k} - x_{n_k} + x - a_{n_k} \right| \leq a_{n_k} - x_{n_k} + x - a_{n_k} < \frac{1}{k} + \frac{\epsilon}{2} \leq \frac{1}{M_2} + \frac{\epsilon}{2} = \epsilon.
\]

We leave the statement for \( \liminf \) as an exercise.

### Using limit inferior and limit superior

The advantage of \( \liminf \) and \( \limsup \) is that we can always write them down for any (bounded) sequence. If we could somehow compute them, we could also compute the limit of the sequence if it exists, or show that the sequence diverges.

Working with \( \liminf \) and \( \limsup \) is a little bit like working with limits, although there are subtle differences.

\[[\text{liminfsupconv}:\text{thm}]\] Let \( \{ x_n \} \) be a bounded sequence. Then \( \{ x_n \} \) converges if and only if
\[
\liminf_{n \to \infty} x_n = \limsup_{n \to \infty} x_n.
\]
Furthermore, if \( \{ x_n \} \) converges, then
\[
\lim_{n \to \infty} x_n = \liminf_{n \to \infty} x_n = \limsup_{n \to \infty} x_n.
\]

Define \( (a_n) \) and \( (b_n) \) as in . Note that \( \{ b_n \} \) and \( \{ a_n \} \) have limits and that these two limits are the same. By the squeeze lemma (), \( \{ x_n \} \) converges and
\[
\lim_{n \to \infty} b_n = \lim_{n \to \infty} a_n = \lim_{n \to \infty} x_n.
\]

Now suppose \( \{ x_n \} \) converges to \( x \). We know by that there exists a subsequence \( \{ x_{n_k} \} \) that converges to \( \limsup x_n \). As \( \{ x_{n_k} \} \) converges to \( x \), every subsequence converges to \( x \) and therefore \( \limsup x_n = \lim x_{n_k} = x \). Similarly \( \liminf x_n = x \).

Limit superior and limit inferior behave nicely with subsequences.

\[[\text{prop:subseqlimsupinf}]\] Suppose \( \{ x_n \} \) is a bounded sequence and \( \{ x_{n_k} \} \) is a subsequence. Then
\[
\liminf_{n \to \infty} x_n \leq \liminf_{k \to \infty} x_{n_k} \leq \limsup_{k \to \infty} x_{n_k} \leq \limsup_{n \to \infty} x_n.
\]
The middle inequality has been proved already. We will prove the third inequality, and leave the first inequality as an exercise.

We want to prove that \( \limsup x_{n_k} \leq \limsup x_n \). Define \( a_j := \sup \{ x_k : k \geq j \} \) as usual. Also define \( c_j := \sup \{ x_{n_k} : k \geq j \} \). It is not true that \( c_j \) is necessarily a subsequence of \( a_j \). However, as \( \{ x_{n_k} : k \geq j \} \) is a subset of \( \{ x_k : k \geq j \} \), a supremum of a subset is less than or equal to the supremum of the set and therefore \( c_j \leq a_j \). We apply to conclude
\[
\lim_{j \to \infty} c_j \leq \lim_{j \to \infty} a_j ,
\]
which is the desired conclusion.

Limit superior and limit inferior are the largest and smallest subsequential limits. If the subsequence in the previous proposition is convergent, then we have that \( \liminf x_{n_k} = \lim x_{n_k} = \limsup x_{n_k} \). Therefore,
\[
\liminf_{n \to \infty} x_n \leq \lim_{k \to \infty} x_{n_k} \leq \limsup_{n \to \infty} x_n .
\]
Similarly we get the following useful test for convergence of a bounded sequence. We leave the proof as an exercise.

A bounded sequence \( \{ x_n \} \) is convergent and converges to \( x \) if and only if every convergent subsequence \( \{ x_{n_k} \} \) converges to \( x \).

**Bolzano-Weierstrass theorem**

While it is not true that a bounded sequence is convergent, the Bolzano-Weierstrass theorem tells us that we can at least find a convergent subsequence. The version of Bolzano-Weierstrass that we present in this section is the Bolzano-Weierstrass for sequences.

Suppose a sequence \( \{ x_n \} \) of real numbers is bounded. Then there exists a convergent subsequence \( \{ x_{n_i} \} \).

We use . It says that there exists a subsequence whose limit is \( \limsup x_n \).

The reader might complain right now that is strictly stronger than the Bolzano-Weierstrass theorem as presented above. That is true. However, only applies to the real line, but Bolzano-Weierstrass applies in more general contexts (that is, in \( \{ \mathbb{R} \} \)) with pretty much the exact same statement.

As the theorem is so important to analysis, we present an explicit proof. The following proof generalizes more easily to different contexts.

As the sequence is bounded, then there exist two numbers \( a_1 < b_1 \) such that \( a_1 \leq x_n \leq b_1 \) for all \( n \in \mathbb{N} \).

We will define a subsequence \( \{ x_{n_i} \} \) and two sequences \( \{ a_i \} \) and \( \{ b_i \} \), such that \( \{ a_i \} \) is monotone increasing, \( \{ b_i \} \) is monotone decreasing, \( a_i \leq x_{n_i} \leq b_i \) and such that \( \lim a_i = \lim b_i \).

That \( \{ x_{n_i} \} \) converges follows by the .
We define the sequences inductively. We will always have that \( a_i < b_i \), and that \( x_n \in [a_i, b_i] \) for infinitely many \( n \in \mathbb{N} \). We have already defined \( a_1 \) and \( b_1 \). We take \( n_1 := 1 \), that is \( x_{n_1} = x_1 \).

Now suppose that up to some \( k \in \mathbb{N} \) we have defined the subsequence \( x_{\{n_1\}}, x_{\{n_2\}}, \ldots, x_{\{n_k\}} \), and the sequences \( \{a_1, a_2, \ldots, a_k\} \) and \( \{b_1, b_2, \ldots, b_k\} \). Let \( y := \frac{a_k+b_k}{2} \). Clearly \( a_k < y < b_k \).

If there exist infinitely many \( j \in \mathbb{N} \) such that \( x_j \in [a_k, y] \), then set \( a_{k+1} := a_k \), \( b_{k+1} := y \), and pick \( n_{k+1} > n_k \) such that \( x_{n_{k+1}} \in [a_k, y] \). If there are not infinitely many \( j \) such that \( x_j \in [a_k, y] \), then it must be true that there are infinitely many \( j \in \mathbb{N} \) such that \( x_j \in [y, b_k] \). In this case pick \( a_{k+1} := y \), \( b_{k+1} := b_k \), and pick \( n_{k+1} > n_k \) such that \( x_{n_{k+1}} \in [y, b_k] \).

Now we have the sequences defined. What is left to prove is that \( \lim_{n \to \infty} a_i = \lim_{n \to \infty} b_i \). Obviously the limits exist as the sequences are monotone. From the construction, it is obvious that \( b_i - a_i \) is cut in half in each step. Therefore \( b_{i+1} - a_{i+1} = \frac{b_i-a_i}{2} \).

Let \( x := \lim_{n \to \infty} a_i \). As \( \{a_i\} \) is monotone we have that \( x = \sup \{ a_i : i \in \mathbb{N} \} \). Now let \( y := \lim_{n \to \infty} b_i \). Obviously \( y \leq x \) as \( a_i < b_i \) for all \( i \). As the sequences are monotone, then for any \( i \) we have (why?) \( y-x \leq b_{i-a_i} = \frac{b_i-a_i}{2^{i-1}} \).

Now another proof of the Bolzano-Weierstrass theorem is to show the following claim, which is left as a challenging exercise.

**Claim:** Every sequence has a monotone subsequence.

**Infinite limits**

If we allow \( \liminf \) and \( \limsup \) to take on the values \( \{-\infty, \infty\} \), we can apply \( \liminf \) and \( \limsup \) to all sequences, not just bounded ones. For any sequence, we write \( \limsup x_n := \inf \{ a_n : n \in \mathbb{N} \} \), \( \liminf x_n := \sup \{ b_n : n \in \mathbb{N} \} \), where \( a_n := \sup \{ x_k : k \geq n \} \) and \( b_n := \inf \{ x_k : k \geq n \} \) as before.

We also often define infinite limits for certain divergent sequences.

We say \( \{x_n\}_n \) **diverges to infinity** if for every \( M \in \mathbb{N} \) there exists an \( N \in \mathbb{N} \) such that for all \( n \geq N \) we have \( x_n > M \). In this case we write \( \lim_{n \to \infty} x_n := \infty \). Similarly if for every \( M \in \mathbb{N} \) there exists an \( N \in \mathbb{N} \) such that for all \( n \geq N \) we have \( x_n < M \), we say \( \{x_n\}_n \) diverges to minus infinity and we write \( \lim_{n \to \infty} x_n := -\infty \).

This definition behaves as expected with \( \limsup \) and \( \liminf \), see exercises [exercise:infseqlimex] and [exercise:infseqlimlims].

If \( \{x_n := 0\}_n \) for odd \( n \) and \( \{x_n := n\}_n \) for even \( n \) then \( \lim_{n \to \infty} n = \infty \).
Exercises

Suppose \( \{ x_n \} \) is a bounded sequence. Define \( \{ a_n \} \) and \( \{ b_n \} \) as in . Show that \( \{ a_n \} \) and \( \{ b_n \} \) are bounded.

Suppose \( \{ x_n \} \) is a bounded sequence. Define \( \{ b_n \} \) as in . Show that \( \{ b_n \} \) is an increasing sequence.

Finish the proof of . That is, suppose \( \{ x_n \} \) is a bounded sequence and \( \{ x_{n_k} \} \) is a subsequence. Prove

\[
\liminf_{n \to \infty} x_n \leq \liminf_{k \to \infty} x_{n_k}.
\]

Prove .

a. Let \( x_n := \frac{(-1)^n}{n} \), find \( \limsup x_n \) and \( \liminf x_n \).

b. Let \( x_n := \frac{(n-1)(-1)^n}{n} \), find \( \limsup x_n \) and \( \liminf x_n \).

Let \( \{ x_n \} \) and \( \{ y_n \} \) be bounded sequences such that \( x_n \leq y_n \) for all \( n \). Then show that \( \limsup_{n \to \infty} x_n \leq \limsup_{n \to \infty} y_n \) and \( \liminf_{n \to \infty} x_n \leq \liminf_{n \to \infty} y_n \).

Let \( \{ x_n \} \) and \( \{ y_n \} \) be bounded sequences.

a. Show that \( \{ x_n + y_n \} \) is bounded.

b. Show that \( \liminf_{n \to \infty} x_n + \liminf_{n \to \infty} y_n \leq \liminf_{n \to \infty} (x_n+y_n) \) Hint: Find a subsequence \( \{ x_{n_i} + y_{n_i} \} \) of \( \{ x_n + y_n \} \) that converges. Then find a subsequence \( \{ x_{n_{m_i}} \} \) of \( \{ x_{n_i} \} \) that converges. Then apply what you know about limits.

c. Find an explicit \( \{ x_n \} \) and \( \{ y_n \} \) such that \( \liminf_{n \to \infty} x_n + \liminf_{n \to \infty} y_n < \liminf_{n \to \infty} (x_n+y_n) \) Hint: Look for examples that do not have a limit.

Let \( \{ x_n \} \) and \( \{ y_n \} \) be bounded sequences (from the previous exercise we know that \( \{ x_n + y_n \} \) is bounded).

a. Show that \( \limsup_{n \to \infty} x_n + \limsup_{n \to \infty} y_n \geq \limsup_{n \to \infty} (x_n+y_n) \) Hint: See previous exercise.

b. Find an explicit \( \{ x_n \} \) and \( \{ y_n \} \) such that \( \limsup_{n \to \infty} x_n + \limsup_{n \to \infty} y_n > \limsup_{n \to \infty} (x_n+y_n) \) Hint: See previous exercise.

If \( S \subset \mathbb{R} \) is a set, then \( x \in \{ \mathbb{B} \} \) is a cluster point if for every \( \epsilon > 0 \), the set \( \{ x - \epsilon, x + \epsilon \} \cap S \setminus \{ x \} \) is not empty. That is, if there are points of \( \{ \} \) arbitrarily close to \( \{ x \} \). For example, \( \{ x = \{ \frac{1}{n} : n \in \{ \mathbb{N} \} \} \} \) has a unique (only one) cluster point \( \{ 0 \} \), but \( \{ 0 \} \) notin \( \{ \} \).

Prove the following version of the Bolzano-Weierstrass theorem:

**Theorem.** Let \( S \subset \{ \mathbb{B} \} \) be a bounded infinite set, then there exists at least one cluster point of \( \{ \} \).
Hint: If \( \langle S \rangle \) is infinite, then \( \langle S \rangle \) contains a countably infinite subset. That is, there is a sequence \( \langle \{ x_n \} \rangle \) of distinct numbers in \( \langle S \rangle \).

a) Prove that any sequence contains a monotone subsequence. Hint: Call \( \langle n \mid \text{in} \{ \text{mathbb \{ N \}} \} \rangle a \) peak if \( \langle a_m \mid \text{leq} a_n \rangle \) for all \( \langle m \mid \text{geq} n \rangle \). There are two possibilities: either the sequence has at most finitely many peaks, or it has infinitely many peaks.

b) Conclude the Bolzano-Weierstrass theorem.

Let us prove a stronger version of . Suppose \( \langle \{ x_n \} \rangle \) is a sequence such that every subsequence \( \langle \{ x_{n_i} \} \rangle \) has a subsequence \( \langle \{ x_{n_{m_i}} \} \rangle \) that converges to \( \langle x \rangle \). a) First show that \( \langle \{ x_n \} \rangle \) is bounded. b) Now show that \( \langle \{ x_n \} \rangle \) converges to \( \langle x \rangle \).

Let \( \langle \{ x_n \} \rangle \) be a bounded sequence.

a) Prove that there exists an \( \langle s \rangle \) such that for any \( \langle r \rangle \) there exists an \( \langle M \mid \text{in} \{ \text{mathbb \{ N \}} \} \rangle \) such that for all \( \langle n \mid \text{geq} M \rangle \) we have \( \langle x_n < r \rangle \).

b) If \( \langle s \rangle \) is a number as in a), then prove \( \langle \text{limsup } x_n \mid \text{leq} s \rangle \).

c) Show that if \( \langle \langle S \rangle \rangle \) is the set of all \( \langle s \rangle \) as in a), then \( \langle \text{limsup } x_n = \text{inf } S \rangle \).

Let \( \langle \{ x_n \} \rangle \) is such that \( \langle \text{liminf } x_n = - \text{infty} \rangle \), \( \langle \text{limsup } x_n = \text{infty} \rangle \). a) Show that \( \langle \{ x_n \} \rangle \) is not convergent, and also that neither \( \langle \text{lim } x_n = \text{infty} \rangle \) nor \( \langle \text{lim } x_n = - \text{infty} \rangle \) is true. b) Find an example of such a sequence.

Let \( \langle \{ x_n \} \rangle \). a) Show that \( \langle \text{lim } x_n = \text{infty} \rangle \) if and only if \( \langle \text{liminf } x_n = \text{infty} \rangle \). b) Then show that \( \langle \text{lim } x_n = - \text{infty} \rangle \) if and only if \( \langle \text{limsup } x_n = - \text{infty} \rangle \). c) If \( \langle \{ x_n \} \rangle \) is monotone increasing, show that either \( \langle \text{lim } x_n \rangle \) exists and is finite or \( \langle \text{lim } x_n = \text{infty} \rangle \).

**Cauchy sequences**

*Note: 0.5–1 lecture*

Often we wish to describe a certain number by a sequence that converges to it. In this case, it is impossible to use the number itself in the proof that the sequence converges. It would be nice if we could check for convergence without knowing the limit.

A sequence \( \langle \{ x_n \} \rangle \) is a *Cauchy sequence* if for every \( \langle \epsilon \rangle > 0 \) there exists an \( \langle M \mid \text{in} \{ \text{mathbb \{ N \}} \} \rangle \) such that for all \( \langle n \mid \text{geq} M \rangle \) and all \( \langle k \mid \text{geq} M \rangle \) we have \( \langle \left| x_n - x_k \right| < \epsilon \rangle \).

Intuitively what it means is that the terms of the sequence are eventually arbitrarily close to each other. We would expect such a sequence to be convergent. It turns out that is true because \( \langle \{ \text{mathbb \{ R \}} \} \rangle \) has the . First, let us look at some examples.

The sequence \( \langle \{ \text{nicefrac \{ 1 \} \{ n \} \} \} \rangle \) is a Cauchy sequence.

Proof: Given \( \langle \epsilon \rangle > 0 \), find \( \langle M \rangle \) such that \( \langle M > \text{nicefrac \{ 2 \} \{ \epsilon \} \} \rangle \). Then for \( \langle n,k \mid \text{geq} M \rangle \) we have that \( \langle \text{nicefrac \{ 1 \} \{ n \} < \text{nicefrac \{ \epsilon \} \{ 2 \} \} \rangle \) and \( \langle \text{nicefrac \{ 1 \} \{ k \} < \text{nicefrac \{ \epsilon \} \{ 2 \} \} \rangle \). Therefore for \( \langle n, k \mid \text{geq} M \rangle \)
we have \[ \left| \frac{1}{n} - \frac{1}{k} \right| \leq \left| \frac{1}{n} \right| + \left| \frac{1}{k} \right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \]

The sequence \( \{ \frac{n+1}{n} \} \) is a Cauchy sequence.

**Proof:** Given \( \epsilon > 0 \), find \( M \) such that \( M > \frac{2}{\epsilon} \). Then for \( n, k \geq M \) we have that
\[
\left| \frac{n}{n} - \frac{k}{k} \right| = \left| \frac{k(n+1)-n(k+1)}{nk} \right| = \left| \frac{k-n}{nk} \right| \leq \frac{1}{n} + \frac{1}{k} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\]

A Cauchy sequence is bounded.

Suppose \( \{ x_n \} \) is Cauchy. Pick \( M \) such that for all \( n, k \geq M \) we have \( \left| x_n - x_k \right| < 1 \). In particular, we have that for all \( n \geq M \) \( \left| x_n - x_M \right| < 1 \). By the reverse triangle inequality, \( \left| x_n - x_M \right| = \left| x_n - x_k + x_k - x_M \right| \leq \left| x_n - x_k \right| + \left| x_k - x_M \right| < 1 + \frac{\epsilon}{3} \). Let \( B := \max \left\{ \left| x_1 \right|, \left| x_2 \right|, \ldots, \left| x_{M-1} \right|, 1 + \left| x_M \right| \right\} \). Then \( \left| x_n \right| < B \) for all \( n \in \mathbb{N} \).

A sequence of real numbers is Cauchy if and only if it converges.

Let \( \epsilon > 0 \) be given and suppose \( \{ x_n \} \) converges to \( x \). Then there exists an \( M \) such that for \( n \geq M \) we have \( \left| x_n - x \right| < \frac{\epsilon}{2} \). Hence for \( n, k \geq M \) we have \( \left| x_n - x_k \right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \). Let \( a := \limsup \ x_n \) and \( b := \liminf \ x_n \). By the reverse triangle inequality, \( \left| a - b \right| = \left| a - x_n + x_n - b \right| \leq \left| a - x_n \right| + \left| x_n - b \right| < \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \).

Alright, that direction was easy. Now suppose \( \{ x_n \} \) is Cauchy. We have shown that \( \{ x_n \} \) must be convergent by . Assuming that \liminf and \limsup exist is where we use the .

Define \( a := \limsup \ x_n \) and \( b := \liminf \ x_n \). By , there exist subsequences \( \{ x_{n_i} \} \) and \( \{ x_{m_i} \} \), such that \( \lim_{i \to \infty} x_{n_i} = a \) and \( \lim_{i \to \infty} x_{m_i} = b \). Given an \( \epsilon > 0 \), there exists an \( M \) such that for all \( i \geq M \) we have \( \left| x_{n_i} - a \right| < \frac{\epsilon}{3} \) and \( \left| x_{m_i} - b \right| < \frac{\epsilon}{3} \). Therefore, \( \left| x_{n_i} - x_{m_i} \right| < \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \).

Let \( M := \max \{ M, M, M \} \). Note that if \( \{ a \} \), then \( \{ n \} \) and \( \{ m \} \). Hence \( \left| a - b \right| \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \), and the sequence converges.
The statement of this proposition is sometimes used to define the completeness property of the real numbers. We say a set is \textit{Cauchy-complete} (or sometimes just \textit{complete}) if every Cauchy sequence converges. Above we proved that as \( (\{\mathbb{R}\}) \) has the , then \( (\{\mathbb{R}\}) \) is Cauchy-complete. We can “complete” \( (\{\mathbb{Q}\}) \) by “throwing in” just enough points to make all Cauchy sequences converge (we omit the details). The resulting field has the least-upper-bound property. The advantage of using Cauchy sequences to define completeness is that this idea generalizes to more abstract settings.

It should be noted that the Cauchy criterion is stronger than just \( \left\lvert x_{n+1} - x_n \right\rvert \) (or \( \left\lvert x_{n+j} - x_n \right\rvert \) for a fixed \( j \)) going to zero as \( n \) goes to infinity. In fact, when we get to the partial sums of the harmonic series (see in the next section), we will have a sequence such that \( \left\lvert x_{n+1} - x_n \right\rvert = \frac{1}{n} \), yet \( \left\{ x_n \right\} \) is divergent. In fact, for that sequence it is true that \( \lim_{n \to \infty} \left\lvert x_{n+j} - x_n \right\rvert = 0 \) for any \( j \) \( \in \mathbb{N} \) (confer ). The key point in the definition of Cauchy is that \( \left\langle n \right\rangle \) and \( \left\langle k \right\rangle \) vary independently and can be arbitrarily far apart.

\textbf{Exercises}

Prove that \( \left\{ \frac{n^2-1}{n^2} \right\} \) is Cauchy using directly the definition of Cauchy sequences.

Let \( \left\{ x_n \right\} \) be a sequence such that there exists a \( 0 < C < 1 \) such that \( \left\lvert x_{n+1} - x_n \right\rvert \leq C \left\lvert x_{n-1} - x_{n-2} \right\rvert . \) Prove that \( \left\{ x_n \right\} \) is Cauchy. Hint: You can freely use the formula (for \( C \not= 1 \))

\[ 1 + C + C^2 + \cdots + C^n = \frac{1-C^{n+1}}{1-C}. \]

Suppose \( F \) is an ordered field that contains the rational numbers \( (\{\mathbb{Q}\}) \), such that \( (\{\mathbb{Q}\}) \) is dense, that is: whenever \( x, y \in F \) are such that \( x < y \), then there exists a \( q \in \{\mathbb{Q}\} \) such that \( x < q < y \). Say a sequence \( \left\{ x_n \right\} \) of rational numbers is Cauchy if given any \( \epsilon \in \{\mathbb{Q}\} \) with \( \epsilon > 0 \), there exists an \( (\{M\}) \) such that for all \( \left\langle n, k \geq M \right\rangle \) we have \( \left\lvert x_n - x_k \right\rvert < \epsilon \). Suppose any Cauchy sequence of rational numbers has a limit in \( (F) \). Prove that \( (F) \) has the .

Let \( \left\{ x_n \right\} \) and \( \left\{ y_n \right\} \) be sequences such that \( \lim_{n \to \infty} y_n = 0 \). Suppose that for all \( k \in \{\mathbb{N}\} \) and for all \( m \geq k \) we have \( \left\lvert x_m - x_k \right\rvert \leq y_k \). Show that \( \left\{ x_n \right\} \) is Cauchy.

Suppose a Cauchy sequence \( \left\{ x_n \right\} \) is such that for every \( M \in \{\mathbb{N}\} \), there exists a \( \left\langle k \geq M \right\rangle \) and an \( n \geq M \) such that \( x_k < 0 \) and \( x_n > 0 \). Using simply the definition of a Cauchy sequence and of a convergent sequence, show that the sequence converges to \( 0 \).

Suppose \( \left\{ x_n \right\} \) is a Cauchy sequence such that for every \( M \in \{\mathbb{N}\} \), there exists an \( \left\langle k \geq M \right\rangle \) and an \( n \geq M \) such that \( x_k < 0 \) and \( x_n > 0 \). Using only the definition of Cauchy sequence prove that \( \left\{ x_n \right\} \) converges to \( 0 \).

Suppose \( \left\{ x_n \right\} \) is a Cauchy sequence such that for infinitely many \( \langle n \rangle \), \( x_n = c \). Using only the definition of Cauchy sequence prove that \( \langle x_n \rangle \) converges to \( c \).

True/False prove or find a counterexample: If \( \left\{ x_n \right\} \) is a Cauchy sequence then there exists an \( \langle M \rangle \) such that for all \( n \geq M \) we have \( \left\lvert x_{n+1} - x_n \right\rvert \leq \left\lvert x_n - x_{n-1} \right\rvert \).
A fundamental object in mathematics is that of a series. In fact, when foundations of analysis were being developed, the motivation was to understand series. Understanding series is very important in applications of analysis. For example, solving differential equations often includes series, and differential equations are the basis for understanding almost all of modern science.

**Definition**

Given a sequence \( \{x_n\} \), we write the formal object \( \sum_{n=1}^{\infty} x_n \) or sometimes just \( \sum x_n \) and call it a series. A series converges, if the sequence \( \{s_k\} \) defined by \( s_k := \sum_{n=1}^{k} x_n = x_1 + x_2 + \cdots + x_k \), converges. The numbers \( \{s_k\} \) are called partial sums. If \( x := \lim s_k \), we write \( \sum_{n=1}^{\infty} x_n = x \). In this case, we cheat a little and treat \( \sum_{n=1}^{\infty} x_n \) as a number.

On the other hand, if the sequence \( \{s_k\} \) diverges, we say the series is divergent. In this case, \( \sum x_n \) is simply a formal object and not a number.

In other words, for a convergent series we have \( \sum_{n=1}^{\infty} x_n = \lim_{k \to \infty} s_k \). We should be careful to only use this equality if the limit on the right actually exists. That is, the right-hand side does not make sense (the limit does not exist) if the series does not converge.

Before going further, let us remark that it is sometimes convenient to start the series at an index different from 1. That is, for example we can write \( \sum_{n=0}^{\infty} r^n = \sum_{n=1}^{\infty} r^{n-1} \). The left-hand side is more convenient to write. The idea is the same as the notation for the tail of a sequence.

It is common to write the series \( \sum x_n \) as \( x_1 + x_2 + x_3 + \cdots \) with the understanding that the ellipsis indicates a series and not a simple sum. We do not use this notation as it often leads to mistakes in proofs.

The series \( \sum_{n=1}^{\infty} \frac{1}{2^n} \) converges and the limit is 1. That is, \( \sum_{n=1}^{\infty} \frac{1}{2^n} = \lim_{k \to \infty} \sum_{n=1}^{k} \frac{1}{2^n} = 1 \).

Proof: First we prove the following equality \( \left( \sum_{n=1}^{k} \frac{1}{2^n} \right) + \frac{1}{2^k} = 1 \). The equality is easy to see when \( k=1 \). The proof for general \( k \) follows by , which we leave to the reader. Let \( \{s_k\} \) be the partial sum. We write \( \left( \sum_{n=1}^{k} \frac{1}{2^n} \right) + \frac{1}{2^k} = \left( \sum_{n=1}^{k} \frac{1}{2^n} \right) + \frac{1}{2^k} \). The sequence \( \{s_k\} \) converges to zero. So, \( \{s_k\} \) converges to 1.

For \( -1 < r < 1 \), the geometric series \( \sum_{n=0}^{\infty} r^n \) converges. In fact, \( \sum_{n=0}^{\infty} r^n = \frac{1}{1-r} \). The proof is left as an exercise to the reader. The proof consists of showing \( \sum_{n=0}^{k} r^n = \frac{1}{1-r} \). and then taking the limit as \( k \) goes to \( \infty \).
A fact we often use is the following analogue of looking at the tail of a sequence.

Let \(\sum x_n\) be a series. Let \(M \in \mathbb{N}\). Then \(\sum_{n=1}^{\infty} x_n \quad \text{converges if and only if} \quad \sum_{n=M}^{\infty} x_n \quad \text{converges.}\)

We look at partial sums of the two series (for \(k \geq M\)) \(\sum_{n=1}^{k} x_n = \sum_{n=M}^{\infty} x_n \) Note that \(\sum_{n=1}^{M-1} x_n\) is a fixed number. Now use to finish the proof.

**Cauchy series**

A series \(\sum x_n\) is said to be *Cauchy* or a *Cauchy series*, if the sequence of partial sums \(\{s_n\}\) is a Cauchy sequence.

A sequence of real numbers converges if and only if it is Cauchy. Therefore a series is convergent if and only if it is Cauchy.

The series \(\sum x_n\) is Cauchy if for every \(\epsilon > 0\), there exists an \(M \in \mathbb{N}\) such that for every \(n \geq M\) and \(k \geq M\) we have \(\left| \sum_{j=1}^{k} x_j - \sum_{j=1}^{n} x_j \right| < \epsilon\). Without loss of generality we assume \(n < k\). Then we write \(\left| \sum_{j=1}^{k} x_j - \sum_{j=1}^{n} x_j \right| = \left| \sum_{j=n+1}^{k} x_j \right| < \epsilon\). We have proved the following simple proposition.

**[prop:cachyser]** The series \(\sum x_n\) is Cauchy if for every \(\epsilon > 0\), there exists an \(M \in \mathbb{N}\) such that for every \(n \geq M\) and every \(k > n\) we have \(\left| \sum_{j=1}^{k} x_j - \sum_{j=1}^{n} x_j \right| < \epsilon\).

**Basic properties**

Let \(\sum x_n\) be a convergent series. Then the sequence \(\{x_n\}\) is convergent and \(\lim_{n \to \infty} x_n = 0\).

Let \(\epsilon > 0\) be given. As \(\sum x_n\) is convergent, it is Cauchy. Thus we find an \(M\) such that for every \(n \geq M\) we have \(\left| \sum_{j=1}^{n+1} x_j \right| < \epsilon\). Hence for every \(n \geq M\) we have \(\left| \sum_{j=n+1}^{\infty} x_j \right| < \epsilon\).

So if a series converges, the terms of the series go to zero. The implication, however, goes only one way. Let us give an example.

**[example:harmonicseries]** The series \(\sum \frac{1}{n}\) diverges (despite the fact that \(\lim \frac{1}{n} = 0\)). This is the famous harmonic series.

**Proof:** We will show that the sequence of partial sums is unbounded, and hence cannot converge. Write the partial sums \(s_n\) for \(n = 2^k\) as:

\[
\begin{aligned}
\text{s}_1 &= 1, \\
\text{s}_2 &= \text{left}(1 \text{ right}) + \text{left}(\text{frac}\{1\}\{2\} \text{ right}), \\
\text{s}_4 &= \text{left}(1 \text{ right}) + \text{left}(\text{frac}\{1\}\{2\} \text{ right}) + \text{left}(\text{frac}\{1\}\{3\} + \text{frac}\{1\}\{4\} \text{ right}), \\
\text{s}_8 &= \text{left}(1 \text{ right}) + \text{left}(\text{frac}\{1\}\{2\} \text{ right}) + \text{left}(\text{frac}\{1\}\{3\} + \text{frac}\{1\}\{4\} \text{ right}) + \text{left}(\text{frac}\{1\}\{5\} + \text{frac}\{1\}\{6\} + \text{frac}\{1\}\{7\} + \text{frac}\{1\}\{8\} \text{ right}), \\
\vdots \\
\text{s}_{2^k} &= \sum_{j=1}^{k} \left( \sum_{m=2^{j-1}+1}^{2^j} \frac{1}{m} \right) .
\end{aligned}
\]

We note
that \(\frac{1}{3} + \frac{1}{4} \geq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}\) and \(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} \geq \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}\). More generally
\[
\sum_{m=2^{k-1}+1}^{2^k} \frac{1}{m} \geq \sum_{m=2^{k-1}+1}^{2^k} \frac{1}{2^k} = (2^{k-1}) \frac{1}{2^k} = \frac{1}{2} .
\]
Therefore \(s_{2^k} = 1 + \sum_{j=1}^k \left( \sum_{m=2^{k-1}+1}^{2^k} \frac{1}{m} \right) \geq 1 + \sum_{j=1}^k \frac{1}{2} = 1 + \frac{k}{2} .\) As \(\{ \frac{k}{2} \}\) is unbounded by the , that means that \(\{ s_{2^k} \}\) is unbounded, and therefore \(\{ s_n \}\) is unbounded. Hence \(\{ s_n \}\) diverges, and consequently \(\{ \sum \frac{1}{n} \}\) diverges.

Convergent series are linear. That is, we can multiply them by constants and add them and these operations are done term by term.

Let \(\alpha \in \mathbb{R}\) and \(\sum x_n\) and \(\sum y_n\) be convergent series. Then

i. \((\sum \alpha x_n)\) is a convergent series and \(\sum_{n=1}^\infty \alpha x_n = \alpha \sum_{n=1}^\infty x_n .\)

ii. \((\sum (x_n + y_n))\) is a convergent series and \(\sum_{n=1}^\infty (x_n + y_n) = \left( \sum_{n=1}^\infty x_n \right) + \left( \sum_{n=1}^\infty y_n \right) .\)

For the first item, we simply write the \((k)\)th partial sum \(\sum_{n=1}^k \alpha x_n = \alpha \left( \sum_{n=1}^k x_n \right) .\) We look at the right-hand side and note that the constant multiple of a convergent sequence is convergent. Hence, we simply take the limit of both sides to obtain the result.

For the second item we also look at the \((k)\)th partial sum \(\sum_{n=1}^k (x_n + y_n) = \left( \sum_{n=1}^k x_n \right) + \left( \sum_{n=1}^k y_n \right) .\) We look at the right-hand side and note that the sum of convergent sequences is convergent. Hence, we simply take the limit of both sides to obtain the proposition.

Note that multiplying series is not as simple as adding, see the next section. It is not true, of course, that we can multiply term by term, since that strategy does not work even for finite sums. For example, \(^((a+b)(c+d) \not= ac+bd)\).

**Absolute convergence**

Since monotone sequences are easier to work with than arbitrary sequences, it is generally easier to work with series \(\sum x_n\) where \(\langle x_n \rangle \) for all \(\langle n \rangle\). Then the sequence of partial sums is monotone increasing and converges if it is bounded from above. Let us formalize this statement as a proposition.

If \(\langle x_n \rangle \) for all \(\langle n \rangle\), then \(\sum x_n\) converges if and only if the sequence of partial sums is bounded from above.

As the limit of a monotone increasing sequence is the supremum, have the inequality \(\sum_{n=1}^\infty k x_n \leq \sum_{n=1}^\infty \text{inf}y x_n .\)

The following criterion often gives a convenient way to test for convergence of a series.

A series \(\sum x_n\) **converges absolutely** if the series \(\sum \left| x_n \right|\) converges. If a series converges, but does not converge absolutely, we say it is **conditionally convergent**.
If the series $\sum x_n$ converges absolutely, then it converges.

A series is convergent if and only if it is Cauchy. Hence suppose $\sum |x_n|$ is Cauchy. That is, for every $\epsilon > 0$, there exists an $M$ such that for all $k > M$ we have $\sum_{j=k+1}^n |x_j| < \epsilon$. We apply the triangle inequality for a finite sum to obtain $\sum_{j=k+1}^n |x_j| < \epsilon$. Hence $\sum x_n$ is Cauchy and therefore it converges.

Of course, if $\sum x_n$ converges absolutely, the limits of $\sum x_n$ and $\sum |x_n|$ are different. Computing one does not help us compute the other.

Absolutely convergent series have many wonderful properties. For example, absolutely convergent series can be rearranged arbitrarily, or we can multiply such series together easily. Conditionally convergent series on the other hand do not often behave as one would expect. See the next section.

We leave as an exercise to show that $\sum_{n=1}^\infty \frac{(-1)^n}{n}$ converges, although the reader should finish this section before trying. On the other hand we proved $\sum_{n=1}^\infty \frac{1}{n}$ diverges. Therefore $\sum_{n=1}^\infty \frac{(-1)^n}{n}$ is a conditionally convergent subsequence.

**Comparison test and the p-series**

We have noted above that for a series to converge the terms not only have to go to zero, but they have to go to zero “fast enough.” If we know about convergence of a certain series we can use the following comparison test to see if the terms of another series go to zero “fast enough.”

Let $\sum x_n$ and $\sum y_n$ be series such that $0 \leq x_n \leq y_n$ for all $n \in \mathbb{N}$.

i. If $\sum y_n$ converges, then so does $\sum x_n$.

ii. If $\sum x_n$ diverges, then so does $\sum y_n$.

Since the terms of the series are all nonnegative, the sequences of partial sums are both monotone increasing. Since $0 \leq x_n \leq y_n$ for all $n$, the partial sums satisfy for all $k$ $\sum_{n=1}^k x_n \leq \sum_{n=1}^k y_n$. If the series $\sum y_n$ converges the partial sums for the series are bounded. Therefore the right-hand side of $\sum_{n=1}^\infty \frac{(-1)^n}{n}$ is bounded for all $k$. Hence the partial sums for $\sum x_n$ are also bounded. Since the partial sums are a monotone increasing sequence they are convergent. The first item is thus proved.

On the other hand if $\sum x_n$ diverges, the sequence of partial sums must be unbounded since it is monotone increasing. That is, the partial sums for $\sum x_n$ are eventually bigger than any real number. Putting this together with $\sum_{n=1}^\infty \frac{(-1)^n}{n}$ we see that for any $B \in \mathbb{R}$, there is a $k$ such that $\sum_{n=1}^k x_n \geq B$. Hence the partial sums for $\sum y_n$ are also unbounded, and $\sum y_n$ also diverges.

A useful series to use with the comparison test is the $(p)$-series.
For \(p \in \mathbb{R}\), the series \(\sum_{n=1}^\infty \frac{1}{n^p}\) converges if and only if \(p > 1\).

First suppose \(p \leq 1\). As \(n \geq 1\), we have \(\frac{1}{n^p} \geq \frac{1}{n}\). Since \(\sum \frac{1}{n}\) diverges, we see that \(\sum \frac{1}{n^p}\) must diverge for all \(p \leq 1\) by the comparison test.

Now suppose \(p > 1\). We proceed in a similar fashion as we did in the case of the harmonic series, but instead of showing that the sequence of partial sums is unbounded we show that it is bounded. Since the terms of the series are positive, the sequence of partial sums is monotone increasing and will converge if we show that it is bounded above. Let \(s_n\) denote the \(n\)th partial sum. \(s_1 = 1\), \(s_3 = 1 + \frac{1}{3^p}\), \(s_7 = 1 + \frac{1}{3^p} + \frac{1}{7^p}\), \(s_{2^k-1} = 1 + \sum_{j=1}^{k-1} \frac{1}{2^j}\).

Instead of estimating from below, we estimate from above. In particular, as \(p\) is positive, then \(2^p < 3^p\), and hence \(\frac{1}{2^p} + \frac{1}{3^p} < \frac{1}{2^p} + \frac{1}{2^p}\). Similarly \(\frac{1}{4^p} + \frac{1}{5^p} + \frac{1}{6^p} + \frac{1}{7^p} < \frac{1}{4^p} + \frac{1}{4^p}\). Therefore \(s_{2^k-1} < 1 + \sum_{j=1}^{k-1} \frac{1}{2^j}\).

As \(p > 1\), then \(\frac{1}{2^p} < 1\). Then by using the result of , we note that \(\sum_{j=1}^\infty \frac{1}{2^j}\) converges. Therefore \(s_{2^k-1} < 1 + \sum_{j=1}^{k-1} \left( \frac{1}{2^j} \right)^n\).

As \(p > 1\), then \(\frac{1}{2^p} < 1\). Then by using the result of , we note that \(\sum_{j=1}^\infty \frac{1}{2^j}\) converges. Therefore \(s_{2^k-1} < 1 + \sum_{j=1}^{k-1} \left( \frac{1}{2^j} \right)^n\).

As \(p > 1\), then \(\frac{1}{2^p} < 1\). Then by using the result of , we note that \(\sum_{j=1}^\infty \frac{1}{2^j}\) converges. Therefore \(s_{2^k-1} < 1 + \sum_{j=1}^{k-1} \left( \frac{1}{2^j} \right)^n\).

The sequence of partial sums is bounded and hence converges.

Note that neither the \((p)\)-series test nor the comparison test tell us what the sum converges to. They only tell us that a limit of the partial sums exists. For example, while we know that \(\sum \frac{1}{n^3}\) converges it is far harder to find the limit is \(\frac{\pi^3}{36}\).

The series \(\sum \frac{1}{n^2+1}\) converges. Proof: First note that \(\frac{1}{n^2+1} < \frac{1}{n^2}\) for all \(n \in \mathbb{N}\). Note that \(\sum \frac{1}{n^2}\) converges by the \((p)\)-series test. Therefore, by the comparison test, \(\sum \frac{1}{n^2+1}\) converges.

**Ratio test**

Let \(\sum x_n\) be a series such that \(L := \lim_{n \to \infty} \frac{x_{n+1}}{x_n}\). If \(L < 1\), then \(\sum x_n\) converges absolutely.

If \(L > 1\), then \(\sum x_n\) diverges.

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
From we note that if \( L > 1 \), then \( x_n \) diverges. Since it is a necessary condition for the convergence of series that the terms go to zero, we know that \( \sum x_n \) must diverge.

Thus suppose \( L < 1 \). We will argue that \( \sum |x_n| \) must converge. The proof is similar to that of . Of course \( L \geq 0 \). Pick \( r \) such that \( L < r < 1 \). As \( (r - L) > 0 \), there exists an \( (n \geq M) \) such that for all \( n \geq M \):

\[
\left| \frac{|x_{n+1}|}{|x_n|} - L \right| < r - L.
\]

Therefore, \( \sum |x_n| \) converges. For \( n > M \) (that is for \( n \geq M+1 \)) write

\[
\frac{|x_{n+1}|}{|x_n|} < r.
\]

For \( n > M \) (that is for \( n \geq M + 1 \)) write

\[
\sum_{n=1}^k |x_n| = \sum_{n=1}^M |x_n| + \sum_{n=M+1}^k |x_n| < \sum_{n=1}^M |x_n| + (|x_M| r^{-M}) \left( \sum_{n=M+1}^k r^n \right) < \sum_{n=1}^M |x_n| + (|x_M| r^{-M}) \left( \sum_{n=M+1}^\infty r^n \right) .
\]

As \( 0 < r < 1 \) the geometric series \( \sum |n| = \sum_{n=0}^\infty |n| \) converges, so \( \sum_{n=1}^\infty |x_n| \) converges as well (why?). We take the limit as \( k \) goes to infinity on the right-hand side above to obtain

\[
\lim_{k \to \infty} \frac{\sum_{n=1}^k |x_n|}{|x_n|} \leq \frac{\sum_{n=1}^M |x_n|}{|x_n|} + \frac{|x_M| r^{-M}) \left( \sum_{n=M+1}^\infty r^n \right)}{r^n} \text{ for } n > M \].

The series \( \sum_{n=1}^\infty |x_n| \) converges absolutely.

Proof: We note that \( \sum_{n=1}^\infty \frac{3}{9n+1} \) converges absolutely.

For \( \frac{\sum_{k=0}^{n-1} r^k}{n!} \) converges absolutely. Therefore, the series converges absolutely by the ratio test.

**Exercises**

For \( r \not= 1 \), prove \( \frac{\sum_{k=0}^{n-1} r^k}{n!} \) converges. Hint: Let \( s := \sum_{k=0}^{n-1} r^k \), then compute \( s(1-r) = s-rs \), and solve for \( s \).

[geometric:exr] Prove that for \( -1 < r < 1 \) we have \( \sum_{n=0}^{\infty} r^n = \frac{1}{1-r} \). Hint: Use the previous exercise.

Decide the convergence or divergence of the following series.

\[
\begin{align*}
a) \sum_{n=1}^{\infty} \frac{3}{9n+1} \\
b) \sum_{n=1}^{\infty} \frac{1}{2n-1} \\
c) \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}
\end{align*}
\]
a. Prove that if \(\sum_{n=1}^{\infty} x_n\) converges, then \(\sum_{n=1}^{\infty} (x_{2n} + x_{2n+1})\) also converges.

b. Find an explicit example where the converse does not hold.

Prove the following stronger version of the ratio test: Let \(\sum x_n\) be a series.

a. If there is an \(N\) and a \(\rho < 1\) such that for all \(n \geq N\) we have \(\left|\frac{x_{n+1}}{x_n}\right| < \rho\), then the series converges absolutely.

b. If there is an \(N\) such that for all \(n \geq N\) we have \(\left|\frac{x_{n+1}}{x_n}\right| \geq 1\), then the series diverges.

Let \(\{x_n\}\) be a decreasing sequence such that \(\sum x_n\) converges. Show that \(\lim_{n \to \infty} n x_n = 0\).

Show that \(\sum_{n=1}^{\infty} \frac{(-1)^n}{n}\) converges. Hint: consider the sum of two subsequent entries.

a. Prove that if \(\sum x_n\) and \(\sum y_n\) converge absolutely, then \(\sum x_n y_n\) converges absolutely.

b. Find an explicit example where the converse does not hold.

c. Find an explicit example where all three series are absolutely convergent, are not just finite sums, and \(\sum x_n (\sum y_n) \not= \sum x_n y_n\). That is, show that series are not multiplied term-by-term.

Prove the triangle inequality for series, that is if \(\sum x_n\) converges absolutely then \(\left|\sum_{n=1}^{\infty} x_n\right| \leq \sum_{n=1}^{\infty} |x_n|\).

Prove the limit comparison test. That is, prove that if \(\{a_n > 0\}\) and \(\{b_n > 0\}\) for all \(n\), and \(0 < \lim_{n \to \infty} \frac{a_n}{b_n} < \infty\), then either \(\sum a_n\) and \(\sum b_n\) both converge or both diverge.

[exercise:badnoaucy] Let \(x_n = \sum_{j=1}^{n} \frac{1}{j}\). Show that for every \(k\) we have \(\lim_{n \to \infty} \left|\frac{x_n+k-x_n}{\sqrt{k}}\right| = 0\), yet \(\{x_n\}\) is not Cauchy.
Let \( s_k \) be the \( k \)th partial sum of \( \sum x_n \).

a) Suppose that there exists a \( m \in \mathbb{N} \) such that \( \lim_{k \to \infty} s_{mk} \) exists and \( \lim x_n = 0 \). Show that \( \sum x_n \) converges.

b) Find an example where \( \lim_{k \to \infty} s_{2k} \) exists and \( \lim x_n \not= 0 \) (and therefore \( \sum x_n \) diverges).

c) (Challenging) Find an example where \( \lim x_n = 0 \), and there exists a subsequence \( \{ s_{k_j} \} \) such that \( \lim_{j \to \infty} s_{k_j} \) exists, but \( \sum x_n \) still diverges.

More on series

Note: up to 2–3 lectures (optional, can safely be skipped or covered partially)

Root test

We have seen the ratio test before. There is one more similar test called the root test. In fact, the proof of this test is similar and somewhat easier.

Let \( \sum x_n \) be a series and let 
\[
L := \limsup_{n \to \infty} \sqrt[n]{|x_n|}.
\]
Then
i. If \( L < 1 \) then \( \sum x_n \) converges absolutely.

ii. If \( L > 1 \) then \( \sum x_n \) diverges.

If \( L > 1 \), then there exists a subsequence \( \{ x_{n_k} \} \) such that \( \lim_{k \to \infty} \sqrt[n_k]{|x_{n_k}|} = L > 1 \). Let \( r \) be such that \( L > r > 1 \). There exists an \( M \) such that for all \( k \geq M \), we have \( \sqrt[n_k]{|x_{n_k}|} > r \), or in other words \( \sqrt[n_k]{|x_{n_k}|} > r^{n_k} \). The subsequence \( \{ \sqrt[n_k]{|x_{n_k}|} \} \), and therefore also \( \{ |x_{n_k}| \} \), cannot possibly converge to zero, and so the series diverges.

Now suppose \( L < 1 \). Pick \( r \) such that \( L < r < 1 \). By definition of limit supremum, pick \( M \) such that for all \( n \geq M \) we have \( \sqrt[n]{|x_n|} < r \). Therefore, for all \( n \geq M \) we have \( \sqrt[n]{|x_n|} < r^n \). Let \( k > M \) and let us estimate \( k \)th partial sum \( \sum_{n=1}^k x_n \). By the geometric series, \( \sum_{n=1}^k r^n = \frac{r^{k+1} - 1}{r - 1} \). As \( 0 < r < 1 \), the geometric series converges to \( \frac{r^{k+1} - 1}{r - 1} \) as \( k \to \infty \). Hence, \( \sum_{n=1}^k x_n \) converges to \( \lim_{k \to \infty} \frac{r^{k+1} - 1}{r - 1} \). As everything is positive, we have \( \sum_{n=1}^k x_n \) is dominated by \( \frac{r^{k+1} - 1}{r - 1} \), and so the series converges. Therefore \( \sum x_n \) converges absolutely.
**Alternating series test**

The tests we have so far only addressed absolute convergence. The following test gives a large supply of conditionally convergent series.

Let \(\{x_n\}\) be a monotone decreasing sequence of positive real numbers such that \(\lim x_n = 0\). Then \(\sum_{n=1}^{\infty} (-1)^n x_n\) converges.

Write \(s_m := \sum_{k=1}^{m} (-1)^k x_k\) be the \(m\)th partial sum. Then write \(\sum_{n=1}^{2n} (-1)^n x_n = \sum_{n=1}^{2k} (-1)^n x_n\). The sequence \(\{x_n\}\) is decreasing and \(-x_{2k-1} + x_{2k}\) is bounded by \(0\) for all \(k\). Therefore the subsequence \(\{s_{2n}\}\) of partial sums is a decreasing sequence. Similarly, \(\sum_{n=1}^{\infty} (-1)^n x_n\) is decreasing and bounded below, so it converges. Let \(a := \lim s_{2n}\).

We wish to show that \(\lim s_m = a\) (not just for the subsequence). Notice \(\sum_{n=1}^{2n+1} = \sum_{n=2n} + x_{2n+1}\). Given \(\epsilon > 0\), pick \(M\) such that \(\left|\sum_{n=1}^{2n} - a\right| < \frac{\epsilon}{2}\) whenever \(2n \geq M\). Since \(\lim x_n = 0\), we also make \(M\) possibly larger to obtain \(\left|\sum_{n=2n+1} - \frac{\epsilon}{2}\right| < \frac{\epsilon}{2}\) whenever \(2n + 1 \geq M\). If \(2n \geq M\), we have \(\left|\sum_{n=2n} - a\right| < \frac{\epsilon}{2}\) whenever \(2n \geq M\), so we just need to check the situation for \(\sum_{n=2n+1}\): \(\left|\sum_{n=2n+1} - a\right| = \left|\sum_{n=2n} + x_{2n+1} - a\right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon\). \(\square\)

In particular, there exist conditionally convergent series where the absolute values of the terms go to zero arbitrarily slowly. For example, \(\sum_{n=1}^{\infty} \frac{(-1)^n}{n^p}\) converges for arbitrarily small \(p > 0\), but it does not converge absolutely when \(p \leq 1\).

**Rearrangements**

Generally, absolutely convergent series behave as we imagine they should. For example, absolutely convergent series can be summed in any order whatsoever. Nothing of the sort holds for conditionally convergent series (see and ).

Take a series \(\sum_{n=1}^{\infty} x_n\). Given a bijective function \(\sigma: \mathbb{N} \to \mathbb{N}\), the corresponding rearrangement is the following series: \(\sum_{n=1}^{\infty} x_{\sigma(n)}\). We simply sum the series in a different order.

Let \(\sum x_n\) be an absolutely convergent series converging to a number \(\ell\). Let \(\sigma: \mathbb{N} \to \mathbb{N}\) be a bijection. Then \(\sum x_{\sigma(n)}\) is absolutely convergent and converges to \(\ell\).

In other words, a rearrangement of an absolutely convergent series converges (absolutely) to the same number.

Let \(\epsilon > 0\) be given. Then take \(M\) to be such that \(\left|\sum_{n=M+1}^{\infty} x_n - \ell\right| < \frac{\epsilon}{2}\). As
\(\sigma\) is a bijection, there exists a number \(K\) such that for each \(n \leq M\), there exists \(k \leq K\) such that \(\sigma(k) = n\). In other words \(\{1, 2, \ldots, M\} \subset \sigma(\{1, 2, \ldots, K\})\).

Then for any \(N \geq K\), let \(Q := \max \{\sigma\} \{1, 2, \ldots, K\}\) and compute \(\begin{split}
\sum_{n=1}^N x_{\sigma(n)} - x
& = \sum_{n=1}^M x_n + \sum_{\sigma(n) > M} x_{\sigma(n)} - x
& \leq \sum_{n=1}^M x_n + \sum_{n=M+1}^Q x_n
& < \epsilon
\end{split}\)

So \(\sum x_{\sigma(n)}\) converges. To see that the convergence is absolute, we apply the above argument to \(\sum |x_n|\) to show that \(\sum |x_{\sigma(n)}|\) converges.

[example:harmonsumanything] Let us show that the alternating harmonic series \(\sum \frac{1}{n}\), which does not converge absolutely, can be rearranged to converge to anything. The odd terms and the even terms both diverge to infinity (prove this!): \(\sum \frac{1}{2n-1} = \infty\), \(\sum \frac{1}{2n} = \infty\). Let \(a_n := \frac{1}{n}\) for simplicity, let an arbitrary number \(L \in \mathbb{R}\) be given, and set \(\sigma(1) := 1\). Suppose we have defined \(\sigma(n)\) for all \(n \leq N\). Let \(\sum_{n=1}^N a_{\sigma(n)} \leq L\), then let \(\sigma(N+1) := k\) be the smallest odd \(k \in \mathbb{N}\) that we have not used yet, that is \(\sigma(n) \neq k\) for all \(n \leq N\). Otherwise let \(\sigma(N+1) := k\) be the smallest even \(k\) that we have not yet used.

By construction \(\sigma\colon \mathbb{N} \to \mathbb{N}\) is one-to-one. It is also onto, because if we keep adding either odd (resp. even) terms, eventually we will pass \(\sum L\) and switch to the evens (resp. odds). So we switch infinitely many times.

Finally, let \(N\) be the \(\{N\}\) where we just pass \(\sum L\) and switch. For example suppose we have just switched from odd to even (so we start subtracting), and let \(N' > N\) be where we first switch back from even to odd. Then \(\sum_{n=1}^N a_{\sigma(n)} > \sum_{n=1}^{N-1} a_{\sigma(n)}\).

And similarly for switching in the other direction. Therefore, the sum up to \(\{N' - 1\}\) is within \(\sum_{\sigma(n) \notin \{N', \ldots, \infty\}} a_{\sigma(n)}\) of \(\sum L\). As we switch infinitely many times we obtain that \(\sum a_{\sigma(n)}\) converges.
\[ \sigma(n) = L. \]

Here is an example to illustrate the proof. Suppose \( L = 1.2 \), then the order is

\[ 1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{9} + \frac{1}{11} + \frac{1}{13} - \frac{1}{2} + \frac{1}{15} + \frac{1}{17} + \frac{1}{19} - \frac{1}{4} + \frac{1}{20} + \cdots. \]

At this point we are no more than \( \frac{1}{8} \) from the limit.

### Multiplication of series

As we have already mentioned, multiplication of series is somewhat harder than addition. If we have that at least one of the series converges absolutely, than we can use the following theorem. For this result it is convenient to start the series at 0, rather than at 1.

Suppose \( \sum_{n=0}^\infty a_n \) and \( \sum_{n=0}^\infty b_n \) are two convergent series, converging to \( A \) and \( B \) respectively. If at least one of the series converges absolutely, then the series \( \sum_{n=0}^\infty c_n \) where

\[ c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0 = \sum_{j=0}^n a_j b_{n-j}, \]

converges to \( AB \).

The series \( \sum c_n \) is called the Cauchy product of \( \sum a_n \) and \( \sum b_n \).

Suppose \( \sum a_n \) converges absolutely, and let \( \epsilon > 0 \) be given. In this proof instead of picking complicated estimates just to make the final estimate come out as less than \( \epsilon \), let us simply obtain an estimate that depends on \( \epsilon \) and can be made arbitrarily small.

Write \( A_m := \sum_{n=0}^m a_n, \quad B_m := \sum_{n=0}^m b_n \). We rearrange the \((m)\)th partial sum of \( \sum c_n \):

\[
\begin{split}
\left| \left( \sum_{n=0}^m c_n \right) - AB \right| &= \left| \left( \sum_{n=0}^m \left( \sum_{j=0}^n a_j b_{n-j} \right) \right) - AB \right| \\
&= \left| \left( \sum_{n=0}^m B_n a_{m-n} \right) - AB \right| \\
&= \left| \left( \sum_{n=0}^m (B_n - B) a_{m-n} \right) + B A_m - AB \right| \\
&\leq \left( \sum_{n=0}^m |B_n - B| |a_{m-n}| \right) + |B| |A_m - A|.
\end{split}
\]

We can surely make the second term on the right hand side go to zero. The trick is to handle the first term. Pick \( K \) such that for all \( m \geq K \) we have \( |\left( \sum_{n=0}^m a_n \right) - A| < \epsilon \) and also \( |\left( \sum_{n=0}^m b_n \right) - B| < \epsilon \). Finally, as \( \sum a_n \) converges absolutely, make sure that \( K \) is large enough such that for all \( \sum_{n=0}^K \), \( \sum_{n=0}^m B_n a_{m-n} \) is small enough, and hence \( \sum_{n=0}^m |B_n - B| |a_{m-n}| \) is small enough. As \( \sum b_n \) converges, then we have that \( B_n \leq \text{max}_{n=0,1,2,\ldots} \) is finite. Take \( \sum_{n=0}^{2K} \), then in particular \( (m+K+1 > K) \). So \( |\left( \sum_{n=0}^m a_n \right) - A| < \epsilon \).

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
Therefore, for \(m \geq 2K\) we have \(\left\lvert \sum_{n=0}^m c_n \right\rvert - AB\) \begin{align*}
& \leq \sum_{n=0}^m \left\lvert B_n - B \right\rvert + \left\lvert A_m - A \right\rvert \leq \epsilon, \\
& < \epsilon B_{\text{max}} + \epsilon \left( \sum_{n=0}^\infty \left\lvert a_n \right\rvert \right) + \left\lvert B \right\rvert \epsilon = \epsilon \left( B_{\text{max}} + \left( \sum_{n=0}^\infty \left\lvert a_n \right\rvert \right) + \left\lvert B \right\rvert \right) .
\end{align*}

The expression in the parenthesis on the right hand side is a fixed number. Hence, we can make the right hand side arbitrarily small by picking a small enough \(\epsilon > 0\). So \(\sum_{n=0}^\infty c_n\) converges to \((AB)\).

If both series are only conditionally convergent, the Cauchy product series need not even converge. Suppose we take \(a_n = b_n = (-1)^n \frac{1}{\sqrt{n+1}}\). The series \(\sum_{n=0}^\infty a_n = \sum_{n=0}^\infty b_n\) converges by the alternating series test, however, it does not converge absolutely as can be seen from the \((p)\)-test. Let us look at the Cauchy product.

\[c_n = (-1)^n \left( \frac{1}{\sqrt{n+1}} + \frac{1}{\sqrt{2n}} + \frac{1}{\sqrt{3(n-1)}} + \cdots + \frac{1}{\sqrt{2n}} + \frac{1}{\sqrt{n+1}} \right) = (-1)^n \sum_{j=0}^n \frac{1}{\sqrt{(j+1)(n-j+1)}} .\]

Therefore \[
\left\lvert c_n \right\rvert = \sum_{j=0}^n \frac{1}{\sqrt{(j+1)(n-j+1)}} \geq \sum_{j=0}^\infty \frac{1}{\sqrt{(n+1)(n+1)}} = 1 .\]

The terms do not go to zero and hence \(\sum c_n\) cannot converge.

**Power series**

Fix \(x_0 \in \mathbb{R}\). A **power series** about \(x_0\) is a series of the form \(\sum_{n=0}^\infty a_n (x-x_0)^n\). A power series is really a function of \(x\), and many important functions in analysis can be written as a power series.

We say that a power series is **convergent** if there is at least one \(x \not= x_0\) that makes the series converge. Note that it is trivial to see that if \(x=x_0\) then the series always converges since all terms except the first are zero. If the series does not converge for any point \(x \not= x_0\), we say that the series is **divergent**.

**[ps:expex]** The series \(\sum_{n=0}^\infty \frac{1}{n!} x^n\) is absolutely convergent for all \(x \in \mathbb{R}\). This can be seen using the ratio test: For any \(x\) notice that \(\lim_{n \to \infty} \left| \frac{x^n}{x^{n+1}} \right| = 0\). In fact, you may recall from calculus that this series converges to \(e^x\).

**[ps:1kex]** The series \(\sum_{n=0}^\infty \frac{1}{n!} x^n\) converges absolutely for all \(x \in (-1,1)\) via the ratio test: \(\lim_{n \to \infty} \left| \frac{x^{n+1}}{x^n} \right| = 1\). It converges at \(x=-1\), as \(\sum_{n=0}^\infty \frac{1}{n!}\) converges by the alternating series test. But the power series does not converge absolutely at \(x=-1\), because \(\sum_{n=0}^\infty \frac{1}{n!}\) does not converge. The series diverges at \(x=1\). When \(\lim_{n \to \infty} \left| \frac{x^{n+1}}{x^n} \right| > 1\), then the series diverges via the ratio test.

**[ps:divergeex]** The series \(\sum_{n=0}^\infty n^n x^n\) diverges for all \(x \not= 0\). Let us apply the root test \(\lim_{n \to \infty} \left| \frac{n^n}{n^n} \right| = 1\). Therefore the series diverges for all \(x \not= 0\).

In fact, convergence of power series in general always works analogously to one of the three examples above.
Let \(\sum a_n (x-x_0)^n\) be a power series. If the series is convergent, then either it converges at all \(x \in \mathbb{R}\), or there exists a number \(\rho\), such that the series converges absolutely on the interval \((x_0-\rho, x_0+\rho)\) and diverges when \(x < x_0-\rho\) or \(x > x_0+\rho\).

The number \(\rho\) is called the **radius of convergence** of the power series. We write \(\rho = \infty\) if the series converges for all \(x\), and we write \(\rho = 0\) if the series is divergent. See . In the radius of convergence is \(\rho=1\). In the radius of convergence is \(\rho=\infty\), and in the radius of convergence is \(\rho=0\).

Write \[R := \limsup_{n \to \infty} |a_n|^{1/n} .\] We use the root test to prove the proposition: \[L = \limsup_{n \to \infty} |a_n (x-x_0)^n|^{1/n} = \limsup_{n \to \infty} |a_n|^{1/n} |x-x_0|^n = \rho \limsup_{n \to \infty} |a_n|^{1/n} \] In particular if \(R = \infty\), then \(L=\infty\) for any \(x \not= x_0\), and the series diverges by the root test. On the other hand if \(R = 0\), then \(L=0\) for any \(x\), and the series converges absolutely for all \(x\).

Suppose \(0 < R < \infty\). The series converges absolutely if \(1 > L = R |x-x_0| \limsup_{n \to \infty} |a_n|^{1/n}\), or in other words when \(L = R |x-x_0| \limsup_{n \to \infty} |a_n|^{1/n} > 0\). The series diverges when \(1 < L = R |x-x_0| \limsup_{n \to \infty} |a_n|^{1/n}\), or \(L = R |x-x_0| \limsup_{n \to \infty} |a_n|^{1/n} < 0\). Letting \(\rho = \limsup_{n \to \infty} |a_n|^{1/n}\) completes the proof.

It may be useful to restate what we have learned in the proof as a separate proposition.

Let \(\sum a_n (x-x_0)^n\) be a power series, and let \[R := \limsup_{n \to \infty} |a_n|^{1/n} .\] If \(R = \infty\), the power series is divergent. If \(R=0\), then the power series converges everywhere. Otherwise the radius of convergence \(\rho = \frac{1}{R}\).

Often, radius of convergence is written as \(\rho = \limsup_{n \to \infty} |a_n|^{1/n}\) in all three cases, with the obvious understanding of what \(\rho\) should be if \(R = 0\) or \(R = \infty\).

Convergent power series can be added and multiplied together, and multiplied by constants. The proposition has an easy proof using what we know about series in general, and power series in particular. We leave the proof to the reader.

Let \(\sum_{n=0}^{\infty} a_n (x-x_0)^n\) and \(\sum_{n=0}^{\infty} b_n (x-x_0)^n\) be two convergent power series with radius of convergence at least \(\rho > 0\) and \(\alpha \in \mathbb{R}\). Then for all \(x\) such that \(|x-x_0| < \rho\), we have \(\alpha \sum_{n=0}^{\infty} a_n (x-x_0)^n + \sum_{n=0}^{\infty} b_n (x-x_0)^n\) convergent. That is, after performing the algebraic operations, the radius of convergence of the resulting series is at least \(\rho\). For all \(x\) with \(|x-x_0| < \rho\), we have two convergent series so their term by term addition and multiplication by constants follows by what we learned in the last section. For multiplication of two power series, the series are absolutely convergent inside the radius of convergence and that is why for those \(x\) we can apply Mertens’ theorem. Note that after applying an algebraic operation the radius of convergence could increase. See the exercises.
Let us look at some examples of power series. Polynomials are simply finite power series. That is, a polynomial is a power series where the \(a_n\)'s are zero for all \(n\) large enough. We expand a polynomial as a power series about any point \(x_0\) by writing the polynomial as a polynomial in \((x-x_0)\). For example, \((2x^2-3x+4)\) as a power series around \(x_0 = 1\) is \(2x^2-3x+4 = 3 + (x-1) + 2{(x-1)}^2 .\)

We can also expand rational functions, that is, ratios of polynomials as power series, although we will not completely prove this fact here. Notice that a series for a rational function only defines the function on an interval even if the function is defined elsewhere. For example, for the geometric series we have that for \(|x| < 1\) \(\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \).

The series diverges when \(|x| > 1\), even though \(\frac{1}{1-x}\) is defined for all \(x \neq 1\).

We can use the geometric series together with rules for addition and multiplication of power series to expand rational functions as power series around \(x_0\), as long as the denominator is not zero at \(x_0\). We state without proof that this is always possible, and we give an example of such a computation using the geometric series.

Let us expand \(\frac{x}{1+2x+x^2}\) as a power series around the origin \((x_0 = 0)\) and find the radius of convergence.

Write \(1+2x+x^2 = (1+x)^2\), and suppose \(|x| < 1\). Compute \(\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n\) \(\frac{x}{1+2x+x^2} = x \frac{1}{1-x}\) \(\frac{1}{1+2x+x^2} = x \left( \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{n+1} \right) = \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{n+1} \)

The radius of convergence is at least 1. We leave it to the reader to verify that the radius of convergence is exactly equal to 1.

You can use the method of partial fractions you know from calculus. For example, to find the power series for \(\frac{x^3+x}{x^2-1}\) at 0, write \(\frac{x^3+x}{x^2-1} = x + \frac{1}{1+x} - \frac{1}{1-x}\) \(\frac{x^3+x}{x^2-1} = x + \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{n+1} \sum_{n=0}^{\infty} x^n \)

### Exercises

Decide the convergence or divergence of the following series.

a) \(\sum_{n=1}^{\infty} \frac{1}{2^{2n+1}}\)

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

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

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

---

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
Suppose both $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ converge absolutely. Show that the product series, $\sum_{n=0}^{\infty} c_n$ where $c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0$, also converges absolutely.

[exercise:seriesconvergestoanything] Let $\sum a_n$ be conditionally convergent. Show that given any number $\langle x \rangle$ there exists a rearrangement of $\sum a_n$ such that the rearranged series converges to $\langle x \rangle$. Hint: See .

a) Show that the alternating harmonic series $\sum \frac{(-1)^n}{n}$ has a rearrangement such that for any $\langle x < y \rangle$, there exists a partial sum $\sum s_n$ of the rearranged series such that $\langle x < s_n < y \rangle$. b) Show that the rearrangement you found does not converge. See . c) Show that for any $\langle x \in \mathbb{R} \rangle$, there exists a subsequence of partial sums $\langle s_{n_k} \rangle$ of your rearrangement such that $\langle \lim \, s_{n_k} = x \rangle$.

For the following power series, find if they are convergent or not, and if so find their radius of convergence.

- a) $\sum_{n=0}^{\infty} 2^n x^n$
- b) $\sum_{n=0}^{\infty} n x^n$
- c) $\sum_{n=0}^{\infty} n! x^n$
- d) $\sum_{n=0}^{\infty} \frac{1}{(2k)!} (x-10)^n$
- e) $\sum_{n=0}^{\infty} x^{2n}$
- f) $\sum_{n=0}^{\infty} n! x^{n!}$

Suppose $\sum a_n x^n$ converges for $\langle x=1 \rangle$. a) What can you say about the radius of convergence? b) If you further know that at $\langle x=1 \rangle$ the convergence is not absolute, what can you say?

Expand $\langle \frac{x}{4-x^2} \rangle$ as a power series around $\langle x = 0 \rangle$ and compute its radius of convergence.

a) Find an example where the radius of convergence of $\sum a_n x^n$ and $\sum b_n x^n$ are 1, but the radius of convergence of the sum of the two series is infinite. b) (Trickier) Find an example where the radius of convergence of $\sum a_n x^n$ and $\sum b_n x^n$ are 1, but the radius of convergence of the product of the two series is infinite.

Figure out how to compute the radius of convergence using the ratio test. That is, suppose $\langle \sum a_n x^n \rangle$ is a power series and $\langle R := \lim \frac{\left|a_{n+1}\right|}{\left|a_n\right|} \rangle$ exists or is $\langle \infty \rangle$. Find the radius of convergence and prove your claim.

a) Prove that $\langle \lim \frac{n^{1/n}}{n^{1/n}} = 1 \rangle$. Hint: Write $\langle n^{1/n} = 1+b_n \rangle$ and note $\langle b_n > 0 \rangle$. Then show that $\langle \frac{(1+b_n)^n}{\frac{\left|b_n\right|}{\left|b_n^2\right|}} \rangle$ and use this to show that $\langle \lim \frac{n^{1/n}}{b_n^2} = 0 \rangle$. b) Use the result of part a) to show that if $\langle \sum a_n x^n \rangle$ is a convergent power series with radius of convergence $\langle R \rangle$, then $\langle \sum n a_n x^n \rangle$ is also convergent with the same radius of convergence.
There are different notions of summability (convergence) of a series than just the one we have seen. A common one is Cesàro summability. Let $\sum a_n$ be a series and let $s_n$ be the $n$th partial sum. The series is said to be Cesàro summable to $(a)$ if

\[ a = \lim_{n \to \infty} \frac{s_1 + s_2 + \cdots + s_n}{n}. \]

a) If $(\sum a_n)$ is convergent to $(a)$ (in the usual sense), show that $(\sum a_n)$ is Cesàro summable to $(a)$. b) Show that in the sense of Cesàro $(\sum (-1)^n)$ is summable to $\frac{1}{2}$. c) Let $(a_n := k)$ when $(n = k^3)$ for some $k \in \{\mathbb{N}\}$, $(a_n := -k)$ when $(n = k^3+1)$ for some $k \in \{\mathbb{N}\}$, otherwise let $(a_n := 0)$. Show that $(\sum a_n)$ diverges in the usual sense, (partial sums are unbounded), but it is Cesàro summable to 0 (seems a little paradoxical at first sight).

Show that the monotonicity in the alternating series test is necessary. That is, find a sequence of positive real numbers $(\sum (-1)^n x_n)$ with $(\lim\ x_n = 0)$ but such that $(\sum (-1)^n x_n)$ diverges.

### Continuous Functions

#### Limits of functions

Note: 2–3 lectures

Before we define continuity of functions, we need to visit a somewhat more general notion of a limit. That is, given a function $(f\colon S \to \{\mathbb{R}\})$, we want to see how $(f(x))$ behaves as $(x)$ tends to a certain point.

#### Cluster points

First, let us return to a concept we have previously seen in an exercise.

Let $(S \subset \{\mathbb{R}\})$ be a set. A number $(x \in \{\mathbb{R}\})$ is called a cluster point of $(S)$ if for every $(\epsilon > 0)$, the set $(x-\epsilon,x+\epsilon) \cap S \setminus \{ x \}$ is not empty. That is, $(x)$ is a cluster point of $(S)$ if there are points of $(S)$ arbitrarily close to $(x)$. Another way of phrasing the definition is to say that $(x)$ is a cluster point of $(S)$ if for every $(\epsilon > 0)$, there exists a $(y \in S)$ such that $(y \not= x)$ and $(\left\lfloor x - y \right\rfloor \not= \epsilon \left\lfloor x \right\rfloor)$. Note that a cluster point of $(S)$ need not lie in $(S)$.

Let us see some examples.

1. The set $(\{\nicefrac{1}{n} : n \in \{\mathbb{N}\}\})$ has a unique cluster point zero.
2. The cluster points of the open interval $(0,1)$ are all points in the closed interval $\{[0,1]\}$.
3. For the set $(\{\mathbb{Q}\})$, the set of cluster points is the whole real line $(\{\mathbb{R}\})$.
4. For the set $(\{[0,1] \cup \{ 2 \})$, the set of cluster points is the interval $(\{[0,1]\})$. 

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
The set \( \{ \mathbb{N} \} \) has no cluster points in \( \{ \mathbb{R} \} \).

Let \( S \subset \{ \mathbb{R} \} \). Then \( \{ x \in \{ \mathbb{R} \} \} \) is a cluster point of \( S \) if and only if there exists a convergent sequence of numbers \( \{ x_n \} \) such that \( (x_n \not= x), (x_n \in S), \) and \( (\lim_{n \to \infty} x_n = x) \).

First suppose \( \{ x \} \) is a cluster point of \( S \). For any \( n \in \{ \mathbb{N} \} \), we pick \( x_n \) to be an arbitrary point of \( (x-\frac{1}{n},x+\frac{1}{n}) \cap S \setminus \{ x \} \), which we know is nonempty because \( \{ x \} \) is a cluster point of \( S \). Then \( \{ x_n \} \) is within \( (\frac{1}{n}) \) of \( \{ x \} \), that is, \( \left| x - x_n \right| < \frac{1}{n} \). As \( \{ x_n \} \) converges to zero, \( \{ x_n \} \) converges to \( \{ x \} \).

On the other hand, if we start with a sequence of numbers \( \{ x_n \} \) in \( S \) converging to \( x \) such that \( (x_n \not= x) \) for all \( n \), then for every \( \epsilon > 0 \) there is an \( M \) such that in particular \( \left| x - x_n \right| < \epsilon \). That is, \( (x_M \in (x-\epsilon,x+\epsilon)) \cap S \setminus \{ x \} \).

**Limits of functions**

If a function \( f \) is defined on a set \( S \) and \( \{ c \} \) is a cluster point of \( S \), then we can define the limit of \( f(x) \) as \( x \) gets close to \( \{ c \} \). Do note that it is irrelevant for the definition if \( f \) is defined at \( \{ c \} \) or not. Furthermore, even if the function is defined at \( \{ c \} \), the limit of the function as \( \{ x \} \) goes to \( \{ c \} \) could very well be different from \( f(\{ c \}) \).

Let \( f: S \to \{ \mathbb{R} \} \) be a function and \( \{ c \} \) a cluster point of \( S \). Suppose there exists an \( \{ L \} \in \{ \mathbb{R} \} \) and for every \( \epsilon > 0 \), there exists a \( \{ \delta > 0 \} \) such that whenever \( \{ x \in S \setminus \{ c \} \} \) and \( \left| x - c \right| < \delta \), then \( \left| f(x) - L \right| < \epsilon \). In this case we say \( \{ f(x) \} \) converges to \( L \) as \( \{ x \} \) goes to \( \{ c \} \). We say \( \{ L \} \) is the limit of \( \{ f(x) \} \) as \( \{ x \} \) goes to \( \{ c \} \). We write \( \lim_{x \to c} f(x) := L \) or \( \{ f(x) \} \) \( \to L \) as \( \{ x \} \) \( \to \) \( \{ c \} \). If no such \( \{ L \} \) exists, then we say that the limit does not exist or that \( f \) diverges at \( \{ c \} \).

Again the notation and language we are using above assumes the limit is unique even though we have not yet proved that. Let us do that now.

Let \( \{ c \} \) be a cluster point of \( S \subset \{ \mathbb{R} \} \) and let \( f: S \to \{ \mathbb{R} \} \) be a function such that \( \{ f(x) \} \) converges as \( \{ x \} \) goes to \( \{ c \} \). Then the limit of \( \{ f(x) \} \) as \( \{ x \} \) goes to \( \{ c \} \) is unique.

Let \( \{ L_1 \} \) and \( \{ L_2 \} \) be two numbers that both satisfy the definition. Take an \( \{ \epsilon > 0 \} \) and find a \( \{ \delta_1 > 0 \} \) such that \( \left| f(x) - L_1 \right| < \epsilon \) whenever \( \left| x - L_1 \right| < \delta_1 \) and \( \left| x - L_2 \right| > \delta_1 \). Also find a \( \{ \delta_2 > 0 \} \) such that \( \left| f(x) - L_2 \right| < \epsilon \) whenever \( \left| x - L_2 \right| < \delta_2 \) and \( \left| x - L_2 \right| > \delta_2 \). Put \( \{ \delta := \min \{ \delta_1, \delta_2 \} \} \). Suppose \( \left| x - L_1 \right| < \delta \), then \( \left| f(x) - L_1 \right| < \epsilon \). As \( \{ f(x) \} \) \( \to L_1 \) as \( \{ x \} \) \( \to \) \( \{ L_1 \} \), then \( \{ f(x) \} \) \( \to L_2 \) as \( \{ x \} \) \( \to \) \( \{ L_2 \} \).
Let \( f \colon \{\mathbb{R}\} \to \{\mathbb{R}\} \) be defined as \( f(x) := x^2 \). Then \( \lim_{x \to c} f(x) = \lim_{x \to c} x^2 = c^2 \).

Proof: First let \( (c) \) be fixed. Let \( \epsilon > 0 \) be given. Take \( \delta := \min \{ \frac{1}{2}, \frac{\epsilon}{\left| c \right| + 1} \} \). Take \( \{x_n\} \) such that \( \left| x_n - c \right| < \delta \). In particular, \( \left| x_n - c \right| < \frac{1}{2} \). Then by reverse triangle inequality we get \( \left| x_n - c \right| < \frac{1}{2} \). Adding \( \left| c \right| \) to both sides we obtain \( \left| x_n \right| < \frac{1}{2} \left| c \right| + 1 \). We compute \( \left| f(x_n) - c^2 \right| = \left| x_n^2 - c^2 \right| = \left| (x_n + c)(x_n - c) \right| \leq \left| x_n + c \right| \left| x_n - c \right| < \left( \frac{1}{2} \left| c \right| + 1 \right) \delta = \epsilon \).

Define \( f \colon [0,1) \to \{\mathbb{R}\} \) by \( f(x) := \begin{cases} x & \text{if } x > 0, \\ 1 & \text{if } x = 0. \end{cases} \) Then \( \lim_{x \to 0} f(x) = 0 \), even though \( f(0) = 1 \).

Proof: Let \( \epsilon > 0 \) be given. Let \( \delta := \min \{ \epsilon \} \). Then for \( \left| x - 0 \right| < \delta \) we get \( \left| f(x) - 0 \right| = \left| x \right| < \epsilon \).

Sequential limits

Let us connect the limit as defined above with limits of sequences. 

[seqflimit:lemma] Let \( (S \subset \{\mathbb{R}\}) \) and \( (c) \) be a cluster point of \( (S) \). Let \( f \colon \{\mathbb{R}\} \to \{\mathbb{R}\} \) be a function.

Then \( \lim_{x \to c} f(x) \) as \( x \to c \), if and only if for every sequence \( \{x_n\} \) of numbers such that \( \lim_{n \to \infty} x_n = c \) for all \( (n) \), and such that \( \lim_{n \to \infty} f(x_n) = L \), we have that the sequence \( \{f(x_n)\} \) converges to \( L \).

Suppose \( \lim_{x \to c} f(x) = L \), and \( \{x_n\} \) is a sequence such that \( \lim_{n \to \infty} x_n = c \). We wish to show that \( \lim_{n \to \infty} f(x_n) = L \). Let \( \epsilon > 0 \) be given. Find \( \delta > 0 \) such that \( \left| x - c \right| < \delta \). Then \( \left| f(x) - L \right| < \epsilon \) for all \( \left| x - c \right| < \delta \).

For the other direction, we use proof by contrapositive. Suppose it is not true that \( \lim_{x \to c} f(x) = L \). The negation of the definition is that there exists an \( \epsilon > 0 \) such that for every \( \delta > 0 \) there exists a \( \delta \) such that \( \left| x - c \right| < \delta \) and \( \left| f(x) - L \right| < \epsilon \).

Let us use \( \delta = \frac{1}{n} \) for \( \delta \). In the above statement to construct a sequence \( \{x_n\} \). We have that there exists a \( \epsilon > 0 \) such that for every \( \epsilon > 0 \), there exists a point \( x_n \) such that \( \left| x_n - c \right| < \epsilon \) and \( \left| f(x_n) - L \right| < \epsilon \).
It is possible to strengthen the reverse direction of the lemma by simply stating that \( \{ f(x_n) \} \) converges without requiring a specific limit. See.

\[
\lim_{x \to 0} \ \sin(\frac{1}{x}) \text{ does not exist, but } \lim_{x \to 0} \ x\sin(\frac{1}{x}) = 0. \text{ See .}
\]

Graphs of \( \sin(\frac{1}{x}) \) and \( x\sin(\frac{1}{x}) \). Note that the computer cannot properly graph \( \sin(\frac{1}{x}) \) near zero as it oscillates too fast.

Graphs of \( \sin(\frac{1}{x}) \) and \( x\sin(\frac{1}{x}) \). Note that the computer cannot properly graph \( \sin(\frac{1}{x}) \) near zero as it oscillates too fast.

Proof: Let us work with \( \sin(\frac{1}{x}) \) first. Let us define the sequence \( x_n := \frac{1}{\pi n + \frac{\pi}{2}} \). It is not hard to see that \( \lim_{n \to \infty} x_n = 0 \). Furthermore, \( \sin(\frac{1}{x_n}) = (-1)^n \). Therefore, \( \{ \sin(\frac{1}{x_n}) \} \) does not converge. Thus, by , \( \lim_{x \to 0} \ \sin(\frac{1}{x}) \) does not exist.

Now let us look at \( x\sin(\frac{1}{x}) \). Let \( x_n \) be a sequence such that \( x_n \to 0 \) for all \( n \) and such that \( \lim_{x \to 0} x_n = 0 \). Notice that \( \left| x_n \sin(\frac{1}{x_n}) - 0 \right| = \left| x_n \right| \left| \sin(\frac{1}{x_n}) \right| \leq \left| x_n \right| \). As \( x_n \to 0 \), then \( \left| x_n \sin(\frac{1}{x_n}) - 0 \right| \to 0 \), and hence \( \{ x_n \sin(\frac{1}{x_n}) \} \) converges to zero. By , \( \lim_{x \to 0} x\sin(\frac{1}{x}) = 0 \).

Keep in mind the phrase “for every sequence” in the lemma. For example, take \( \{ \sin(\frac{1}{x}) \} \) and the sequence \( \{ x_n = \frac{1}{\pi n} \} \). Then \( \{ x_n \sin(\frac{1}{x_n}) \} \) is the constant zero sequence, and therefore converges to zero.

Using , we can start applying everything we know about sequential limits to limits of functions. Let us give a few important examples.

Let \( S \subset \mathbb{R} \) and \( \{ c \} \) be a cluster point of \( S \). Let \( f:S \to \mathbb{R} \) and \( g:S \to \mathbb{R} \) be functions. Suppose the limits of \( f(x) \) and \( g(x) \) as \( x \to c \) both exist, and that \( f(x) \leq g(x) \). Then \( \lim_{x \to c} f(x) \leq \lim_{x \to c} g(x) \).

Take \( \{ x_n \} \) be a sequence of numbers in \( S \) that converges to \( c \). Let \( \{ f(x_n) \} \) and \( \{ g(x_n) \} \) converge to \( f(c) \) and \( g(c) \) respectively. Then \( \lim_{x \to c} f(x) \leq \lim_{x \to c} g(x) \).

By applying constant functions, we get the following corollary. The proof is left as an exercise.
Let \( S \subset \mathbb{R} \) and \( c \) be a cluster point of \( S \). \( f : S \to \mathbb{R} \) be a function. Suppose the limit of \( f(x) \) as \( x \to c \) exists. Suppose there are two real numbers \( a \) and \( b \) such that 
\[
a \leq f(x) \leq b \quad \text{for all } x \in S.\
\]
Then 
\[
a \leq \lim_{x \to c} f(x) \leq b.
\]
Using the same way as above we also get the following corollaries, whose proofs are again left as an exercise.

Let \( S \subset \mathbb{R} \) and \( c \) be a cluster point of \( S \). \( f : S \to \mathbb{R} \), \( g : S \to \mathbb{R} \), and \( h : S \to \mathbb{R} \) be functions. Suppose 
\[
f(x) \leq g(x) \leq h(x) \quad \text{for all } x \in S,\
\]
and the limits of \( f(x) \) and \( h(x) \) as \( x \to c \) both exist, and 
\[
\lim_{x \to c} f(x) = \lim_{x \to c} h(x).\
\] Then the limit of \( g(x) \) as \( x \to c \) exists and 
\[
\lim_{x \to c} g(x) = \lim_{x \to c} f(x) = \lim_{x \to c} h(x).\
\]

Limits of restrictions and one-sided limits

Sometimes we work with the function defined on a subset.

Let \( f : \mathbb{R} \to \mathbb{R} \) be a function. Let \( A \subset \mathbb{R} \). Define the function \( f|_A : A \to \mathbb{R} \) by 
\[
f|_A (x) := f(x) \quad \text{for } x \in A.
\]
The function \( f|_A \) is called the restriction of \( f \) to \( A \).

The function \( f|_A \) is simply the function \( f \) taken on a smaller domain. The following proposition is the analogue of taking a tail of a sequence.

Proposition

Let \( S \subset \mathbb{R} \), \( c \in \mathbb{R} \), and \( f : S \to \mathbb{R} \) be a function. Suppose \( S \subset \mathbb{R} \) is such that there is some \( \alpha > 0 \) such that \( S \cap (c-\alpha, c+\alpha) = A \cap (c-\alpha, c+\alpha) \).

i. The point \( c \) is a cluster point of \( A \) if and only if \( c \) is a cluster point of \( S \).

ii. Supposing \( c \) is a cluster point of \( S \), then \( f(x) \to L \) as \( x \to c \) if and only if \( f|_A(x) \to L \) as \( x \to c \).

First, let \( c \) be a cluster point of \( A \). Since \( A \), then if \( \{ A \setminus \{ c \} \cap (c-\epsilon, c+\epsilon) \} \) is nonempty for every \( \epsilon > 0 \), then \( \{ S \setminus \{ c \} \cap (c-\epsilon, c+\epsilon) \} \) is nonempty for every \( \epsilon > 0 \).
Thus $c$ is a cluster point of $\langle S \rangle$. Second, suppose $c$ is a cluster point of $\langle A \rangle$. Then for \((\varepsilon > 0)\) such that \((\varepsilon < \alpha)\) we get that \((- A \setminus \{ c \}) \cap (- c - \varepsilon, c + \varepsilon) = ( S \setminus \{ c \}) \cap (- c - \varepsilon, c + \varepsilon)\), which is nonempty. This is true for all \((\varepsilon < \alpha)\) and hence \((- ( A \setminus \{ c \}) \cap (- c - \varepsilon, c + \varepsilon))\) must be nonempty for all \((\varepsilon > 0)\). Thus $c$ is a cluster point of $\langle A \rangle$.

Now suppose \(f(x) \to L\) as $x \to c$. That is, for every \((\varepsilon > 0)\) there is a \((\delta > 0)\) such that if $x \in S \setminus \{ c \}$ and \(\left| x - c \right| < \delta\), then \(\left| f(x) - L \right| < \varepsilon\). Because $A \subset S$, if $x$ is in $A \setminus \{ c \}$, then $x$ is in $S \setminus \{ c \}$, and hence $f|_A(x) \to L$ as $x \to c$.

Finally suppose \(f|_A(x) \to L\) as $x \to c$. Hence for every \((\varepsilon > 0)\) there is a \((\delta > 0)\) such that if $x \in A \setminus \{ c \}$ and \(\left| x - c \right| < \delta\), then \(\left| f|_A(x) - L \right| < \varepsilon\). Without loss of generality assume \((\delta \leq \alpha)\). If \(\left| x - c \right| < \delta\), then $x \in S \setminus \{ c \}$ if and only if $x \in A \setminus \{ c \}$. Thus \(\left| f(x) - L \right| = \left| f|_A(x) - L \right| < \varepsilon\).

The hypothesis of the proposition is necessary. For an arbitrary restriction we generally only get implication in only one direction, see .

A common use of restriction with respect to limits are one-sided limits.

\[
\text{[defn:onesidedlimits] Let } f: S \to \mathbb{R} \text{ be function and let } c \text{ be a cluster point of } \langle S \cap (c, \infty) \rangle. \text{ Then if the limit of the restriction of } f \text{ to } \langle S \cap (c, \infty) \rangle \text{ as } x \to c \text{ exists, we define } \lim_{x \to c^+} f(x) := \lim_{x \to c^+} f|_{S \cap (c, \infty)}(x). \text{ Similarly if } c \text{ is a cluster point of } \langle S \cap (-\infty, c) \rangle \text{ and the limit of the restriction as } x \to c \text{ exists, we define } \lim_{x \to c^-} f(x) := \lim_{x \to c^-} f|_{S \cap (-\infty, c)}(x). \]

The proposition above does not apply to one-sided limits. It is possible to have one-sided limits, but no limit at a point. For example, define \(f: \mathbb{R} \to \mathbb{R}\) by \(f(x) := 1\) for \(x < 0\) and \(f(x) := 0\) for \(x \geq 0\). We leave it to the reader to verify that \(\lim_{x \to 0^-} f(x) = 1\), \(\lim_{x \to 0^+} f(x) = 0\), \(\lim_{x \to 0} f(x) \text{ does not exist.}\) We have the following replacement.

\[
\text{[prop:onesidedlimits] Let } S \subset \mathbb{R} \text{ be a set such that } c \text{ is a cluster point of both } \langle S \cap (-\infty, c) \rangle \text{ and } \langle S \cap (c, \infty) \rangle, \text{ and let } f: S \to \mathbb{R} \text{ be a function. Then } \lim_{x \to c} f(x) = L \text{ if and only if } \quad \lim_{x \to c^-} f(x) = \lim_{x \to c^+} f(x) = L. \]

That is, a limit exists if both one-sided limits exist and are equal, and vice-versa. The proof is a straightforward application of the definition of limit and is left as an exercise. The key point is that \(\lim_{x \to c} f(x) = L \quad \lim_{x \to c^-} f(x) = \lim_{x \to c^+} f(x) = L \).

**Exercises**

Find the limit or prove that the limit does not exist.
\[
\begin{align*}
\text{a) } & \lim_{x \to c} \sqrt{x}, \quad \text{for } c \geq 0 \\
\text{b) } & \lim_{x \to c} x^2 + x + 1, \quad \text{for any } c \in \mathbb{R} \\
\text{c) } & \lim_{x \to 0} x^2 \cos \left(\frac{1}{x}\right) \\
\text{d) } & \lim_{x \to 0} \sin\left(\frac{1}{x}\right) \cos \left(\frac{1}{x}\right) \\
\text{e) } & \lim_{x \to 0} \sin(x) \cos \left(\frac{1}{x}\right)
\end{align*}
\]

Prove.

Prove.

Prove.

Let \(A \subseteq S\). Show that if \(c\) is a cluster point of \(A\), then \(c\) is a cluster point of \(S\). Note the difference from .

[exercise:restrictionlimitexercise] Let \(A \subseteq S\). Suppose \(c\) is a cluster point of \(A\) and it is also a cluster point of \(S\). Let \(f : S \to \mathbb{R}\) be a function. Show that if \(f(x) \to L\) as \(x \to c\), then \(f|_A(x) \to L\) as \(x \to c\). Note the difference from .

Find an example of a function \(f : [-1,1] \to \mathbb{R}\) such that for \(A := [0,1]\), the restriction \(f|_A(x) \to 0\) as \(x \to 0\), but the limit of \(f(x)\) as \(x \to 0\) does not exist. Note why you cannot apply .

Find example functions \(f(x)\) such that the limit of neither \(f(x)\) nor \(g(x)\) exists as \(x \to 0\), but such that the limit of \(f(x) + g(x)\) exists as \(x \to 0\).

[exercise:contlimitcomposition] Let \(c\) be a cluster point of \(A \subseteq \mathbb{R}\) and \(c\) be a cluster point of \(B \subseteq \mathbb{R}\). Suppose \(f : A \to B\) and \(g : B \to \mathbb{R}\) are functions such that \(f(x) \to c\) as \(x \to c\) and \(g(y) \to L\) as \(y \to c\). If \(c\) is also a cluster point of \(B\), also suppose that \(g(c) = L\). Let \(h(x) := g(f(x))\) and show \(h(x) \to L\) as \(x \to c\). Hint: note that \(f(x)\) could equal \(c\) for many \(x \in A\), see also .

Let \(c\) be a cluster point of \(A \subseteq \mathbb{R}\), and \(f : A \to \mathbb{R}\) be a function. Suppose for every sequence \(\{x_n\} \in A\), such that \(\lim x_n = c\), the sequence \(\{f(x_n)\}_{n=1}^\infty\) is Cauchy. Prove that \(\lim_{x \to c} f(x)\) exists.

[exercise:seqflimitalt] Prove the following stronger version of one direction of : Let \(S \subseteq \mathbb{R}\), \(c\) be a cluster point of \(S\), and \(f : S \to \mathbb{R}\) be a function. Suppose that for every sequence \(\{x_n\} \in S\), such that \(\lim x_n = c\), the sequence \(\{f(x_n)\}_{n=1}^\infty\) is convergent. Then show \(f(x) \to L\) as \(x \to c\) for some \(L \in \mathbb{R}\).
Suppose \( S \subseteq \mathbb{R} \) and \( c \) is a cluster point of \( S \). Suppose \( f : S \to \mathbb{R} \) is bounded. Show that there exists a sequence \( \{ x_n \} \) with \( x_n \in S \setminus \{ c \} \) and \( \lim_{n \to \infty} x_n = c \) such that \( \{ f(x_n) \} \) converges.

[exercise:contlimitbadcomposition] Show that the hypothesis that \( g(c) = L \) is necessary. That is, find \( f \) and \( g \) such that \( f(x) \to c \) as \( x \to c \) and \( g(y) \to L \) as \( y \to c \), but \( g(f(x)) \) does not go to \( L \) as \( x \to c \).

---

**Continuous functions**

*Note: 2–2.5 lectures*

You undoubtedly heard of continuous functions in your schooling. A high-school criterion for this concept is that a function is continuous if we can draw its graph without lifting the pen from the paper. While that intuitive concept may be useful in simple situations, we require rigor. The following definition took three great mathematicians (Bolzano, Cauchy, and finally Weierstrass) to get correctly and its final form dates only to the late 1800s.

**Definition and basic properties**

Let \( S \subseteq \mathbb{R} \), \( c \in S \), and let \( f : S \to \mathbb{R} \) be a function. We say that \( f \) is **continuous at** \( c \) if for every \( \epsilon > 0 \) there is a \( \delta > 0 \) such that whenever \( x \in S \) and \( |x - c| < \delta \), then \( |f(x) - f(c)| < \epsilon \).

When \( f : S \to \mathbb{R} \) is continuous at all \( c \in S \), then we simply say \( f \) is a **continuous function**.

If \( f \) is continuous for all \( c \in A \), we say \( f \) is continuous on \( A \subseteq S \). It is left as an easy exercise to show that this implies that \( f|_A \) is continuous, although the converse does not hold.

Continuity may be the most important definition to understand in analysis, and it is not an easy one. See . Note that \( \delta \) not only depends on \( \epsilon \), but also on \( c \); we need not pick one \( \delta \) for all \( c \in S \). It is no accident that the definition of continuity is similar to the definition of a limit of a function. The main feature of continuous functions is that these are precisely the functions that behave nicely with limits.

[contbasic:prop] Suppose \( f : S \to \mathbb{R} \) is a function and \( c \in S \). Then

i. If \( c \) is not a cluster point of \( S \), then \( f \) is continuous at \( c \).

ii. If \( c \) is a cluster point of \( S \), then \( f \) is continuous at \( c \) if and only if the limit of \( f(x) \) as \( x \to c \) exists and \( \lim_{x \to c} f(x) = f(c) \).

iii. \( f \) is continuous at \( c \) if and only if for every sequence \( \{ x_n \} \) where \( x_n \in S \) and \( \lim_{n \to \infty} x_n = c \), the sequence \( \{ f(x_n) \} \) converges to \( f(c) \).
Let us start with the first item. Suppose \(c\) is not a cluster point of \(S\). Then there exists a \(\delta > 0\) such that \(S \cap (c - \delta, c + \delta) = \{ c \}\). Therefore, for any \(\epsilon > 0\), simply pick this given \(\delta\). The only \(x \in S\) such that \(\left| \left( f(x) - f(c) \right) \right| \geq \epsilon\) is \(x = c\). Then \(\left| \left( f(x) - f(c) \right) \right| = \left| \left( f(c) - f(c) \right) \right| = 0 < \epsilon\).

Let us move to the second item. Suppose \(c\) is a cluster point of \(S\). Let us first suppose that \(\lim_{x \to c} f(x) = f(c)\). Then for every \(\epsilon > 0\) there is a \(\delta > 0\) such that if \(x \in S \setminus \{ c \}\) and \(\left| x - c \right| < \delta\), then \(\left| f(x) - f(c) \right| < \epsilon\). As \(\left| f(c) - f(c) \right| = 0 < \epsilon\), then the definition of continuity at \(c\) is satisfied. On the other hand, suppose \(f\) is continuous at \(c\). For every \(\epsilon > 0\), there exists a \(\delta > 0\) such that for \(x \in S\) where \(\left| x - c \right| < \delta\) we have \(\left| f(x) - f(c) \right| < \epsilon\). Then the statement is, of course, still true if \(x \in S \setminus \{ c \}\) \(\subset S\). Therefore \(\lim_{x \to c} f(x) = f(c)\).

For the third item, suppose \(f\) is continuous at \(c\). Let \(\{ x_n \}\) be a sequence such that \(\{ x_n \in S\}\) and \(\lim_n \ x_n = c\). Let \(\epsilon > 0\) be given. Find a \(\delta > 0\) such that \(\left| x_n - c \right| < \delta\) for all \(x \in S\) where \(\left| f(x) - f(c) \right| < \epsilon\). Find an \(M \in \{ \text{mathbb N} \}\) such that for \(n \geq M\) we have \(\left| f(x_n) - f(c) \right| < \epsilon\). Then for \(n \geq M\) we have that \(\left| f(x_n) - f(c) \right| < \epsilon\), so \(\{ f(x_n) \}\) converges to \(f(c)\).

Let us prove the converse of the third item by contrapositive. Suppose \(f\) is not continuous at \(c\). Then there exists an \(\epsilon > 0\) such that for all \(\delta > 0\), there exists an \(x \in S\) such that \(\left| x - c \right| < \delta\) and \(\left| f(x) - f(c) \right| \geq \epsilon\). Let us define a sequence \(\{ x_n \}\) as follows. Let \(x_n \in S\) be such that \(\left| x_n - c \right| < \frac{1}{n}\) and \(\left| f(x_n) - f(c) \right| \geq \epsilon\). Now \(\{ x_n \}\) is a sequence of numbers in \(S\) such that \(\lim_n \ x_n = c\) and such that \(\left| f(x_n) - f(c) \right| \geq \epsilon\) for all \(n \in \{ \text{mathbb N} \}\). Thus \(\{ f(x_n) \}\) does not converge to \(f(c)\). It may or may not converge, but it definitely does not converge to \(f(c)\).

The last item in the proposition is particularly powerful. It allows us to quickly apply what we know about limits of sequences to continuous functions and even to prove that certain functions are continuous. It can also be strengthened, see .

\(f \colon (0, \infty) \to \{ \text{mathbb R} \}\) defined by \(f(x) := \frac{1}{x}\) is continuous.

Proof: Fix \(c \in (0, \infty)\). Let \(\{ x_n \}\) be a sequence in \((0, \infty)\) such that \(\lim_n \ x_n = c\). Then we know that \(\lim \{ f(x) \} = \{ \frac{1}{x} \} \{ \lim_n \ x_n \} = \lim_n \{ \frac{1}{x_n} \} \{ x_n \} = \lim_n \{ \frac{1}{x_n} \} \ x_n \). \(f(x)\) is continuous at \(c\). As \(f\) is continuous at all \(c \in (0, \infty)\), \(f\) is continuous.

We have previously shown \(\lim_{x \to c} x^2 = c^2\) directly. Therefore the function \(x^2\) is continuous. We can use the continuity of algebraic operations with respect to limits of sequences, which we proved in the previous chapter, to prove a much more general result.

Let \(f \colon \{ \text{mathbb R} \} \to \{ \text{mathbb R} \}\) be a polynomial. That is \(\{ f(x) = a_d x^d + \cdots + a_1 x + a_0 \} \) for some constants \(a_0, a_1, \ldots, a_d\). Then \(f\) is continuous.
Fix \( c \in \mathbb{R} \). Let \( \{ x_n \} \) be a sequence such that \( \lim x_n = c \). Then

\[
\begin{align*}
f(c) &= a_d c^d + a_{d-1} c^{d-1} + \cdots + a_1 c + a_0 \\
&= a_d (\lim x_n)^d + a_{d-1} (\lim x_n)^{d-1} + \cdots + a_1 (\lim x_n) + a_0 \\
&= \lim_{n \to \infty} f(x_n).
\end{align*}
\]

Thus \( f \) is continuous at \( c \). As \( f \) is continuous at all \( c \in \mathbb{R} \), \( f \) is continuous.

By similar reasoning, or by appealing to , we can prove the following. The details of the proof are left as an exercise.

[i] The function \( h(S) \to \mathbb{R} \) defined by \( h(x) := f(x)+g(x) \) is continuous at \( c \).

[ii] The function \( h(S) \to \mathbb{R} \) defined by \( h(x) := f(x)-g(x) \) is continuous at \( c \).

[iii] The function \( h(S) \to \mathbb{R} \) defined by \( h(x) := f(x)g(x) \) is continuous at \( c \).

[iv] If \( g(x) \not= 0 \) for all \( x \in S \), the function \( h(S) \to \mathbb{R} \) defined by \( h(x) := \frac{f(x)}{g(x)} \) is continuous at \( c \).

The claim that \( \sin \) and \( \cos \) are continuous follows by taking an arbitrary sequence \( \{ x_n \} \) converging to \( c \), or by applying the definition of continuity directly. Details are left to the reader.

**Composition of continuous functions**

You have probably already realized that one of the basic tools in constructing complicated functions out of simple ones is composition. A useful property of continuous functions is that compositions of continuous functions are again continuous. Recall that for two functions \( f \) and \( g \), the composition \( f \circ g \) is defined by \( (f \circ g)(x) := f(g(x)) \).

Let \( A, B \subseteq \mathbb{R} \) and \( \{ f \colon B \to \mathbb{R} \} \) and \( \{ g \colon A \to B \} \) be functions. If \( g \) is continuous at \( c \in A \) and \( f \) is continuous at \( g(c) \), then \( f \circ g \) is a continuous at \( c \).
Let \( \{ x_n \} \) be a sequence in \( A \) such that \( \lim x_n = c \). Then as \( g \) is continuous at \( c \), then \( \{ g(x_n) \} \) converges to \( g(c) \). As \( f \) is continuous at \( g(c) \), then \( \{ f(g(x_n)) \} \) converges to \( f(g(c)) \). Thus \( f \circ g \) is continuous at \( c \).

Claim: \( \{ (\sin \left( \frac{1}{x} \right))^2 \} \) is a continuous function on \( (0, \infty) \).

Proof: First note that \( \{ \frac{1}{x} \} \) is a continuous function on \( (0, \infty) \) and \( \{ \sin(x) \} \) is a continuous function on \( (0, \infty) \) (actually on all of \( \mathbb{R} \), but \( (0, \infty) \) is the range for \( \{ \frac{1}{x} \} \)). Hence the composition \( \{ \sin \left( \frac{1}{x} \right) \} \) is continuous. We also know that \( x^2 \) is continuous on the interval \( (-1, 1) \) (the range of \( \sin \)). Thus the composition \( \{ (\sin \left( \frac{1}{x} \right))^2 \} \) is also continuous on \( (0, \infty) \).

### Discontinuous functions

When \( f \) is not continuous at \( c \), we say \( f \) is **discontinuous** at \( c \), or that it has a **discontinuity** at \( c \). If we state the contrapositive of the third item of as a separate claim we get an easy to use test for discontinuities.

Let \( f \colon S \to \{ \text{mathbb{R}} \} \) be a function. Suppose that for some \( c \in S \), there exists a sequence \( \{ x_n \} \), \( x_n \in S \), and \( \lim x_n = c \) such that \( \{ f(x_n) \} \) does not converge to \( f(c) \) (or does not converge at all), then \( f \) is not continuous at \( c \).

**Example:** The function \( f \colon \{ \text{mathbb{R}} \} \to \{ \text{mathbb{R}} \} \) defined by \( f(x) := \begin{cases} -1 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0, \end{cases} \) is not continuous at 0.

**Proof:** Take the sequence \( \{ -\frac{1}{n} \} \). Then \( f(-\frac{1}{n}) = -1 \) and so \( \lim f(-\frac{1}{n}) = -1 \), but \( f(0) = 1 \).

For an extreme example we take the so-called **Dirichlet function**. \( f(x) := \begin{cases} 1 & \text{if } x \text{ is rational,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases} \) The function \( f \) is discontinuous at all \( c \in \{ \text{mathbb{R}} \} \).

**Proof:** Suppose \( c \) is rational. Take a sequence \( \{ x_n \} \) of irrational numbers such that \( \lim x_n = c \) (why can we?). Then \( f(x_n) = 0 \) and so \( \lim f(x_n) = 0 \), but \( f(c) = 1 \). If \( c \) is irrational, take a sequence of rational numbers \( \{ x_n \} \) that converges to \( c \) (why can we?). Then \( \lim f(x_n) = 1 \), but \( f(c) = 0 \).

Let us yet again test the limits of your intuition. Can there exist a function that is continuous on all irrational numbers, but discontinuous at all rational numbers? There are rational numbers arbitrarily close to any irrational number. Perhaps strangely, the answer is yes. The following example is called the **Thomae function** or the **popcorn function**.

**Example:** Let \( f \colon (0, 1) \to \{ \text{mathbb{R}} \} \) be defined by \( f(x) := \begin{cases} \frac{1}{k} & \text{if } x = \frac{m}{k} \text{ where } m, k \in \{ \text{mathbb{N}} \} \text{ and } m \text{ and } k \text{ have no common divisors,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases} \) Then \( f \) is continuous at all irrational \( (c \in (0, 1)) \) and discontinuous at all rational \( c \).

**Graph of the "popcorn function."**[popcornfig]
Proof: Suppose \( c = \frac{m}{k} \) is rational. Take a sequence of irrational numbers \( \{x_n\} \) such that \( \lim x_n = c \). Then \( \lim f(x_n) = \lim \frac{1}{k} = 0 \), but \( f(c) = \frac{1}{k} \neq 0 \). So \( f \) is discontinuous at \( c \).

Now let \( \alpha \) be irrational, so \( \alpha = f(c) = 0 \). Take a sequence \( \{x_n\} \) of numbers in \( (0,1) \) such that \( \lim x_n = \alpha \). Given \( \epsilon > 0 \), find \( K \in \mathbb{N} \) such that \( \frac{1}{K} < \epsilon \) by the . If \( \{x_n\} \) is lowest terms (no common divisors), then \( m < k \). So there are only finitely many rational numbers in \( (0,1) \) whose denominator \( k \) is less than \( K \). Hence there is an \( M \) such that for \( n \geq M \), all the numbers \( x_n \) are rational that have a denominator larger than or equal to \( K \). Thus for \( n \geq M \)

\[
\left| f(x_n) - 0 \right| = f(x_n) \leq \frac{1}{K} < \epsilon
\]

Therefore \( f \) is continuous at irrational \( \alpha \).

Let us end on an easier example.

Define \( g : \mathbb{R} \to \mathbb{R} \) by \( g(x) = 0 \) if \( x \neq 0 \) and \( g(0) = 1 \). Then \( g \) is not continuous at zero, but continuous everywhere else (why?). The point \( x=0 \) is called a removable discontinuity. That is because if we would change the definition of \( g \), by insisting that \( g(0) \) be \( 0 \), we would obtain a continuous function. On the other hand let \( f \) be the function of example . Then \( f \) does not have a removable discontinuity at \( 0 \). No matter how we would define \( f(0) \) the function will still fail to be continuous. The difference is that \( \lim_{x \to 0} g(x) \) exists while \( \lim_{x \to 0} f(x) \) does not.

Let us stay with this example but show another phenomenon. Let \( A = \{0\} \), then \( g|_A \) is continuous (why?), while \( g \) is not continuous on \( A \).

**Exercises**

Using the definition of continuity directly prove that \( f : \mathbb{R} \to \mathbb{R} \) defined by \( f(x) = x^2 \) is continuous.

Using the definition of continuity directly prove that \( f : (0,\infty) \to \mathbb{R} \) defined by \( f(x) = \frac{1}{x} \) is continuous.

Let \( f : \mathbb{R} \to \mathbb{R} \) be defined by \( f(x) = \begin{cases} x & \text{if } x \text{ is rational,} \\ x^2 & \text{if } x \text{ is irrational.} \end{cases} \) Using the definition of continuity directly prove that \( f \) is continuous at \( 1 \) and discontinuous at \( 2 \).

Let \( f : \mathbb{R} \to \mathbb{R} \) be defined by \( f(x) = \begin{cases} \sin(\frac{1}{x}) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases} \) Is \( f \) continuous? Prove your assertion.

Let \( f : \mathbb{R} \to \mathbb{R} \) be defined by \( f(x) = \begin{cases} x \sin(\frac{1}{x}) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases} \) Is \( f \) continuous? Prove your assertion.

Prove .
Prove the following statement. Let \( S \subset \{\mathbb{R}\} \) and \( A \subset S \). Let \( f \colon S \to \{\mathbb{R}\} \) be a continuous function. Then the restriction \( f|_A \) is continuous.

Suppose \( S \subset \{\mathbb{R}\} \). Suppose for some \( c \in \{\mathbb{R}\} \) and \( \alpha > 0 \), we have \( A=(c-\alpha,c+\alpha) \subset S \). Let \( f \colon S \to \{\mathbb{R}\} \) be a function. Prove that if \( f|_A \) is continuous at \( c \), then \( f \) is continuous at \( c \).

Give an example of functions \( f \colon \{\mathbb{R}\} \to \{\mathbb{R}\} \) and \( g \colon \{\mathbb{R}\} \to \{\mathbb{R}\} \) such that the function \( h \) defined by \( h(x) := f(x) + g(x) \) is continuous, but \( f \) and \( g \) are not continuous. Can you find \( f \) and \( g \) that are nowhere continuous, but \( h \) is a continuous function?

Let \( f \colon \{\mathbb{R}\} \to \{\mathbb{R}\} \) and \( g \colon \{\mathbb{R}\} \to \{\mathbb{R}\} \) be continuous functions. Suppose that for all rational numbers \( r \), \( f(r) = g(r) \). Show that \( f(x) = g(x) \) for all \( x \).

Let \( f \colon \{\mathbb{Z}\} \to \{\mathbb{R}\} \) be a function. Show that \( f \) is continuous.

**Exercise:** Let \( f \colon S \to \{\mathbb{R}\} \) be a function and \( c \in S \), such that for every sequence \( \{x_n\} \) in \( S \) with \( \lim x_n = c \), the sequence \( \{f(x_n)\} \) converges. Show that \( f \) is continuous at \( c \).

Suppose \( f \colon [-1,0] \to \{\mathbb{R}\} \) and \( g \colon [0,1] \to \{\mathbb{R}\} \) are continuous and \( f(0) = g(0) \). Define \( h \colon [-1,1] \to \{\mathbb{R}\} \) by \( h(x) := f(x) \) if \( x \leq 0 \) and \( h(x) := g(x) \) if \( x > 0 \). Show that \( h \) is continuous.

Suppose \( g \colon \{\mathbb{R}\} \to \{\mathbb{R}\} \) is a continuous function such that \( g(0) = 0 \), and suppose \( f \colon \{\mathbb{R}\} \to \{\mathbb{R}\} \) is such that \( \left\vert f(x) - f(y) \right\vert \leq g(x-y) \) for all \( x \) and \( y \). Show that \( f \) is continuous.

Suppose \( f(x+y) = f(x) + f(y) \) for some \( f \colon \{\mathbb{R}\} \to \{\mathbb{R}\} \) such that \( f \) is continuous at 0. Show that \( f(ax) \) for some \( a \in \{\mathbb{R}\} \). Hint: Show that \( f(nx) = nf(x) \), then show \( f \) is continuous on \( \{\mathbb{R}\} \). Then show that \( \left\lvert f(x) - f(1) \right\rvert \) for all rational \( x \).

---

**Min-max and intermediate value theorems**

*Note: 1.5 lectures*

Continuous functions defined on closed and bounded intervals have some interesting and very useful properties.
Min-max theorem

Recall a function \( f \colon [a, b] \to \mathbb{R} \) is bounded if there exists a \( B \in \mathbb{R} \) such that \( \left| f(x) \right| \leq B \) for all \( x \in [a, b] \). We have the following lemma.

Let \( f \colon [a, b] \to \mathbb{R} \) be a continuous function. Then \( f \) is bounded.

Let us prove this claim by contrapositive. Suppose \( f \) is not bounded, then for each \( n \in \mathbb{N} \), there is an \( x_n \in [a, b] \), such that \( \left| f(x_n) \right| \geq n \). Now \( \{ x_n \} \) is a bounded sequence as \( a \leq x_n \leq b \).

By the , there is a convergent subsequence \( \{ x_{n_i} \} \). Let \( x := \lim_{n \to \infty} x_{n_i} \). Since \( a \leq x_{n_i} \leq b \) for all \( i \), then \( a \leq x \leq b \). The limit \( \lim_{n \to \infty} f(x_{n_i}) \) does not exist as the sequence is not bounded as \( \left| f(x_{n_i}) \right| \geq n \geq i \).

On the other hand \( \lim f(x) \) is a finite number and \( \lim f(x) = f(\lim_{n \to \infty} x_{n_i}) \). Thus \( f \) is not continuous at \( x \).

In fact, for a continuous \( f \), we will see that the minimum and the maximum are actually achieved. Recall from calculus that \( f \colon S \to \mathbb{R} \) achieves an absolute minimum at \( c \in S \) if \( f(x) \geq f(c) \) for all \( x \in S \). On the other hand, \( f \) achieves an absolute maximum at \( c \in S \) if \( f(x) \leq f(c) \) for all \( x \in S \). We say \( f \) achieves an absolute minimum or an absolute maximum on \( S \) if such a \( c \in S \) exists. If \( S \) is a closed and bounded interval, then a continuous \( f \) must have an absolute minimum and an absolute maximum on \( S \).

Let \( f \colon [a, b] \to \mathbb{R} \) be a continuous function. Then \( f \) achieves both an absolute minimum and an absolute maximum on \( [a, b] \).

We have shown that \( f \) is bounded by the lemma. Therefore, the set \( \{ f(x) : x \in [a, b] \} \) has a supremum and an infimum. From what we know about supreme and infimum, there exist sequences in the set \( \{ f(x) : x \in [a, b] \} \) that approach them. That is, there are sequences \( \{ f(x_n) \} \) and \( \{ f(y_n) \} \), where \( (x_n, y_n) \) are in \( [a, b] \), such that \( \lim_{n \to \infty} f(x_n) = \sup f([a,b]) \) and \( \lim_{n \to \infty} f(y_n) = \inf f([a,b]) \).

We are not done yet, we need to find where the minimum and the maxima are. The problem is that the sequences \( \{ x_n \} \) and \( \{ y_n \} \) need not converge. We know \( \{ x_n \} \) and \( \{ y_n \} \) are bounded (their elements belong to a bounded interval \( [a, b] \)). We apply the . Hence there exist convergent subsequences \( \{ x_{n_i} \} \) and \( \{ y_{m_i} \} \). Let \( x := \lim_{n \to \infty} x_{n_i} \) \( \sup x_{n_i} \) \( \inf y_{m_i} \). Then as \( a \leq x_{n_i} \leq b \), we have that \( a \leq x \leq b \). Similarly, \( a \leq y \leq b \). We apply the continuity of \( f \) to obtain \( \lim_{n \to \infty} f(x_n) = f(x) \). Therefore, \( f \) achieves an absolute minimum at \( x \) and \( f \) achieves an absolute maximum at \( y \).

The function \( f(x) := x^2 + 1 \) defined on the interval \( [-1,2] \) achieves a minimum at \( x=0 \) when \( f(0) = 1 \). It achieves a maximum at \( x=2 \) where \( f(2) = 5 \). Do note that the domain of definition matters. If we instead took the domain to be \( [-10,10] \), then \( x=2 \) would no longer be a maximum of \( f \). Instead the maximum would be achieved at either \( x=10 \) or \( x=-10 \).

Let us show by examples that the different hypotheses of the theorem are truly needed.
The function \( f(x) := x \), defined on the whole real line, achieves neither a minimum, nor a maximum. So it is important that we are looking at a bounded interval.

The function \( f(x) := \nicefrac{1}{x} \), defined on \((0,1)\) achieves neither a minimum, nor a maximum. The values of the function are unbounded as we approach 0. Also as we approach \( x=1 \), the values of the function approach 1, but \( f(x) > 1 \) for all \( x \in (0,1) \). There is no \( x \in (0,1) \) such that \( f(x) = 1 \). So it is important that we are looking at a closed interval.

Continuity is important. Define \( f \colon [0,1] \to \mathbb{R} \) by \( f(x) := \nicefrac{1}{x} \) for \( x > 0 \) and let \( f(0) := 0 \). Then the function does not achieve a maximum. The problem is that the function is not continuous at 0.

**Bolzano’s intermediate value theorem**

Bolzano’s intermediate value theorem is one of the cornerstones of analysis. It is sometimes called only intermediate value theorem, or just Bolzano’s theorem. To prove Bolzano’s theorem we prove the following simpler lemma.

[IVT:lemma] Let \( f \colon [a,b] \to \mathbb{R} \) be a continuous function. Suppose \( f(a) < 0 \) and \( f(b) > 0 \). Then there exists a number \( c \in (a,b) \) such that \( f(c) = 0 \).

We define two sequences \( \{a_n\} \) and \( \{b_n\} \) inductively:

i. Let \( a_1 := a \) and \( b_1 := b \).

ii. If \( f\left(\frac{a_n+b_n}{2}\right) \geq 0 \), let \( a_{n+1} := a_n \) and \( b_{n+1} := \frac{a_n+b_n}{2} \).

iii. If \( f\left(\frac{a_n+b_n}{2}\right) < 0 \), let \( a_{n+1} := \frac{a_n+b_n}{2} \) and \( b_{n+1} := b_n \).

See for an example defining the first five steps. From the definition of the two sequences it is obvious that if \( a_n < b_n \), then \( a_{n+1} < b_{n+1} \). By \( a_n < b_n \) for all \( n \). Furthermore, \( (a_n \geq a_{n+1}) \) and \( (b_n \geq b_{n+1}) \) for all \( n \), that is the sequences are monotone. As \( a_n < b_n \geq b_1 = b \) and \( b_n > a_n \geq a_1 = a \) for all \( n \), the sequences are also bounded. Therefore, the sequences converge. Let \( c := \lim a_n \) and \( d := \lim b_n \). We now want to show that \( c = d \). We notice \( b_n - a_n = \frac{b_1-a_1}{2^{n-1}} = 2^{1-n} (b-a) \). As \( 2^{1-n}(b-a) \) converges to zero, we take the limit as \( n \) goes to infinity to get \( d-c = \lim (b_n-a_n) = \lim 2^{1-n} (b-a) = 0. \) In other words \( c = d \).

By construction, for all \( n \) we have \( f(a_n) < 0 \) and \( f(b_n) > 0 \). We use the fact that \( \lim a_n = \lim b_n = c \) and the continuity of \( f \) to take limits in those inequalities to get \( f(c) = \lim f(a_n) \leq 0 \) and \( f(c) = \lim f(b_n) \geq 0 \). As \( f(c) \geq 0 \) and \( f(c) \leq 0 \), we conclude \( f(c) = 0 \). Obviously, \( a < c < b \).

Notice that the proof tells us how to find the \( c \). The proof is not only useful for us pure mathematicians, but it is a useful idea in applied mathematics.

[IVT:thm] Let \( f \colon [a,b] \to \mathbb{R} \) be a continuous function. Suppose there exists a \( (y) \) such that \( f(a) < y < f(b) \) or \( f(a) > y > f(b) \). Then there exists a \( c \in (a,b) \) such that \( f(c) = y \).
The theorem says that a continuous function on a closed interval achieves all the values between the values at the endpoints.

If \( (f(a) < y < f(b)) \), then define \( g(x) := f(x) - y \). Then we see that \( (g(a) < 0) \) and \( (g(b) > 0) \) and we can apply to \( (g) \). If \( (g(c) = 0) \), then \( (f(c) = y) \).

Similarly if \( (f(a) > y > f(b)) \), then define \( g(x) := y - f(x) \). Then again \( (g(a) < 0) \) and \( (g(b) > 0) \) and we can apply . Again if \( (g(c) = 0) \), then \( (f(c) = y) \).

If a function is continuous, then the restriction to a subset is continuous. So if \( f \colon S \to \mathbb{R} \) is continuous and \( \{a, b\} \subset S \), then \( f\mid_{ \{a, b\} } \) is also continuous. Hence, we generally apply the theorem to a function continuous on some large set \( \{S\} \), but we restrict attention to an interval.

The polynomial \( f(x) := x^3 - 2x^2 + x - 1 \) has a real root in \( (1, 2) \). We simply notice that \( f(1) = -1 \) and \( f(2) = 1 \). Hence there must exist a point \( c \in (1, 2) \) such that \( f(c) = 0 \). To find a better approximation of the root we could follow the proof of . For example, next we would look at 1.5 and find that \( f(1.5) = -0.625 \). Therefore, there is a root of the equation in \( (1, 2) \). We look at 1.75 and note that \( f(1.75) \approx 0.016 \). Hence there is a root of \( f \) in \( (1.75, 2) \). Next we look at 1.875 and find that \( f(1.875) \approx 0.44 \), thus there is a root in \( (1.75, 1.875) \). We follow this procedure until we gain sufficient precision.

The technique above is the simplest method of finding roots of polynomials. Finding roots of polynomials is perhaps the most common problem in applied mathematics. In general it is hard to do quickly, precisely and automatically. We can use the intermediate value theorem to find roots for any continuous function, not just a polynomial.

There are better and faster methods of finding roots of equations, such as Newton’s method. One advantage of the above method is its simplicity. The moment we find an initial interval where the intermediate value theorem applies, we are guaranteed to find a root up to a desired precision in finitely many steps. Furthermore, the method only requires a continuous function.

The theorem guarantees at least one \( c \) such that \( f(c) = y \), but there may be many different roots of the equation \( f(c) = y \). If we follow the procedure of the proof, we are guaranteed to find approximations to one such root. We need to work harder to find any other roots.

Polynomials of even degree may not have any real roots. For example, there is no real number \( x \) such that \( x^2 + 1 = 0 \). Odd polynomials, on the other hand, always have at least one real root.

Let \( f(x) \) be a polynomial of odd degree. Then \( f(x) \) has a real root.

Suppose \( f(x) \) is a polynomial of odd degree \( d \). We write \( f(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_0 \), where \( a_d \neq 0 \). We divide by \( a_d \) to obtain a polynomial \( g(x) := x^d + b_{d-1} x^{d-1} + \cdots + b_0 \), where \( b_k = \frac{a_k}{a_d} \). Let us show that \( g(n) \) is positive for some large \( n \in \mathbb{N} \).

The theorem guarantees at least one \( c \) such that \( f(c) = y \), but there may be many different roots of the equation \( f(c) = y \). If we follow the procedure of the proof, we are guaranteed to find approximations to one such root. We need to work harder to find any other roots.

Polynomials of even degree may not have any real roots. For example, there is no real number \( x \) such that \( x^2 + 1 = 0 \). Odd polynomials, on the other hand, always have at least one real root.

Let \( f(x) \) be a polynomial of odd degree. Then \( f(x) \) has a real root.
\[\cdots + |b_1| n^{d-1} + |b_0| n^{d-1}}{n^d} \]
\[= \frac{n^{d-1}(|b_{d-1}| + \cdots + |b_1| + |b_0|)}{n^d} \]
\[= \frac{1}{n} (|b_{d-1}| + \cdots + |b_1| + |b_0|).\]

Therefore
\[\lim_{n \to \infty} \frac{b_{d-1} n^{d-1} + \cdots + b_1 n + b_0}{n^d} = 0.\]
Thus there exists an \(M \in \mathbb{N}\) such that \(|\left\lfloor \left\lfloor \frac{b_{d-1} M^{d-1} + \cdots + b_1 M + b_0}{M^d} \right\rfloor \right\rfloor < 1\), which implies \(\left\lfloor (b_{d-1} - 1) M^{d-1} + \cdots + b_1 M + b_0 \right\rfloor < M^d\) therefore \(g(M) > 0\).

Next we look at \(g(n)\) for \(n \in \{1, \ldots, M\}\). By a similar argument (exercise) we find that there exists some \(K \in \mathbb{N}\) such that \(|\left\lfloor \left\lfloor \frac{b_{d-1} (-K)^{d-1} + \cdots + b_1 (-K) + b_0}{(-K)^d} \right\rfloor \right\rfloor < K^d\) and therefore \(g(-K) < 0\) (why?). In the proof make sure you use the fact that \(d\) is odd. In particular, if \(n\) is odd then \(\left\lfloor (n^d) \right\rfloor = -(n^d)\).

We appeal to the intermediate value theorem, to find a \(c \in [-K, M]\) such that \(g(c) = 0\). As \(g(x) = \left\lfloor \left\lfloor \frac{f(x)}{a_d} \right\rfloor \right\rfloor\), we see that \(f(c) = 0\), and the proof is done.

An interesting fact is that there do exist discontinuous functions that have the intermediate value property. The function \(f(x) := \begin{cases} \sin\left(\frac{1}{x}\right) & \text{ if } x \neq 0, \\ 0 & \text{ if } x = 0, \end{cases}\) is not continuous at 0, however, it has the intermediate value property. That is, for any \((a < b)\), and any \((y)\) such that \(f(a) < y < f(b)) or \((f(a) > y > f(b))\), there exists a \(c\) such that \(f(c) = y\). Proof is left as an exercise.

**Exercises**

Find an example of a discontinuous function \(f : [0,1] \to \mathbb{R}\) where the intermediate value theorem fails.

Find an example of a **bounded** discontinuous function \(f : [0,1] \to \mathbb{R}\) that has neither an absolute minimum nor an absolute maximum.

Let \(f : (0,1) \to \mathbb{R}\) be a continuous function such that \(\lim_{x \to 0^+} f(x) = \lim_{x \to 1^-} f(x) = 0\). Show that \(f\) achieves either an absolute minimum or an absolute maximum on \((0,1)\) (but perhaps not both).

Let \(f : (0,1) \to \mathbb{R}\) be a continuous function such that \(\lim_{x \to 0^+} f(x) = \lim_{x \to 1^-} f(x) = 0\). Show that \(f\) achieves either an absolute minimum or an absolute maximum on \((0,1)\) (but perhaps not both).

Let \(f(x) := \begin{cases} \sin\left(\frac{1}{x}\right) & \text{ if } x \neq 0, \\ 0 & \text{ if } x = 0, \end{cases}\) Show that \(f\) has the intermediate value property. That is, for any \((a < b)\), if there exists a \(y\) such that \(f(a) < y < f(b)\) or \((f(a) > y > f(b))\), then there exists a \(c\) such that \(f(c) = y\).

Suppose \(g(x)\) is a polynomial of odd degree \(d\) such that \(g(x) = x^d + b_{d-1} x^{d-1} + \cdots + b_1 x + b_0\), for some real numbers \(b_{d-1}, b_{d-2}, \ldots, b_1, b_0\). Show that there exists a \(K \in \mathbb{N}\) such that \(g(-K) < 0\). Hint: Make sure to use the fact that \(d\) is odd. You will have to use that \(\left\lfloor (n^d) \right\rfloor = -(n^d)\).

Suppose \(g(x)\) is a polynomial of positive even degree \(d\) such that \(g(x) = x^d + b_{d-1} x^{d-1} + \cdots + b_1 x + b_0\), for some real numbers \(b_{d-1}, b_{d-2}, \ldots, b_1, b_0\). Suppose \(g(0) < 0\). Show that \(g\) has at least two distinct real roots.
Suppose \( f : [a,b] \to \mathbb{R} \) is a continuous function. Prove that the direct image \( f([a,b]) \) is a closed and bounded interval or a single number.

Suppose \( f : \mathbb{R} \to \mathbb{R} \) is continuous and periodic with period \( P > 0 \). That is, \( f(x+P) = f(x) \) for all \( x \in \mathbb{R} \). Show that \( f \) achieves an absolute minimum and an absolute maximum.

Suppose \( f : [0,1] \to [0,1] \) is continuous. Show that \( f \) has a fixed point, in other words, show that there exists an \( x \) such that \( f(x) = x \).

Find an example of a bounded function \( f : \mathbb{R} \to \mathbb{R} \) that does not achieve an absolute minimum nor an absolute maximum on \( \mathbb{R} \).

Suppose \( f : \mathbb{R} \to \mathbb{R} \) is a continuous function such that \( x \leq f(x) \leq x+1 \) for all \( x \in \mathbb{R} \). Find \( f(\mathbb{R}) \).

**Uniform continuity**

*Note: 1.5–2 lectures (Continuous extension and Lipschitz can be optional)*

**Uniform continuity**

We made a fuss of saying that the \( \delta \) in the definition of continuity depended on the point \( c \). There are situations when it is advantageous to have a \( \delta \) independent of any point. Let us give a name to this concept.

Let \( S \subset \mathbb{R} \), and let \( f : S \to \mathbb{R} \) be a function. Suppose for any \( \epsilon > 0 \) there exists a \( \delta > 0 \) such that whenever \( x, c \in S \) and \( |x-c| < \delta \), then \( |f(x)-f(c)| < \epsilon \). Then we say \( f \) is uniformly continuous.

It is not hard to see that a uniformly continuous function must be continuous. The only difference in the definitions is that for a given \( \epsilon > 0 \) we pick a \( \delta > 0 \) that works for all \( c \in S \). That is, \( \delta \) can no longer depend on \( c \), it must be the same for all points in \( S \).
only depends on \(\epsilon\). The domain of definition of the function makes a difference now. A function that is not uniformly continuous on a larger set, may be uniformly continuous when restricted to a smaller set.

The function \(f \colon (0,1) \to \mathbb{R}\), defined by \(f(x) := \frac{1}{x}\) is not uniformly continuous, but it is continuous.

Proof: Given \(\epsilon > 0\), then for \(\epsilon > \left| \frac{1}{x} - \frac{1}{y} \right|\) to hold we must have \(\epsilon > \frac{|y-x|}{|xy|} = \frac{|y-x|}{xy}\). Therefore, to satisfy the definition of uniform continuity we would have to have \(|\delta| \leq xy\epsilon\) for all \((x,y)\) in \(((0,1))\), but that would mean that \(|\delta| \leq 0\). Therefore there is no single \(|\delta| > 0\).

\(f \colon [0,1] \to \mathbb{R}\), defined by \(f(x) := x^2\) is uniformly continuous.

Proof: Note that \(0 \leq x,c \leq 1\). Then \(\left| x^2 - c^2 \right| = \left( x+c \right) \left| x-c \right| \leq (1+1) \left| x-c \right|\). Therefore given \(\epsilon > 0\), let \(\delta := \frac{\epsilon}{2}\). If \(\left| x-c \right| < \delta\), then \(\left| x^2 - c^2 \right| < \epsilon\).

On the other hand, \(f \colon \mathbb{R} \to \mathbb{R}\), defined by \(f(x) := x^2\) is not uniformly continuous.

Proof: Suppose it is uniformly continuous, then for all \(\epsilon > 0\), there would exist a \(\delta > 0\) such that if \(\left| x-c \right| < \delta\), then \(\left| x^2 - c^2 \right| < \epsilon\). Take \(x > 0\) and let \(c := x+\nicefrac{\epsilon}{2}\). Write \(\left| x^2 - c^2 \right| = \frac{(2x+c)\epsilon}{2}\). Therefore \(x < \nicefrac{\epsilon}{\delta}\) for all \(x > 0\), which is a contradiction.

We have seen that if \(f\) is defined on an interval that is either not closed or not bounded, then \(f\) can be continuous, but not uniformly continuous. For a closed and bounded interval \([a,b]\), we can, however, make the following statement.

[unifcont:thm] Let \(f \colon [a,b] \to \mathbb{R}\) be a continuous function. Then \(f\) is uniformly continuous.

We prove the statement by contrapositive. Suppose \(f\) is not uniformly continuous. We will prove that there is some \(c \in [a,b]\) where \(f\) is not continuous. Let us negate the definition of uniformly continuous. There exists an \(\epsilon > 0\) such that for every \(\delta > 0\), there exist points \((x,y)\) in \((S)\) with \(\left| x-y \right| < \delta\) and \(\left| f(x) - f(y) \right| \geq \epsilon\). By , there exists a convergent subsequence \(\{x_n\}\) such that \(\left| x_n - n \right| \to \epsilon\). As \(\left| x_n - n \right| \to \epsilon\), then \(\left| x_n - n \right| \to \epsilon\)

So for the \(\epsilon > 0\) above, we find sequences \(\{x_n\}\) and \(\{y_n\}\) such that \(\left| x_n - y_n \right| \to \epsilon\) and such that \(\left| f(x_n) - f(y_n) \right| \to \epsilon\). By , there exists a convergent subsequence \(\{x_{n_k}\}\) such that \(\left| x_{n_k} - n \right| \to \epsilon\). As \(\left| x_{n_k} - n \right| \to \epsilon\), then \(\left| x_{n_k} - n \right| \to \epsilon\)

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
converges and the limit is \(c\). We now show that \(f\) is not continuous at \(c\). We estimate
\[
\left| f(x_{n_k}) - f(c) \right| = \left| f(x_{n_k}) - f(y_{n_k}) + f(y_{n_k}) - f(c) \right| \\
\geq \left| f(x_{n_k}) - f(y_{n_k}) \right| - \left| f(y_{n_k}) - f(c) \right|
\]
Or in other words
\[
\left| f(x_{n_k}) - f(c) \right| + \left| f(y_{n_k}) - f(c) \right| \geq \epsilon.
\]
At least one of the sequences \(\{f(x_{n_k})\}\) or \(\{f(y_{n_k})\}\) cannot converge to \(f(c)\), otherwise the left hand side of the inequality would go to zero while the right-hand side is positive. Thus \(f\) cannot be continuous at \(c\).

**Continuous extension**

Before we get to continuous extension, we show the following useful lemma. It says that uniformly continuous functions behave nicely with respect to Cauchy sequences. The new issue here is that for a Cauchy sequence we no longer know where the limit ends up; it may not end up in the domain of the function.

Let \(f : S \to \mathbb{R}\) be a uniformly continuous function. Let \(\{x_n\}\) be a Cauchy sequence in \(S\). Then \(\{f(x_n)\}\) is Cauchy.

Let \(\epsilon > 0\) be given. Then there is a \(\delta > 0\) such that \(\left| f(x) - f(y) \right| < \epsilon\) whenever \(\left| x - y \right| < \delta\). Now find an \(M \in \mathbb{N}\) such that for all \(n, k \geq M\) we have \(\left| f_n(x) - f_n(y) \right| < \epsilon\).

An application of the above lemma is the following theorem. It says that a function on an open interval is uniformly continuous if and only if it can be extended to a continuous function on the closed interval.

Let \(\epsilon > 0\) be given. Then there is a \(\delta > 0\) such that \(\left| f(x) - f(y) \right| < \epsilon\) whenever \(\left| x - y \right| < \delta\). Now find an \(M \in \mathbb{N}\) such that for all \(n, k \geq M\) we have \(\left| f_n(x) - f_n(y) \right| < \epsilon\).

One direction is not hard to prove. If \(\widetilde{f}\) is continuous, then it is uniformly continuous by \(\widetilde{f}\). As \(f\) is the restriction of \(\widetilde{f}\) to \((a,b)\), then \(f\) is also uniformly continuous (easy exercise).
\[
\nicefrac{\delta}{2} = \delta.\]

So
\[
\left\lvert {L_1-L_2} \right\rvert = \left\lvert {L_1-f(x_n)+f(x_n)- f(y_n)+f(y_n)-L_2} \right\rvert 
\leq \left\lvert {L_1-f(x_n)} \right\rvert + \left\lvert {f(x_n)-f(y_n)} \right\rvert + \left\lvert {f(y_n)-L_2} \right\rvert 
\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.
\]

Therefore \(L_1 = L_2\). Thus \(L_a\) exists. To show that \(L_b\) exists is left as an exercise.

Now that we know that the limits \((L_a)\) and \((L_b)\) exist, we are done. If \(\lim_{x \to a} f(x)\) exists, then \(\lim_{x \to a} \widetilde{f}(x)\) exists (See ). Similarly with \((L_b)\). Hence \((\widetilde{f})\) is continuous at \((a)\) and \((b)\). And since \(f\) is continuous at \((c \in (a,b))\), then \(\widetilde{f}\) is continuous at \((c \in (a,b))\).

**Lipschitz continuous functions**

Let \(f : S \to \mathbb{R}\) be a function such that there exists a number \(K\) such that for all \(x, y \in S\) we have
\[
\left\lvert {f(x)-f(y)} \right\rvert \leq K \left\lvert {x-y} \right\rvert.
\]

Then \(f\) is said to be **Lipschitz continuous**.

A large class of functions is Lipschitz continuous. Be careful, just as for uniformly continuous functions, the domain of definition of the function is important. See the examples below and the exercises. First we justify the use of the word continuous.

A Lipschitz continuous function is uniformly continuous.

Let \(f : S \to \mathbb{R}\) be a function and let \(K\) be a constant such that for all \(x, y \in S\) we have
\[
\left\lvert {f(x)-f(y)} \right\rvert \leq K \left\lvert {x-y} \right\rvert.
\]

Let \(\epsilon > 0\) be given. Take \(\delta := \frac{\epsilon}{K}\). For any \(x, y \in S\) such that \(\left\lvert {x-y} \right\rvert < \delta\) we have that
\[
\left\lvert {f(x)-f(y)} \right\rvert \leq K \left\lvert {x-y} \right\rvert < K \delta = K \frac{\epsilon}{K} = \epsilon.
\]

Therefore \(f\) is uniformly continuous.

We interpret Lipschitz continuity geometrically. If \(f\) is a Lipschitz continuous function with some constant \(K\). We rewrite the inequality to say that for \(x \neq y\) we have
\[
\left\lvert {\frac{f(x)-f(y)}{x-y}} \right\rvert \leq K.
\]

The quantity \(\frac{f(x)-f(y)}{x-y}\) is the slope of the line between the points \((x,f(x))\) and \((y,f(y))\), that is, a secant line. Therefore, \(f\) is Lipschitz continuous if and only if every line that intersects the graph of \(f\) in at least two distinct points has slope less than or equal to \(K\). See .

The functions \(\sin(x)\) and \(\cos(x)\) are Lipschitz continuous. We have seen () the following two inequalities.
\[
\left\lvert {\sin(x)} - \sin(y) \right\rvert \leq \left\lvert {x-y} \right\rvert \qquad \text{and} \qquad \left\lvert {\cos(x)} - \cos(y) \right\rvert \leq \left\lvert {x-y} \right\rvert.
\]

Hence \(\sin\) and \(\cos\) are Lipschitz continuous with \((K=1)\).

The function \(f : [1,\infty) \to \mathbb{R}\) defined by \(f(x) := \sqrt{x}\) is Lipschitz continuous. Proof:
\[
\left\lvert \sqrt{x} - \sqrt{y} \right\rvert = \left\lvert \frac{x-y}{\sqrt{x} + \sqrt{y}} \right\rvert \leq \frac{1}{\sqrt{1} + \sqrt{1}} \left\lvert {x-y} \right\rvert.
\]

As \((x \geq 1)\) and \((y \geq 1)\), we see that \(\frac{1}{2} \left\lvert {x-y} \right\rvert \leq \frac{1}{2} \left\lvert {\sqrt{x} - \sqrt{y}} \right\rvert \leq \frac{1}{\sqrt{1} + \sqrt{1}} \left\lvert {x-y} \right\rvert\).
Therefore \[
\left| \sqrt{x} - \sqrt{y} \right| = \left| \frac{x-y}{\sqrt{x}+\sqrt{y}} \right| \leq \frac{1}{2} \left| x-y \right|.
\]

On the other hand \((f \colon [0,\infty) \to \mathbb{R})\) defined by \(f(x) := \sqrt{x}\) is not Lipschitz continuous. Let us see why: Suppose we have \((\left| \sqrt{x} - \sqrt{y} \right| \leq K \left| x-y \right|)\) for some \(K\). Let \(y=0\) to obtain \((\left| \sqrt{x} \right| \leq K x)\). If \(K > 0\), then for \((x > 0)\) we then get \((\frac{1}{K} \leq \sqrt{x})\). This cannot possibly be true for all \((x > 0)\). Thus no such \((K > 0)\) exists and \((f)\) is not Lipschitz continuous.

The last example is a function that is uniformly continuous but not Lipschitz continuous. To see that \((\sqrt{x})\) is uniformly continuous on \([0,\infty)\) note that it is uniformly continuous on \([0,1]\) by . It is also Lipschitz (and therefore uniformly continuous) on \([1,\infty)\). It is not hard (exercise) to show that this means that \((\sqrt{x})\) is uniformly continuous on \([0,\infty)\).

**Exercises**

Let \((f \colon S \to \mathbb{R})\) be uniformly continuous. Let \((A \subseteq S)\). Then the restriction \((f|_A)\) is uniformly continuous.

Let \((f \colon (a,b) \to \mathbb{R})\) be a uniformly continuous function. Finish the proof of by showing that the limit \((\lim_{x \to b} f(x))\) exists.

Show that \((f \colon (c,\infty) \to \mathbb{R})\) for some \((c > 0)\) and defined by \((f(x) := \frac{1}{x})\) is Lipschitz continuous.

Show that \((f \colon (0,\infty) \to \mathbb{R})\) defined by \((f(x) := \frac{1}{x})\) is not Lipschitz continuous.

Let \((A, B)\) be intervals. Let \((f \colon A \to \mathbb{R})\) and \((g \colon B \to \mathbb{R})\) be uniformly continuous functions such that \((f(x) = g(x))\) for \((x \in A \cap B)\). Define the function \((h \colon A \cup B \to \mathbb{R})\) by \((h(x) := f(x))\) if \((x \in A)\) and \((h(x) := g(x))\) if \((x \in B \setminus \text{setminus } A)\). a) Prove that if \((A \cap B \not= \emptyset)\), then \((h)\) is uniformly continuous. b) Find an example where \((A \cap B = \emptyset)\) and \((h)\) is not even continuous.

Let \((f \colon \mathbb{R} \to \mathbb{R})\) be a polynomial of degree \((d \geq 2)\). Show that \((f)\) is not Lipschitz continuous.

Let \((f \colon (0,1) \to \mathbb{R})\) be a bounded continuous function. Show that the function \((g(x) := x(1-x)f(x))\) is uniformly continuous.

Show that \((f \colon (0,\infty) \to \mathbb{R})\) defined by \((f(x) := \sin(\frac{1}{x}))\) is not uniformly continuous.

Let \((f \colon \mathbb{Q} \to \mathbb{R})\) be a uniformly continuous function. Show that there exists a uniformly continuous function \((\tilde{f} \colon \mathbb{Q} \to \mathbb{R})\) such that \((f(x) = \tilde{f}(x))\) for all \((x \in \mathbb{Q})\).
a) Find a continuous \( f \colon (0,1) \to \mathbb{R} \) and a sequence \( \{ x_n \} \) in \((0,1)\) that is Cauchy, but such that \( \{ f(x_n) \} \) is not Cauchy. b) Prove that if \( f \colon \mathbb{R} \to \mathbb{R} \) is continuous, and \( \{ x_n \} \) is Cauchy, then \( \{ f(x_n) \} \) is Cauchy.

a) If \( f \colon S \to \mathbb{R} \) and \( g \colon S \to \mathbb{R} \) are uniformly continuous, then show that \( h \colon S \to \mathbb{R} \) given by \( h(x) := f(x) + g(x) \) is uniformly continuous.

b) If \( f \colon S \to \mathbb{R} \) is uniformly continuous and \( a \in \mathbb{R} \), then show that \( h \colon S \to \mathbb{R} \) given by \( h(x) := a f(x) \) is uniformly continuous.

da) If \( f \colon [0,1] \to \mathbb{R} \) is given by \( f(x) := x^m \) for an integer \( m \geq 0 \), show \( f \) is Lipschitz and find the best (the smallest) Lipschitz constant \( K \) (depending on \( m \) of course). Hint: \((x-y)(x^{m-1} + x^{m-2}y + x^{m-3}y^2 + \cdots + x y^{m-2} + y^{m-1}) = x^m - y^m\).

b) Using the previous exercise, show that if \( f \colon [0,1] \to \mathbb{R} \) is a polynomial, that is, \( f(x) := a_m x^m + a_{m-1} x^{m-1} + \cdots + a_0 \), then \( f \) is Lipschitz.

Suppose for \( f \colon [0,1] \to \mathbb{R} \) we have \( \left| f(x)-f(y) \right| \leq K \left| x-y \right| \) for some \( K \). Further show by example that \( K \) is the best possible, that is, there exists such a continuous function for which \( \left| f(x) \right| = K/2 \) for some \( x \in [0,1] \).

**Limits at infinity**

*Note: less than 1 lecture (optional, can safely be omitted unless or is also covered)*

**Limits at infinity**

As for sequences, a continuous variable can also approach infinity. Let us make this notion precise.

We say \((\infty)\) is a cluster point of \( S \subset \mathbb{R} \), if for every \( R \in (M \geq 0) \), there exists an \( x \in S \) such that \( x \geq R \). Similarly \((-\infty)\) is a cluster point of \( S \subset \mathbb{R} \), if for every \( M \in \mathbb{R} \), there exists an \( x \in S \) such that \( x \leq M \).

Let \( f \colon S \to \mathbb{R} \) be a function, where \((\infty)\) is a cluster point of \( S \). If there exists a \( L \in \mathbb{R} \) such that \( \lim_{x \to \infty} f(x) = L \), we call \( L \) the limit and write \( \lim_{x \to \infty} f(x) = L \). Alternatively we write \( f(x) \to L \) as \( x \to \infty \).
Similarly, if \((-\infty)\) is a cluster point of \((S)\) and there exists an \((L)\) in \((\mathbb{R})\) such that for every \((\epsilon > 0)\), there is an \((M)\) in \((\mathbb{R})\) such that \(\lvert f(x) - L \rvert < \epsilon\) whenever \((x \geq M)\), then we say \((f(x))\) converges to \((L)\) as \((x)\) goes to \((-\infty)\). We call \((L)\) the limit and write \([\lim_{x \to -\infty} f(x) := L]\). Alternatively we write \((f(x) \to L)\) as \((x \to -\infty)\).

We cheated a little bit again and said the limit. We leave it as an exercise for the reader to prove the following proposition.

\[\text{[liminfty:unique]}\] The limit at \((-\infty)\) or \((\infty)\) as defined above is unique if it exists.

Let \((f(x) := \frac{1}{\lvert x \rvert + 1})\). Then \([\lim_{x \to \infty} f(x) = 0, \lim_{x \to -\infty} f(x) = 0]\).

Proof: Let \((\epsilon > 0)\) be given. Find \((M > 0)\) large enough so that \((\frac{1}{M+1} < \epsilon)\). If \((x \geq M)\), then \((\frac{1}{\lvert x \rvert + 1} < \epsilon)\). Since \((\frac{1}{\lvert x \rvert + 1} > 0)\) for all \((x)\) the first limit is proved. The proof for \((-\infty)\) is left to the reader.

Let \((f(x) := \sin(\pi x))\). Then \((\lim_{x \to \infty} f(x))\) does not exist. To prove this fact note that if \((x = 2n+\frac{1}{2})\) for some \((n)\) in \((\mathbb{N})\) then \((f(x)=1)\), while if \((x = 2n+\frac{3}{2})\) then \((f(x)=-1)\), so they cannot both be within a small \((\epsilon)\) of a single real number.

We must be careful not to confuse continuous limits with limits of sequences. For \((f(x) = \sin(\pi x))\) we could say \([\lim_{n \to \infty} f(n) = 0, \lim_{x \to \infty} f(x) \text{ does not exist}]\). Of course the notation is ambiguous. We are simply using the convention that \((n)\) in \((\mathbb{N})\), while \((x)\) in \((\mathbb{R})\). When the notation is not clear, it is good to explicitly mention where the variable lives, or what kind of limit are you using.

There is a connection of continuous limits to limits of sequences, but we must take all sequences going to infinity, just as before in .

\[\text{[seqflimitinf:lemma]}\] Suppose \((f: S \to \mathbb{R})\) is a function, \((-\infty)\) is a cluster point of \((S)\) and \((L)\) in \((\mathbb{R})\). Then \((\lim_{x \to -\infty} f(x) = L \iff \lim_{n \to \infty} f(x_n) = L\) for all sequences \((\{ x_n \})\) such that \((\lim \ x_n = \infty)\).

The lemma holds for the limit as \((x \to -\infty)\). Its proof is almost identical and is left as an exercise.

First suppose \((f(x) \to L)\) as \((x \to \infty)\). Given an \((\epsilon > 0)\), there exists an \((M)\) such that for all \((x \geq M)\) we have \((\lvert f(x) - L \rvert < \epsilon)\). Let \((\{ x_n \})\) be a sequence in \((S)\) such that \((\lim \ x_n = \infty)\). Then there exists an \((N)\) such that for all \((n \geq N)\) we have \((\lvert f(x_n) - L \rvert < \epsilon)\). And thus \((\lim_{x \to \infty} f(x) = L)\).

We prove the converse by contrapositive. Suppose \((f(x))\) does not go to \((L)\) as \((x \to \infty)\). This means that there exists an \((\epsilon > 0)\), such that for every \((M)\) in \((\mathbb{N})\), there exists an \((x)\) in \((S)\), \((x \geq M)\), let us call it \((x_M)\),
such that \( \left| f(x_M) - L \right| \geq \epsilon \). Consider the sequence \( \{ x_n \} \). Clearly \( \{ f(x_n) \} \) does not converge to \( L \). It remains to note that \( \lim_{n \to \infty} x_n = \infty \), because \( (x_n \geq n) \) for all \( n \).

Using the lemma, we again translate results about sequential limits into results about continuous limits as \( x \) goes to infinity. That is, we have almost immediate analogues of the corollaries in . We simply allow the cluster point \( c \) to be either \( \infty \) or \( -\infty \), in addition to a real number. We leave it to the student to verify these statements.

### Infinite limit

Just as for sequences, it is often convenient to distinguish certain divergent sequences, and talk about limits being infinite almost as if the limits existed.

Let \( f \colon S \to \mathbb{R} \) be a function and suppose \( S \) has \( \infty \) as a cluster point. We say \( f(x) \) **diverges to infinity** as \( x \) goes to \( \infty \), if for every \( N \in \mathbb{R} \) there exists an \( M \in \mathbb{R} \) such that \( f(x) > N \) whenever \( x \in S \) and \( x \geq M \). We write \( \lim_{x \to \infty} f(x) := \infty \) or we say that \( f(x) \to \infty \) as \( x \to \infty \).

A similar definition can be made for limits as \( x \to -\infty \) or as \( x \to c \) for a finite \( c \). Also similar definitions can be made for limits being \( -\infty \). Stating these definitions is left as an exercise. Note that sometimes *converges to infinity* is used. We can again use sequential limits, and an analogue of is left as an exercise.

Let us show that \( \lim_{x \to \infty} \frac{1+x^2}{1+x} = \infty \).

**Proof:** For \( x \geq 1 \) we have \( \frac{1+x^2}{1+x} \geq \frac{x^2}{x+x} = \frac{x}{2} \). Given \( N \in \mathbb{R} \), take \( M = \max \{ 2N+1, 1 \} \). If \( x \geq M \), then \( x \geq 1 \) and \( \frac{x}{2} > N \). So \( \frac{1+x^2}{1+x} \geq \frac{x}{2} > N \).

### Compositions

Finally, just as for limits at finite numbers we can compose functions easily.

[prop:inflimcompositions] Suppose \( f \colon A \to B \), \( g \colon B \to \mathbb{R} \), \( A, B \subseteq \mathbb{R} \), \( a \in \mathbb{R} \cup \{ -\infty, \infty \} \) is a cluster point of \( A \), and \( b \in \mathbb{R} \cup \{ -\infty, \infty \} \) is a cluster point of \( B \). Suppose \( \lim_{x \to a} f(x) = b \) and \( \lim_{y \to b} g(y) = c \) for some \( c \in \mathbb{R} \cup \{ -\infty, \infty \} \). If \( b \in B \), then suppose \( g(b) = c \). Then \( \lim_{x \to a} g(f(x)) = c \).

The proof is straightforward, and left as an exercise. We already know the proposition when \( a, b, c \in \mathbb{R} \), see Exercises [exercise:contlimitcomposition] and [exercise:contlimitbadcomposition]. Again the requirement that \( g \) is continuous at \( b \), if \( b \in B \), is necessary.

Let \( h(x) := e^{x^2+2x} \). Then \( \lim_{x \to \infty} h(x) = 0 \).

**Proof:** The claim follows once we know \( \lim_{x \to \infty} -x^2+2x = -\infty \) and \( \lim_{y \to -\infty} e^y = 0 \), which is usually proved when the exponential function is defined.
Exercises

Prove.

Let \( f : [1, \infty) \to \mathbb{R} \) be a function. Define \( g : (0, 1] \to \mathbb{R} \) via \( g(x) := f(\frac{1}{x}) \). Using the definitions of limits directly, show that \( \lim_{x \to 0^+} g(x) \) exists if and only if \( \lim_{x \to \infty} f(x) \) exists, in which case they are equal.

Prove.

Let us justify terminology. Let \( f : \mathbb{R} \to \mathbb{R} \) be a function such that \( \lim_{x \to \infty} f(x) = \infty \) (diverges to infinity). Show that \( f(x) \) diverges (i.e. does not converge) as \( x \to \infty \).

Come up with the definitions for limits of \( f(x) \) going to \(-\infty\) as \( x \to \infty \), \( x \to -\infty \), and as \( x \to c \) for a finite \( c \in \mathbb{R} \). Then state the definitions for limits of \( f(x) \) going to \(-\infty\) as \( x \to -\infty \), and as \( x \to c \) for a finite \( c \in \mathbb{R} \).

Suppose \( P(x) := x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \) is a monic polynomial of degree \( n \geq 1 \) (monic means that the coefficient of \( x^n \) is 1). a) Show that if \( n \) is even then \( \lim_{x \to \infty} P(x) = \lim_{x \to -\infty} P(x) = \infty \). b) Show that if \( n \) is odd then \( \lim_{x \to \infty} P(x) = \infty \) and \( \lim_{x \to -\infty} P(x) = -\infty \) (see previous exercise).

Let \( \{ x_n \} \) be a sequence. Consider \( S := \mathbb{N} \subset \mathbb{R} \), and \( f : S \to \mathbb{R} \) defined by \( f(n) := x_n \). Show that the two notions of limit, \( \lim_{n \to \infty} x_n \) and \( \lim_{x \to \infty} f(x) \), are equivalent. That is, show that if one exists so does the other one, and in this case they are equal.

Extend as follows. Suppose \( S \subset \mathbb{R} \) has a cluster point \( c \in \mathbb{R} \), \( c = \infty \), or \( c = -\infty \). Let \( f : S \to \mathbb{R} \) be a function and let \( L = \infty \) or \( L = -\infty \). Show that \( \lim_{x \to c} f(x) = L \) if and only if \( \lim_{n \to \infty} f(x_n) = L \) for all sequences \( \{ x_n \} \) such that \( \lim x_n = c \).

Monotone functions and continuity

Note: 1 lecture (optional, can safely be omitted unless is also covered, requires )

Let \( S \subset \mathbb{R} \). We say \( f : S \to \mathbb{R} \) is increasing (resp. strictly increasing) if \( (x, y \in S) \) with \( x < y \) implies \( f(x) \leq f(y) \) (resp. \( f(x) < f(y) \)). We define decreasing and strictly decreasing in the same way by switching the inequalities for \( f \).

If a function is either increasing or decreasing we say it is monotone. If it is strictly increasing or strictly decreasing we say it is strictly monotone.
Sometimes nondecreasing (resp. nonincreasing) is used for increasing (resp. decreasing) function to emphasize it is not strictly increasing (resp. strictly decreasing).

**Continuity of monotone functions**

It is easy to compute one-sided limits for monotone functions.

[prop:monotlimits] Let \( S \subset \mathbb{R} \), \( c \in \mathbb{R} \), and \( f \colon S \to \mathbb{R} \) be increasing.

If \((c)\) is a cluster point of \((S \cap (-\infty,c))\), then \(\lim_{x \to c^-} f(x) = \sup \{ f(x) : x < c, x \in S \} \) and if \((c)\) is a cluster point of \((S \cap (c,\infty))\), then \(\lim_{x \to c^+} f(x) = \inf \{ f(x) : x > c, x \in S \} \).

Similarly, if \((c)\) is decreasing and \((c)\) is a cluster point of \((S \cap (-\infty,c))\), then \(\lim_{x \to c^-} f(x) = \inf \{ f(x) : x < c, x \in S \} \) and if \((c)\) is a cluster point of \((S \cap (c,\infty))\), then \(\lim_{x \to c^+} f(x) = \sup \{ f(x) : x > c, x \in S \} \).

In particular all the one-sided limits exist whenever they make sense. If we from now on say that say a the left hand limit \(\lim_{x \to c^-} f(x)\) exists we mean that \(c\) is a cluster point of \((S \cap (-\infty,c))\).

Let us assume \((f)\) is increasing, and we will show the first equality. The rest of the proof is very similar and is left as an exercise.

Let \(a := \sup \{ f(x) : x < c, x \in S \} \). If \((a = \infty)\), then for every \(M \in \mathbb{R}\), there exists an \(x_M \) such that \(f(x_M) > M\). As \((f)\) is increasing, \(f(x) \geq f(x_M) > M\) for all \(x \in S\) with \(x > x_M\). If we take \((\delta = c - x_M)\) we obtain the definition of the limit going to infinity.

So assume \((a < \infty)\). Let \(\epsilon > 0\) be given. Because \((a)\) is the supremum, there exists an \((x_{\epsilon} < c)\), \((x_{\epsilon} \in S)\), such that \((f(x_{\epsilon}) > a - \epsilon)\). As \((f)\) is increasing, \((f(x) \geq f(x_{\epsilon}) > a - \epsilon)\) for all \((x \in S)\) with \((x > x_{\epsilon})\). If we take \((\delta = c - x_{\epsilon})\) we obtain the definition of the limit going to infinity.

Suppose \((f)\) is not constant and \((c)\) is a cluster point of \((S \cap (c,\infty))\) whenever \((x < c < y)\), taking the limits we obtain \(\lim_{x \to c^-} f(x) \leq f(c) \leq \lim_{x \to c^+} f(x)\). Then \((f)\) is continuous at \((c)\) if and only if both limits are equal to each other (and hence equal to \((f(c))\)). See also . See to get an idea of what a discontinuity looks like.

[cor:continterval] If \((I \subset \mathbb{R})\) is an interval and \((f)\) is monotone and not constant, then \((f(I))\) is an interval if and only if \((f)\) is continuous.

Assuming \((f)\) is not constant is to avoid the technicality that \((f(I))\) is a single point in that case; \((f(I))\) is a single point if and only if \((f)\) is constant. A constant function is continuous.

If \((f)\) is continuous then \((f(I))\) being an interval is a consequence of . See also .
\begin{align*}
\quad \text{As } \gamma(c) \text{ is a discontinuity, } \gamma(a < b). \text{ If } \gamma(x < c), \text{ then } \\
\gamma(f(x) \leq a), \text{ and if } \gamma(x > c), \text{ then } \gamma(f(x) \geq b). \text{ Therefore any point in } \gamma,(a,b)\setminus\{f(c)\} \text{ is not in } \gamma(f(I)). \text{ However there exists } \\
\gamma(x_1 \in \gamma(S)), \gamma(x_1 < c) \text{ so } f(x_1) \leq a, \text{ and there exists } \gamma(x_2 \in \gamma(S)), \gamma(x_2 > c) \text{ so } f(x_2) \geq b. \text{ Both } \\
\gamma(f(x_1)) \text{ and } \gamma(f(x_2)) \text{ are in } \gamma(f(I)), \text{ and so } \gamma(f(I)) \text{ is not an interval. See } .
\end{align*}

When \(c \in \gamma(I)\) is an endpoint, the proof is similar and is left as an exercise.

A striking property of monotone functions is that they cannot have too many discontinuities.

[cor:monotcountcont] Let \(\gamma(I) \subset \mathbb{R}\) be an interval and \(\gamma(f) \colon \gamma(I) \to \mathbb{R}\) be monotone. Then 
\(\gamma(f)\) has at most countably many discontinuities.

Let \(\gamma(E \subset \gamma(I))\) be the set of all discontinuities that are not endpoints of \(\gamma(I)\). As there are only two endpoints, it is enough to show that \(\gamma(E)\) is countable. Without loss of generality, suppose \(\gamma(f)\) is increasing. We will define an injection \(\gamma(h) \colon \gamma(E) \to \mathbb{Q}\). For each \(c \in \gamma(E)\) the one-sided limits of \(\gamma(f)\) both exist as \(\gamma(c)\) is not an endpoint. Let \(\gamma(a := \lim\{x \to c^-} f(x) = \sup\{f(x) : x \in \gamma(I), x < c\}\), \gamma(b := \lim\{x \to c^+} f(x) = \inf\{f(x) : x \in \gamma(I), x > c\}\). As \(\gamma(c)\) is a discontinuity, we have \(\gamma(a < b)\). There exists a rational number \(\gamma(q \in (a,b))\), so let \(\gamma(h(c) := q)\). Because \(\gamma(f)\) is increasing, \(\gamma(q)\) cannot correspond to any other discontinuity, so after making this choice for all \(\gamma(c \in \gamma(E))\), we have that \(\gamma(h)\) is one-to-one (injective). Therefore, \(\gamma(E)\) is countable.

[example:countdiscont] By \(\lfloor x \rfloor\) denote the largest integer less than or equal to \(\gamma(x)\). Define \(\gamma(f) \colon [0,1] \to \mathbb{R}\) by \(\gamma(f(x) := x + \sum\{n=0}^\lfloor 1/(1-x) \rfloor 2^{-n} ,\) for \(x < 1\) and \(\gamma(f(1) = 3)\). It is left as an exercise to show that \(\gamma(f)\) is strictly increasing, bounded, and has a discontinuity at all points \(\gamma(1-\nicefrac{1}{k})\) for \(\gamma(k \in \mathbb{N})\)). In particular, there are countably many discontinuities, but the function is bounded and defined on a closed bounded interval.

### Continuity of inverse functions

A strictly monotone function \(\gamma(f)\) is one-to-one (injective). To see this notice that if \(\gamma(y \not= x)\) then we can assume \(\gamma(y < x)\). Then either \(\gamma(f(y) < f(x))\) if \(\gamma(f)\) is strictly increasing or \(\gamma(f(x) > f(y))\) if \(\gamma(f)\) is strictly decreasing, so \(\gamma(f(x) \not= f(y))\). Hence, it must have an inverse \(\gamma(f^{-1})\) defined on its range.

[prop:invcont] If \(\gamma(I) \subset \mathbb{R}\) is an interval and \(\gamma(f) \colon \gamma(I) \to \mathbb{R}\) is strictly monotone. Then the inverse \(\gamma(f^{-1})\) is continuous.

Let us suppose \(\gamma(f)\) is strictly increasing. The proof is almost identical for a strictly decreasing function. Since \(\gamma(f)\) is strictly increasing, so is \(\gamma(f^{-1})\). That is, if \(\gamma(f(x) < f(y))\), then we must have \(\gamma(x < y)\) and therefore \(\gamma(f(x) \not= f(y))\). Hence, it must have an inverse \(\gamma(f^{-1})\) defined on its range.

Take \(\gamma(c \in \gamma(f(I)))\). If \(\gamma(c)\) is not a cluster point of \(\gamma(f(I))\), then \(\gamma(f^{-1})\) is continuous at \(\gamma(c)\) automatically. So let \(\gamma(c)\) be a cluster point of \(\gamma(f(I))\). Suppose both of the following one-sided limits exist: \
\begin{align*}
\gamma(f_0 := \lim\{y \to c^-} f^{-1}(y) = \sup\{f^{-1}(y) : y < c, y \in \gamma(f(I))\} = \sup\{x : f(x) < c, x \in \gamma(I)\} ,\ \\
\gamma(f_1 := \lim\{y \to c^+} f^{-1}(y) = \inf\{f^{-1}(y) : y > c, y \in \gamma(f(I))\} = \inf\{x : f(x) > c, x \in \gamma(I)\}. \end{align*}
\end{align*}
We have \(\gamma(x_0 \not= x_1)\) as \(\gamma(f^{-1})\) is
increasing. For all \(x > x_0\) with \(x \in \mathbb{I}\), we have \(f(x) \geq c\). As \(f\) is strictly increasing, we must have \(f(x) > c\) for all \(x > x_0\), \(x \in \mathbb{I}\). Therefore, \[\{x : x > x_0, x \in \mathbb{I}\} \subset \{x : f(x) > c, x \in \mathbb{I}\}\.] The infimum of the left hand set is \(x_0\) and the infimum of the right hand set is \(x_1\), so we obtain \(x_0 \leq x_1\). So \(x_1 = x_0\), and \(f^{-1}\) is continuous at \(c\).

If one of the one-sided limits does not exist the argument is similar and is left as an exercise.

The proposition does not require \(f\) itself to be continuous. For example, let \(f : \mathbb{R} \to \mathbb{R}\) \[f(x) := \begin{cases} x & \text{if } x < 0, \\ x+1 & \text{if } x \geq 0. \end{cases}\] The function \(f\) is not continuous at \(0\). The image of \((0) = \mathbb{R}\) is the set \((-\infty,0] \cup [1,\infty)\), not an interval. Then \(f^{-1} : (-\infty,0] \cup [1,\infty) \to \mathbb{R}\) can be written as \[f^{-1}(x) = \begin{cases} x & \text{if } x < 0, \\ x-1 & \text{if } x \geq 1. \end{cases}\] It is not difficult to see that \(f^{-1}\) is a continuous function.

Notice what happens with the proposition if \(f(I)\) is an interval. In that case we could simply apply to both \(f\) and \(f^{-1}\). That is, if \(f : \mathbb{I} \to \mathbb{J}\) is an onto strictly monotone function and \(I\) and \(J\) are intervals, then both \(f\) and \(f^{-1}\) are continuous. Furthermore \(f(I)\) is an interval precisely when \(f\) is continuous.

**Exercises**

Suppose \(f : [0,1] \to \mathbb{R}\) is monotone. Prove \(f\) is bounded.

Finish the proof of .

Finish the proof of .

Prove the claims in .

Finish the proof of .

Suppose \(S \subset \mathbb{R}\) and \(f : \mathbb{R} \to \mathbb{R}\) is an increasing function. a) If \(c\) is a cluster point of \(S\), show that \(\lim\limits_{x \to c^-} f(x) = \infty\). b) If \(c\) is a cluster point of \(S\), show that \(\lim\limits_{x \to c^-} f(x) = \infty\), prove that \(\lim\limits_{x \to c^-} f(x) = \infty\).

Suppose \(f : \mathbb{R} \to \mathbb{R}\) is an interval and \(f : \mathbb{R} \to \mathbb{R}\) is a function such that for each \(c \in \mathbb{I}\), there exist \(a, b \in \mathbb{R}\) such that \(f(x) \geq ax + b\) for all \(x \in \mathbb{I}\). Show that \(f\) is strictly increasing.

Suppose \(f : \mathbb{I} \to \mathbb{J}\) is a continuous, bijective (one-to-one and onto) function for two intervals \(I\) and \(J\). Show that \(f\) is strictly monotone.

Consider a monotone function \(f : \mathbb{I} \to \mathbb{R}\) on an interval \(I\). Prove that there exists a function \(g : \mathbb{I} \to \mathbb{R}\) such that \(\lim\limits_{x \to c^-} g(x) = g(c)\) for all \(c \in \mathbb{I}\), except the smaller (left) endpoint of \(I\), and such that \(g(x) = f(x)\) for all but countably many \(x\).

---

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
a) Let \( S \subset \mathbb{R} \) be any subset. If \( f: S \to \mathbb{R} \) is increasing, then show that there exists an increasing \( F: \mathbb{R} \to \mathbb{R} \) such that \( f(x) = F(x) \) for all \( x \in S \). 
b) Find an example of a strictly increasing \( f: S \to \mathbb{R} \) such that an increasing \( F \) as above is never strictly increasing.

[exercise:increasingfuncdiscatQ] Find an example of an increasing function \( f: [0,1] \to \mathbb{R} \) that has a discontinuity at each rational number. Then show that the image \( f([0,1]) \) contains no interval. Hint: Enumerate the rational numbers and define the function with a series.

---

### The Derivative

#### The derivative

*Note: 1 lecture*

The idea of a derivative is the following. Let us suppose a graph of a function looks locally like a straight line. We can then talk about the slope of this line. The slope tells us the rate at which the value of the function changing at the particular point. Of course, we are leaving out any function that has corners or discontinuities. Let us be precise.

#### Definition and basic properties

Let \( (I) \) be an interval, let \( f: I \to \mathbb{R} \) be a function, and let \( c \in I \). If the limit

\[
\lim_{x \to c} \frac{f(x)-f(c)}{x-c} = L
\]

exists, then we say \( f \) is *differentiable* at \( c \), that \( L \) is the *derivative* of \( f \) at \( c \), and write \( f'(c) := L \).

If \( f \) is differentiable at all \( c \in I \), then we simply say that \( f \) is *differentiable*, and then we obtain a function \( f': I \to \mathbb{R} \).

The expression \( \frac{f(x)-f(c)}{x-c} \) is called the *difference quotient*.

The graphical interpretation of the derivative is depicted in . The left-hand plot gives the line through \( (c,f(c)) \) and \( (x,f(x)) \) with slope \( \frac{f(x)-f(c)}{x-c} \), that is, the so-called *secant line*. When we take the limit as \( x \) goes to \( c \), we get the right-hand plot, where we see that the derivative of the function at the point \( c \) is the slope of the line tangent to the graph of \( f \) at the point \( (c,f(c)) \).

We allow \( I \) to be a closed interval and we allow \( c \) to be an endpoint of \( I \). Some calculus books do not allow \( c \) to be an endpoint of an interval, but all the theory still works by allowing it, and it makes our work easier.

Let \( f(x) := x^2 \) defined on the whole real line. We find that

\[
\lim_{x \to c} \frac{x^2-c^2}{x-c} = \lim_{x \to c} (x+c) = 2c.
\]

Therefore \( f'(c) = 2c \).
The function \( f(x) := |x| \) is not differentiable at the origin. When \( x > 0 \), then \( \frac{|x| - |0|}{x - 0} = \frac{x - 0}{x - 0} = 1 \), and when \( x < 0 \) we have \( \frac{|x| - |0|}{x - 0} = \frac{-x - 0}{x - 0} = -1 \).

A famous example of Weierstrass shows that there exists a continuous function that is not differentiable at any point. The construction of this function is beyond the scope of this book. On the other hand, a differentiable function is always continuous.

Let \( f \colon I \to \mathbb{R} \) be differentiable at \( c \in I \), then it is continuous at \( c \).

We know the limits \( \lim_{x \to c} \frac{f(x) - f(c)}{x - c} = f'(c) \) and \( \lim_{x \to c} (x - c) = 0 \) exist. Furthermore, \( \lim_{x \to c} (f(x) - f(c)) \) exists and \( \lim_{x \to c} (x - c) \) exists. Therefore the limit of \( \frac{f(x) - f(c)}{x - c} \) exists and \( \lim_{x \to c} (x - c) \) exists. Hence \( \lim_{x \to c} f(x) = f(c) \), and \( f \) is continuous at \( c \).

An important property of the derivative is linearity. The derivative is the approximation of a function by a straight line. The slope of a line through two points changes linearly when the \( y \)-coordinates are changed linearly. By taking the limit, it makes sense that the derivative is linear.

Let \( I \) be an interval, let \( f \colon I \to \mathbb{R} \) and \( g \colon I \to \mathbb{R} \) be differentiable at \( c \in I \), and let \( \alpha \in \mathbb{R} \).

i. Define \( h \colon I \to \mathbb{R} \) by \( h(x) := \alpha f(x) \). Then \( h \) is differentiable at \( c \) and \( h'(c) = \alpha f'(c) \).

ii. Define \( h \colon I \to \mathbb{R} \) by \( h(x) := f(x) + g(x) \). Then \( h \) is differentiable at \( c \) and \( h'(c) = f'(c) + g'(c) \).

First, let \( h(x) := \alpha f(x) \). For \( x \in I \), \( x \neq c \) we have \( \lim_{x \to c} (f(x) - f(c)) \) exists. Therefore the limit of \( \frac{f(x) - f(c)}{x - c} \) exists and \( \lim_{x \to c} (x - c) \) exists. Hence \( \lim_{x \to c} (f(x) - f(c)) \) exists and \( \lim_{x \to c} (x - c) \) exists. Therefore \( h \) is differentiable at \( c \), and the derivative is computed as given.

Next, define \( h(x) := f(x) + g(x) \). For \( x \in I \), \( x \neq c \) we have \( \lim_{x \to c} (f(x) - f(c)) \) exists. Therefore the limit of \( \frac{f(x) - f(c)}{x - c} \) exists and \( \lim_{x \to c} (x - c) \) exists. Hence \( \lim_{x \to c} (f(x) - f(c)) \) exists and \( \lim_{x \to c} (x - c) \) exists. Therefore \( h \) is differentiable at \( c \) and the derivative is computed as given.

It is not true that the derivative of a multiple of two functions is the multiple of the derivatives. Instead we get the so-called product rule or the Leibniz rule.
Let \( I \) be an interval, let \( f : I \to \mathbb{R} \) and \( g : I \to \mathbb{R} \) be functions differentiable at \( c \). If \( h : I \to \mathbb{R} \) is defined by \( h(x) := f(x) \cdot g(x) \), then \( h \) is differentiable at \( c \) and \( h'(c) = f(c) \cdot g'(c) + f'(c) \cdot g(c) \).

The proof of the product rule is left as an exercise. The key is to use the identity \( f(x) \cdot g(x) - f(c) \cdot g(c) = f(x) \cdot (g(x) - g(c)) + g(c) \cdot (f(x) - f(c)) \).

Let \( I \) be an interval, let \( f : I \to \mathbb{R} \) and \( g : I \to \mathbb{R} \) be differentiable at \( c \) and \( g(x) \neq 0 \) for all \( x \in I \). If \( h : I \to \mathbb{R} \) is defined by \( h(x) := \frac{f(x)}{g(x)} \), then \( h \) is differentiable at \( c \) and \( h'(c) = \frac{f'(c) \cdot g(c) - f(c) \cdot g'(c)}{(g(c))^2} \).

Again the proof is left as an exercise.

**Chain rule**

A useful rule for computing derivatives is the chain rule.

Let \( I_1, I_2 \) be intervals, let \( g : I_1 \to I_2 \) be differentiable at \( c \in I_1 \), and \( f : I_2 \to \mathbb{R} \) be differentiable at \( g(c) \). If \( h : I_1 \to \mathbb{R} \) is defined by \( h(x) := (f \circ g)(x) = f(g(x)) \), then \( h \) is differentiable at \( c \) and \( h'(c) = f'(g(c)) \cdot g'(c) \).

Let \( d := g(c) \). Define \( u : I_2 \to \mathbb{R} \) and \( v : I_1 \to \mathbb{R} \) by

\[
\begin{aligned}
&u(y) := \\
&v(x) := \\
\end{aligned}
\]

We note that \( f(y) - f(d) = u(y) \cdot (y-d) \quad \text{and} \quad g(x) - g(c) = v(x) \cdot (x-c) \). We plug in to obtain \( h(x) - h(c) = f' \cdot \text{bigl}(g(x) \cdot \text{bigl}) - f' \cdot \text{bigl}(g(c) \cdot \text{bigl}) = u' \cdot \text{bigl}(g(x) \cdot \text{bigl}) - u' \cdot \text{bigl}(g(c) \cdot \text{bigl}) \).

Therefore, \( h'(c) = \lim_{y \to d} u'(g(y)) \cdot v(x) \). We compute the limits \( \lim_{y \to d} u'(g(y)) = f'(d) \cdot g'(c) \) and \( \lim_{x \to c} v(x) = g'(c) \). That is, the functions \( u' \) and \( v \) are continuous at \( d = g(c) \) and \( c \) respectively. Furthermore the function \( g \) is continuous at \( c \). Hence the limit of the right-hand side of \( [\text{eq:chainruleeq}] \) as \( x \) goes to \( c \) exists and is equal to \( (f' \cdot \text{bigl}(g(c) \cdot \text{bigl}) \cdot g'(c)) \). Thus \( h \) is differentiable at \( c \) and the limit is \( (f' \cdot \text{bigl}(g(c) \cdot \text{bigl}) \cdot g'(c)) \).

**Exercises**

Prove the product rule. Hint: Use \( (f(x) \cdot g(x) - f(c) \cdot g(c) = f(x) \cdot \text{bigl}(g(x) - g(c)) + g(c) \cdot \text{bigl}(f(x) - f(c)) \).)

Prove the quotient rule. Hint: You can do this directly, but it may be easier to find the derivative of \( \frac{1}{x} \) and then use the chain rule and the product rule.

[exercise:diffofxn] For \( n \in \mathbb{Z} \), prove that \( x^n \) is differentiable and find the derivative, unless, of course, \( n < 0 \) and \( (x=0) \). Hint: Use the product rule.

Prove that a polynomial is differentiable and find the derivative. Hint: Use the previous exercise.
Define \( f \colon \mathbb{R} \to \mathbb{R} \) by \( f(x) := \begin{cases} x^2 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{otherwise.} \end{cases} \) Prove that \( f \) is differentiable at \( 0 \), but discontinuous at all points except \( 0 \).

Assume the inequality \( \left| x - \sin(x) \right| \leq x^2 \). Prove that \( \sin \) is differentiable at \( 0 \), and find the derivative at \( 0 \).

Using the previous exercise, prove that \( \sin \) is differentiable at all \( x \) and that the derivative is \( \cos(x) \). Hint: Use the sum-to-product trigonometric identity as we did before.

Let \( f \colon \mathbb{R} \to \mathbb{R} \) be differentiable. Given \( n \in \mathbb{Z} \), define \( f^n \) be the function defined by \( f^n(x) := \bigl( f(x) \bigl( \bigl( f(x) \bigl( \cdots \bigl( f(x) \big| \bigr) \big| \bigr) \big| \cdots \bigl( f(x) \big| \bigr) \bigg) \bigg) \bigg) \). If \( n < 0 \) assume \( f(x) \not= 0 \). Prove that \( f^n(x) = n \bigl( \bigl( f(x) \bigl( \bigl( f(x) \bigl( \cdots \bigl( f(x) \big| \bigr) \big| \bigr) \big| \cdots \bigl( f(x) \big| \bigr) \bigg) \bigg) \bigg) \bigg) \) is a differentiable Lipschitz continuous function. Prove that \( f^n \) is a bounded function.

Let \( (I_1, I_2) \) be intervals. Let \( f \colon I_1 \to \mathbb{R} \) be a bijective function and \( g \colon I_2 \to \mathbb{R} \) be the inverse. Suppose that both \( f \) is differentiable at \( c \in I_1 \) and \( f'(c) \not= 0 \) and \( g \) is differentiable at \( f(c) \). Use the chain rule to find a formula for \( (g^n)(x) \) in terms of \( f(x) \).

[exercise:bndmuldiff] Suppose \( f \colon \mathbb{R} \to \mathbb{R} \) is a bounded function and \( g \colon \mathbb{R} \to \mathbb{R} \) is a function differentiable at \( c \) and \( g(c) = f'(c) = 0 \). Show that \( (h)(x) := f(x) g(x) \) is differentiable at \( c \). Hint: Note that you cannot apply the product rule.

[exercise:diffsqueeze] Suppose \( f \colon \mathbb{R} \to \mathbb{R} \), \( g \colon \mathbb{R} \to \mathbb{R} \), and \( h \colon \mathbb{R} \to \mathbb{R} \), are functions. Suppose \( \epsilon \in \mathbb{R} \) is such that \( f(c) = g(c) = h(c) \) are differentiable at \( c \), and \( g'(c) = h'(c) \). Furthermore suppose \( h(x) \leq f(x) \leq g(x) \) for all \( x \in I \). Prove \( f \) is differentiable at \( c \) and \( f'(c) = g'(c) = h'(c) \).

---

**Mean value theorem**

*Note: 2 lectures (some applications may be skipped)*

**Relative minima and maxima**

We talked about absolute maxima and minima. These are the tallest peaks and lowest valleys in the whole mountain range. We might also want to talk about peaks of individual mountains and valleys.

Let \( S \subseteq \mathbb{R} \) be a set and let \( f \colon S \to \mathbb{R} \) be a function. The function \( f \) is said to have a *relative maximum* at \( c \in S \) if there exists a \( \delta > 0 \) such that for all \( x \in S \) such that \( \| x - c \| < \delta \) we have \( f(x) \leq f(c) \). The definition of *relative minimum* is analogous.

[relminmax:thm] Let \( f \colon [a, b] \to \mathbb{R} \) be a function differentiable at \( c \in (a, b) \), and \( f \) is a relative minimum or a relative maximum of \( f \). Then \( f'(c) = 0 \).
We prove the statement for a maximum. For a minimum the statement follows by considering the function \((-f)\).

Let \(c\) be a relative maximum of \(f\). In particular as long as \(|x-c| < \delta\) we have \(f(x) \leq f(c)\). Then we look at the difference quotient. If \(x > c\) we note that \(\frac{f(x)-f(c)}{x-c} \leq 0\) and if \(y < c\) we have \(\frac{f(y)-f(c)}{y-c} \geq 0\). See for an illustration.

We now take sequences \(\{x_n\}\) and \(\{y_n\}\), such that \(x_n > c\), and \(y_n < c\) for all \(n \in \mathbb{N}\), and such that \(\lim x_n = \lim y_n = c\). Since \(f\) is differentiable at \(c\) we know \(0 \geq \lim \frac{f(x_n) - f(c)}{x_n - c} = f'(c) = \lim \frac{f(y_n) - f(c)}{y_n - c} \geq 0\). See for an illustration.

For a differentiable function, a point where \((f'(c) = 0)\) is called a critical point. When \((f)\) is not differentiable at some points, it is common to also say \((c)\) is a critical point if \((f'(c))\) does not exist. The theorem says that a relative minimum or maximum at an interior point of an interval must be a critical point. As you remember from calculus, finding minima and maxima of a function can be done by finding all the critical points together with the endpoints of the interval and simply checking where is the function biggest or smallest.

**Rolle’s theorem**

Suppose a function is zero at both endpoints of an interval. Intuitively it ought to attain a minimum or a maximum in the interior of the interval, then at such a minimum or a maximum, the derivative should be zero. See for the geometric idea. This is the content of the so-called Rolle’s theorem.

\[\text{Let } f: [a, b] \to \mathbb{R} \text{ be continuous function differentiable on } (a, b) \text{ such that } f(a) = f(b). \]

Then there exists a \(c \in (a, b)\) such that \(f'(c) = 0\).

As \((f)\) is continuous on \(([a, b])\) it attains an absolute minimum and an absolute maximum in \(([a, b])\). We wish to apply and so we need a minimum or maximum at some \(c \in (a, b)\). Write \(K := f(a) = f(b)\). If there exists an \(x\) such that \(f(x) > K\), then the absolute maximum is bigger than \(K\) and hence occurs at \(c \in (a, b)\), and therefore we get \(f'(c) = 0\). On the other hand if there exists an \(x\) such that \(f(x) < K\), then the absolute minimum occurs at some \(c \in (a, b)\) and we have that \(f'(c) = 0\). If there is no \(x\) such that \(f(x) > K\) or \(f(x) < K\), then we have that \(f(x) = K\) for all \(x\) and then \(f'(x) = 0\) for all \(x \in [a, b]\), so any \(c\) will work.

Note that it is absolutely necessary for the derivative to exist for all \(x \in (a, b)\). For example take the function \(f(x) = |x|\) on \([-1, 1]\). Clearly \((f(x) = f(1))\), but there is no point where \(f'(c) = 0\).

**Mean value theorem**

We extend to functions that attain different values at the endpoints.

\[\text{Let } f: [a, b] \to \mathbb{R} \text{ be a continuous function differentiable on } (a, b) \text{. Then there exists a point } c \in (a, b) \text{ such that } f(b) - f(a) = f'(c)(b-a).\]

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
The theorem follows from . Define the function \( g \colon [a,b] \to \{ \mathbb{R} \} \) by
\[
g(x) := f(x)-f(b)+\bigl(f(b)-f(a)\bigr)\frac{b-x}{b-a}.
\]
The function \( g \) is a differentiable on \((a,b)\), continuous on \([a,b]\), such that \( g(a) = 0 \) and \( g(b) = 0 \). Thus there exists \( c \in (a,b) \) such that \( (g'(c) = 0) \). Hence \( 0 = g'(c) = f'(c) + \bigl(f(b)-f(a)\bigr)\frac{-1}{b-a} \). Or in other words \( f'(c)(b-a) = f(b)-f(a) \).

For a geometric interpretation of the mean value theorem, see . The idea is that the value \( \frac{f(b)-f(a)}{b-a} \) is the slope of the line between the points \( (a,f(a)) \) and \( (b,f(b)) \). Then \( c \) is the point such that \( f'(c) = \frac{f(b)-f(a)}{b-a} \), that is, the tangent line at the point \( (c,f(c)) \) has the same slope as the line between \( (a,f(a)) \) and \( (b,f(b)) \).

**Applications**

We now solve our very first differential equation.

[prop:derzeroconst] Let \( I \) be an interval and let \( f \colon I \to \{ \mathbb{R} \} \) be a differentiable function such that \( f'(x) = 0 \) for all \( x \in I \). Then \( f \) is constant.

Take arbitrary \( x,y \in I \) with \( x < y \). Then \( f \) restricted to \( [x,y] \) satisfies the hypotheses of the . Therefore there is a \( c \in (x,y) \) such that \( f(y)-f(x) = f'(c)(y-x) \). As \( f'(c) = 0 \), we have \( f(y) = f(x) \). Therefore, the function is constant.

Now that we know what it means for the function to stay constant, let us look at increasing and decreasing functions. We say \( f \colon I \to \{ \mathbb{R} \} \) is *increasing* (resp. *strictly increasing*) if \( x < y \) implies \( f(x) \leq f(y) \) (resp. \( f(x) < f(y) \)). We define *decreasing* and *strictly decreasing* in the same way by switching the inequalities for \( f \).

[incdecdiffprop] Let \( I \) be an interval and let \( f \colon I \to \{ \mathbb{R} \} \) be a differentiable function.

i. \( f \) is increasing if and only if \( f'(x) \geq 0 \) for all \( x \in I \).

ii. \( f \) is decreasing if and only if \( f'(x) \leq 0 \) for all \( x \in I \).

Let us prove the first item. Suppose \( f \) is increasing, then for all \( x \in I \) and \( c \in I \) we have \( f(x) \geq f(c) \) for all \( x \in I \). Taking a limit as \( x \) goes to \( c \) we see that \( f(c) \geq 0 \).

For the other direction, suppose \( f'(x) \geq 0 \) for all \( x \in I \). Take any \( x, y \in I \) where \( x < y \). By the there is some \( c \in (x,y) \) such that \( f(y)-f(x) = f'(c)(y-x) \). As \( f'(x) \geq 0 \), and \( x-y < 0 \), then \( f(x) \geq f(y) \) or \( f(x) \leq f(y) \) and so \( f \) is increasing.

We leave the decreasing part to the reader as exercise.

We can make a similar but weaker statement about strictly increasing and decreasing functions. If \( f'(x) > 0 \) for all \( x \in I \), then \( f \) is strictly increasing. The proof is left as an exercise. The converse is not true. For example, \( f(x) := x^3 \) is a strictly increasing function, but \( f(0) = 0 \).
Another application of the is the following result about location of extrema. The theorem is stated for an absolute minimum and maximum, but the way it is applied to find relative minima and maxima is to restrict \( f(x) \) to an interval \((c-\delta, c+\delta)\).

\[ \text{firstderminmaxtest} \]

Let \( f: \mathbb{R} \to \mathbb{R} \) be continuous. Let \( f \) is differentiable on \((c,c)\) and \((c,b)\).

i. If \( f'(x) \leq 0 \) for \( x \in (a,c) \) and \( f'(x) \geq 0 \) for \( x \in (c,b) \), then \( f \) has an absolute minimum at \( c \).

ii. If \( f'(x) \geq 0 \) for \( x \in (a,c) \) and \( f'(x) \leq 0 \) for \( x \in (c,b) \), then \( f \) has an absolute maximum at \( c \).

Let us prove the first item. The second is left to the reader. Let \( \{ y_n \} \) be a sequence such that \( (x < y_n < c) \) and \( (\lim, y_n = c) \). By the previous proposition, the function is decreasing on \((a,c)\), so \( f(x) \leq f(y_n) \). The function is continuous at \( c \) so we can take the limit to get \( f(x) \leq f(c) \) for all \( x \in (a,b) \).

Similarly take \( (x, y_n) \) a sequence such that \( (c < y_n < x) \) and \( (\lim, y_n = c) \). The function is increasing on \((c,b)\), so \( f(x) \geq f(y_n) \). By continuity of \( f \) we get \( f(x) \geq f(c) \) for all \( x \in (a,b) \). Thus \( f(x) \geq f(c) \) for all \( x \in (a,b) \).

The converse of the proposition does not hold. See below.

**Continuity of derivatives and the intermediate value theorem**

Derivatives of functions satisfy an intermediate value property. The result is usually called Darboux’s theorem.

\[ \text{thm:darboux} \]

Let \( f: [a, b] \to \mathbb{R} \) be differentiable. Suppose that there exists a \( y \in \mathbb{R} \) such that \( f'(a) < y < f'(b) \). Then there exists a \( c \in (a, b) \) such that \( f'(c) = y \).

Suppose without loss of generality that \( f'(a) < y < f'(b) \). Define \( g(x) := yx - f(x) \). As \( g(y) \) is continuous on \((a,b)\), then \( g(x) \) attains a maximum at some \( c \in (a,b) \).

Now compute \( g'(x) = y - f'(x) \). Thus \( g'(a) > 0 \). As the derivative is the limit of difference quotients and is positive, there must be some difference quotient that is positive. That is, there must exist an \( x > a \) such that \( g'(a) > 0 \) or \( g(x) > g(a) \). Thus \( g'(a) \) cannot possibly be a maximum of \( g \). Similarly as \( g'(b) < 0 \), we find an \( x < b \) (a different \( (x) \)) such that \( g'(b) > 0 \) or \( g(x) > g(b) \), thus \( g(b) \) cannot possibly be a maximum.

Therefore \( c \in (a,b) \). Then as \( g'(c) \) is a maximum of \( g \) we find \( g'(c) = 0 \) and \( f'(c) = y \).

We have seen already that there exist discontinuous functions that have the intermediate value property. While it is hard to imagine at first, there also exist functions that are differentiable everywhere and the derivative is not continuous.

\[ \text{baddifffunc:example} \]

Let \( f: \mathbb{R} \to \mathbb{R} \) be the function defined by \( f(x) := \begin{cases} \sin(x) & \text{if } x \not= 0, \\ 0 & \text{if } x = 0. \end{cases} \) We claim that \( f \) is
differentiable everywhere, but \( (f' \colon \mathbb{R} \to \mathbb{R}) \) is not continuous at the origin. Furthermore, \( (f') \) has a minimum at 0, but the derivative changes sign infinitely often near the origin. See .

A function with a discontinuous derivative. The function \( f \)
is on the left and \( f' \) is on the right. Notice that \( f(x) \leq x^2 \)on the left graph.

A function with a discontinuous derivative. The function \( f \)
is on the left and \( f' \) is on the right. Notice that \( f(x) \leq x^2 \)on the left graph.

Proof: It is easy to see from the definition that \( (f') \) has an absolute minimum at 0: we know \( (f(x) \geq 0) \) for all \( x \) and 
\( (f(0) = 0) \).

The function \( (f') \) is differentiable for \( (x \not= 0) \) and the derivative is 
\[ (2 \sin \left( \frac{1}{x} \right) \cos \left( \frac{1}{x} \right) - \cos \left( \frac{1}{x} \right) \sin \left( \frac{1}{x} \right)) \bigg|_{x} \bigg|_{x=0} \].

As an exercise show that for \( x_n = \frac{4}{(8n+1)\pi} \) we have \( (\lim f'(x_n) = -1) \), and
for \( y_n = \frac{4}{(8n+3)\pi} \) we have \( (\lim f'(y_n) = 1) \). Hence if \( (f') \) exists at \( 0 \), then it cannot be continuous.

Let us show that \( (f') \) exists at 0. We claim that the derivative is zero. In other words \( (\left| \frac{f(x)-f(0)}{x-0} \right| - 0) \)
\( \to 0 \) goes to zero as \( (x) \) goes to zero. For \( x \not= 0 \) we have
\( (\left| \frac{f(x)-f(0)}{x-0} \right| - 0) \to 0 \)
\( \lim f'(x) \to 0 \). And, of course, as \( (x) \) tends to zero, then
\( (\left| \frac{f(x)}{x} \right| \lim f'(x) \to 0) \). A key point in the above calculation is that is that \( (\left| \frac{f(x)}{x} \right| \lim f'(x) \to 0) \), see also Exercises
[exercise:bndmuldiff] and [exercise:diffsqueeze].

It is sometimes useful to assume the derivative of a differentiable function is continuous. If \( (f' \colon I \to \mathbb{R}) \) is
differentiable and the derivative \( (f') \) is continuous on \( (I) \), then we say \( (f') \) is \textit{continuously differentiable}. It is common to
write \( (C^1(I)) \) for the set of continuously differentiable functions on \( (I) \).

**Exercises**

Finish the proof of .

Finish the proof of .

Suppose \( (f' \colon \mathbb{R} \to \mathbb{R}) \) is a differentiable function such that \( (f') \) is a bounded function. Prove \( (f') \) is a Lipschitz continuous function.

Suppose \( (f' \colon [a,b] \to \mathbb{R}) \) is differentiable and \( (c \in [a,b]) \). Then show there exists a sequence \( (x_n) \)
converging to \( (c) \), \( (x_n \not= c) \) for all \( (n) \), such that \( \lim f'(x_n) = f'(c) \) Do note this does not imply that \( (f') \) is continuous (why?).
Suppose \( f : \mathbb{R} \to \mathbb{R} \) is a function such that \( |f(x) - f(y)| \leq |x-y|^2 \) for all \( x \) and \( y \). Show that \( f(x) = C \) for some constant \( C \). Hint: Show that \( f \) is differentiable at all points and compute the derivative.

Suppose \( f : (a,b) \to \mathbb{R} \) is a function such that \( f'(x) > 0 \) for all \( x \in (a,b) \), then show that there exists a constant \( C \) such that \( f(x) = g(x) + C \).

Prove the following version of L'Hopital's rule. Suppose \( f, g : (a,b) \to \mathbb{R} \) are differentiable functions. Suppose that at \( c \in (a,b) \), \( f(c) = 0 \), \( g(c) = 0 \), and that the limit of \( \frac{f'(x)}{g'(x)} \) as \( x \) goes to \( c \) exists. Show that \( \lim_{x \to c} \frac{f'(x)}{g'(x)} = \lim_{x \to c} \frac{f(x)}{g(x)} \).

Let \( f : (a,b) \to \mathbb{R} \) be an unbounded differentiable function. Show \( f'(x) \) is unbounded.

Prove the theorem Rolle actually proved in 1691: If \( f \) is a polynomial, \( f(a) = f(b) = 0 \) for some \( a < b \), and there is no \( c \in (a,b) \) such that \( f'(c) = 0 \), then there is at most one root of \( f \) in \( (a,b) \), that is at most one \( x \in (a,b) \) such that \( f(x) = 0 \). In other words, between any two consecutive roots of \( f \) is at most one root of \( f' \). Hint: suppose there are two roots and see what happens.

Taylor's theorem

Note: half a lecture (optional section)

Derivatives of higher orders

When \( f : (a,b) \to \mathbb{R} \) is differentiable, we obtain a function \( f' : (a,b) \to \mathbb{R} \). The function \( f' \) is called the first derivative of \( f \). If \( f' \) is differentiable, we denote by \( f'' \) the derivative of \( f' \). The function \( f'' \) is called the second derivative of \( f \). We similarly obtain \( f''' \), \( f'''' \), and so on. With a larger number of derivatives the notation would get out of hand; we denote by \( f^{(n)} \) the \( n \)th derivative of \( f \).

When \( f \) possesses \( n \) derivatives, we say \( f \) is \( n \) times differentiable.
Taylor’s theorem

Taylor’s theorem \(^{19}\) is a generalization of the Mean value theorem. Mean value theorem says that up to a small error \(f(x)\) for \(x\) near \(x_0\) can be approximated by \(f(x_0)\), that is \([f(x) = f(x_0) + f'(c)(x-x_0),]\) where the “error” is measured in terms of the first derivative at some point \(c\) between \(x\) and \(x_0\). Taylor’s theorem generalizes this result to higher derivatives. It tells us that up to a small error, any \((n)\) times differentiable function can be approximated at a point \(x_0\) by a polynomial. The error of this approximation behaves like \(((x-x_0)^n)\) near the point \(x_0\). To see why this is a good approximation notice that for a big \((n)\), \(((x-x_0)^n)\) is very small in a small interval around \(x_0\).

For an \((n)\) times differentiable function \(f(x)\) defined near a point \((x_0)\in \{\text{mathbb{R}}\}\), define the \((n)\)th **Taylor polynomial** for \(f(x)\) at \((x_0)\) as \([\text{\begin{split} P_n^{x_0}(x) & := \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \end{split}}\)]

Taylor’s theorem says a function behaves like its \((n)\)th Taylor polynomial. The is really Taylor’s theorem for the first derivative.

\[\text{thm: taylor} \]
Suppose \(f: [a,b] \to \{\text{mathbb{R}}\}\) is a function with \((n)\) continuous derivatives on \([(a,b)]\) and such that \(f^{(n+1)}((x_0)+)\) exists on \((a,b)\). Given distinct points \((x_0)\) and \((x)\) in \([(a,b)]\), we can find a point \((c)\) between \((x_0)\) and \((x)\) such that \([f(x)-P_n^{x_0}(x)+{R_n^{x_0}(x)}\) for \(x\) near \(x_0\).

The term \({R_n^{x_0}(x)}(x) := \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}\) is called the **remainder term**. This form of the remainder term is called the Lagrange form of the remainder. There are other ways to write the remainder term, but we skip those. Note that \((c)\) depends on both \((x)\) and \((x_0)\).

Find a number \((\lambda_{x,x_0})\) (depending on \((x)\) and \((x_0)\)) solving the equation \([f(x)-P_n^{x_0}(x)+\lambda_{x,x_0}(x-x_0)^{n+1}\) for \((x)\). Define a function \((g(s))\) by \([g(s):=\frac{f(s)-P_n^{x_0}(s)-\lambda_{x,x_0}(s-x_0)^{n+1}}{\text{\begin{split} (s-x_0)^{n+1} & = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2}(x-x_0)^2 + \frac{f'''(x_0)}{6}(x-x_0)^3 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n . \end{split}}\) for \((x)\)).

In particular \((g(x_0) = 0)\). On the other hand \((g(x) = 0)\). By the there exists an \((\lambda_{x_1})\) between \((x_0)\) and \((x)\) such that \((g(x_1) = 0)\). Applying the to \((g^n)\) we obtain that there exists \((\lambda_{x_2})\) between \((x_0)\) and \((x_1)\) (and therefore between \((x_0)\) and \((x)\)) such that \((g(x_2) = 0)\). We repeat the argument \((n+1)\) times to obtain a number \((\lambda_{x_{n+1}})\) between \((x_0)\) and \((x_n)\) (and therefore between \((x_0)\) and \((x)\)) such that \((g^{(n+1)}(x_{n+1}) = 0)\).

Let \((\lambda:=x_{n+1})\). We compute the \((n+1)\)th derivative of \((g)\) to find \([g^{(n)}(x_0) = f^{(n)}(x_0)\) for \((k=0,1,2,\ldots,n)\). Therefore the Taylor polynomial has the same derivatives as \((g)\) at \((x_0)\) up to the \((n)\)th derivative. That is why the Taylor polynomial is a good approximation to \((f)\).
The definition of derivative says that a function is differentiable if it is locally approximated by a line. Similarly we mention in passing that there exists a converse to Taylor’s theorem, which we will neither state nor prove, saying that if a function is locally approximated in a certain way by a polynomial of degree \(d\), then it has \(d\) derivatives.

**Exercises**

Compute the \(n\)th Taylor Polynomial at \(0\) for the exponential function.

Suppose \(p\) is a polynomial of degree \(d\). Given any \(x_0 \in \mathbb{R}\), show that the \((d+1)\)th Taylor polynomial for \(p\) at \(x_0\) is equal to \(p'(x_0)\).

Let \(f(x) = \left| x \right|^3\). Compute \(f'(x)\) and \(f''(x)\) for all \(x\), but show that \(f^{(3)}(0)\) does not exist.

Suppose \(f\) has \(n\) continuous derivatives. Show that for any \(x_0 \in \mathbb{R}\), there exist polynomials \(P(x)\) and \(Q(x)\) of degree \(n\) such that \(P(x) \leq f(x) \leq Q(x)\) for all \(x \in [x_0-\varepsilon,x_0+\varepsilon]\) and \(Q(x) - P(x) = \lambda (x-x_0)^n\) for some \(\lambda \geq 0\).

If \(f\) has \(n+1\) continuous derivatives and \(f^{(k)}(x_0) = 0\) for all \(k = 0, 1, 2, \ldots, n\) if and only if \(g(x) := \frac{f(x)}{x^n}\) is continuous at \(x_0\).

If \(f\) has \(n+1\) continuous derivatives and \(f^{(k)}(x_0) = 0\) for all \(k = 0, 1, 2, \ldots, n\) if and only if \(g(x) := \frac{f(x)}{x^n}\) is continuous at \(x_0\).

**Inverse function theorem**

Note: less than 1 lecture (optional section, needed for , requires )
Inverse function theorem

The main idea of differentiating inverse functions is the following lemma.

[lemma:ift] Let \( (I,J) \subseteq \mathbb{R} \) be intervals. If \( f : I \to J \) is strictly monotone (hence one-to-one), onto \((f(I) = J)\), differentiable at \( f(x) \) and \( f'(x) \not= 0 \), then the inverse \( f^{-1} \) is differentiable at \( f(x) \) and \( (f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))} = \frac{1}{f'(x)} \). If \( f \) is continuously differentiable and \( f' \) is never zero, then \( f^{-1} \) is continuously differentiable.

By \( f \) has a continuous inverse. Let us call the inverse \( g : J \to I \) for convenience. Let \( (x,y) \) be as in the statement, take \( (t \in I) \) to be arbitrary and let \( (s := f(t)) \). Then \( \frac{g(s) - g(y)}{s - y} = \frac{g(f(t)) - g(f(x))}{f(t) - f(x)} = \frac{t - x}{f(t) - f(x)} \). As \( f \) is differentiable at \( x \) and \( f'(x) \not= 0 \), then \( \lim_{t \to x} \frac{t - x}{f(t) - f(x)} = \frac{1}{f'(x)} \). Because \( g(s) \to g(y) \) as \( s \to y \), we can plug in \( g(s) \) for \( t \), and \( g(y) \) for \( x \) and take the limit as \( s \to y \), that is, the limit exists. In other words, \( \lim_{s \to y} \frac{g(s) - g(y)}{s - y} = \lim_{t \to x} \frac{f(t) - f(x)}{g(f(t)) - g(f(x))} = \frac{1}{f'(x)} = \frac{1}{f'(g(y))} \). See for the geometric idea.

If both \( f' \) and \( g' \) are continuous, \( f' \) is nonzero at all \( x \), then the lemma applies at all points \( x \in I \) and the resulting function \( g' \) must be continuous.

What is usually called the inverse function theorem is the following result.

Let \( f : (a,b) \to \mathbb{R} \) be a continuously differentiable function, \( x_0 \in (a,b) \) a point where \( f'(x_0) \not= 0 \). Then there exists an interval \( (I,J) \subseteq (a,b) \) with \( f(I) \subseteq J \), the restriction \( f|_I \) is injective with an inverse \( g : J \to f(I) \) defined on \( J := f(I) \), which is continuously differentiable and \( g(y) = \frac{1}{f'(g(y))} \) for all \( y \in J \).

Without loss of generality, suppose \( f'(x_0) > 0 \). As \( f' \) is continuous, there must exist an interval \( (I,J) \subseteq (a,b) \) such that \( f'(x) > 0 \) for all \( x \in I \).

By \( f \) is strictly increasing on \( (I,J) \), and hence the restriction \( f|_I \) bijective onto \( (f(I),J) \). As \( f \) is continuous, then by the (see also ), \( f(I) \) is in interval. Now apply .

If you tried to prove the existence of roots directly as in you may have seen how difficult that endeavor is. However, with the machinery we have built for inverse functions it becomes an almost trivial exercise, and with and the inverse function theorem we prove far more than mere existence.

Given any \( n \in \mathbb{N} \) and any \( x \geq 0 \) there exists a unique number \( y \geq 0 \) (denoted \( y := x^{1/n} \)), such that \( y^n = x \). Furthermore, the function \( g : (0,\infty) \to (0,\infty) \) defined by \( g(x) := x^{1/n} \) is continuously differentiable and \( g'(x) = \frac{1}{n} x^{(1-n)/n} \) using the convention \( x^{1/n} := \{ x^{1/n} \} \).

For \( x=0 \) the existence of a unique root is trivial.
Let $f(x) := x^n$. Using product rule, $f'$ is continuously differentiable and $f'(x) = nx^{n-1}$, see . For $(x > 0)$ the derivative $f'(x)$ is strictly positive and so again by , $f'(x)$ is strictly increasing (this can also be proved directly). It is also easy to see that the image of $f(x)$ is the entire interval $((0, \infty))$. We obtain a unique inverse $g(x)$ and so the existence and uniqueness of positive $(n)$th roots. We apply to obtain the derivative.

The corollary provides a good example of where the inverse function theorem gives us an interval smaller than $((a, b))$. Take $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) := x^2$. Then $f'(x) \not= 0$ as long as $x \not= 0$. If $x_0 > 0$, we can take $I=(0, \infty)$, but no larger.

Another useful example is $f(x) := x^3$. The function $f : \mathbb{R} \to \mathbb{R}$ is one-to-one and onto, so $f^{-1}(x) = x^{1/3}$ exists on the entire real line including zero and negative $(x)$. The function $f$ has a continuous derivative, but $f^{-1}$ has no derivative at the origin. The point is that $f'(0) = 0$. See also .

**Exercises**

Suppose $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable such that $f'(x) > 0$ for all $(x)$. Show that $f$ is invertible on the interval $J = f(\mathbb{R})$, the inverse is continuously differentiable, and $(f^{-1})'(y) > 0$ for all $y \in f(\mathbb{R})$.

Suppose $(I, J)$ are intervals and a monotone onto $f : I \to J$ has an inverse $g : J \to I$. Suppose you already know that both $f$ and $g$ are differentiable everywhere and $f'$ is never zero. Using chain rule but not prove the formula $g'(y) = \frac{1}{f'(g(y))}$.

Let $(n \in \mathbb{N})$ be even. Prove that every $(x > 0)$ has a unique negative $(n)$th root. That is, there exists a negative number $(y)$ such that $(y^n = x)$. Compute the derivative of the function $g(x) := y$.

[exercise:oddroot] Let $(n \in \mathbb{N})$ be odd and $(n \geq 3)$. Prove that every $(x)$ has a unique $(n)$th root. That is, there exists a number $(y)$ such that $(y^n = x)$. Prove that the function defined by $(g(x) := y)$ is differentiable except at $(x=0)$ and compute the derivative. Prove that $(g'(x))$ is not differentiable at $(x=0)$.

§3] Show that if in the inverse function theorem $f$ has $(k)$ continuous derivatives, then the inverse function $g$ also has $(k)$ continuous derivatives.

Let $f(x) := x + 2 x^2 \\sin(\nicefrac{1}{x})$ for $x \not= 0$ and $f(0) = 0$. Show that $(f)$ is differentiable at all $(x)$, that $(f'(0) > 0)$, but that $(f)$ is not invertible on any interval containing the origin.

a) Let $(f : \mathbb{R} \to \mathbb{R})$ be a continuously differentiable function and $(k > 0)$ be a number such that $(f(x) \geq k)$ for all $(x \in \mathbb{R})$. Show $(f)$ is one-to-one and onto, and has a continuously differentiable inverse $(f^{-1} : \mathbb{R} \to \mathbb{R})$. b) Find an example $(f : \mathbb{R} \to \mathbb{R})$ where $(f'(x) > 0)$ for all $(x)$, but $(f)$ is not onto.

Suppose $(I, J)$ are intervals and a monotone onto $f : I \to J$ has an inverse $g : J \to I$. Suppose $(x \in I)$ and $(y := f(x) \in J)$, and that $(g)$ is differentiable at $(y)$. Prove:
a) If \(g'(y) \neq 0\), then \(f\) is differentiable at \(x\).

b) If \(g'(y) = 0\), then \(f\) is not differentiable at \(x\).

---

**The Riemann Integral**

---

**The Riemann integral**

*Note: 1.5 lectures*

We now get to the fundamental concept of integration. There is often confusion among students of calculus between *integral* and *antiderivative*. The integral is (informally) the area under the curve, nothing else. That we can compute an antiderivative using the integral is a nontrivial result we have to prove. In this chapter we define the *Riemann integral*\(^{20}\) using the Darboux integral \(^{21}\), which is technically simpler than (but equivalent to) the traditional definition as done by Riemann.

**Partitions and lower and upper integrals**

We want to integrate a bounded function defined on an interval \([a,b]\). We first define two auxiliary integrals that can be defined for all bounded functions. Only then can we talk about the Riemann integral and the Riemann integrable functions.

A *partition* \(P\) of the interval \([a,b]\) is a finite set of numbers \(\{x_0,x_1,x_2,\ldots,x_n\}\) such that \(a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b\). We write \(\Delta x_i := x_i - x_{i-1}\).

Let \(f \colon [a,b] \to \mathbb{R}\) be a bounded function. Let \(P\) be a partition of \([a,b]\). Define \(\inf \{ f(x) : x_{i-1} \leq x \leq x_i \}\) and \(\sup \{ f(x) : x_{i-1} \leq x \leq x_i \}\) for all \(i\). Then \(L(P,f) := \sum_{i=1}^n m_i \Delta x_i\) and \(U(P,f) := \sum_{i=1}^n M_i \Delta x_i\). We call \(L(P,f)\) the *lower Darboux sum* and \(U(P,f)\) the *upper Darboux sum*.

The geometric idea of Darboux sums is indicated in . The lower sum is the area of the shaded rectangles, and the upper sum is the area of the entire rectangles, shaded plus unshaded parts. The width of the \(i\)th rectangle is \((\Delta x_i)\), the height of the shaded rectangle is \((m_i)\), and the height of the entire rectangle is \((M_i)\).

---

**sumulbound:** Let \(f \colon [a,b] \to \mathbb{R}\) be a bounded function. Let \(m, M \in \mathbb{R}\) be such that for all \(x\) we have \(m \leq f(x) \leq M\). For any partition \(P\) of \([a,b]\) we have \(m(b-a) \leq L(P,f) \leq U(P,f) \leq M(b-a)\). In other words, the set of lower and upper sums are bounded sets.
As the sets of lower and upper Darboux sums are bounded, we define
\[
\underline{\int_a^b} f(x) \text{--dx :=} \sup \{ L(P, f) : P \text{ a partition of } [a,b] \} \quad \& \quad \overline{\int_a^b} f(x) \text{--dx :=} \inf \{ U(P, f) : P \text{ a partition of } [a,b] \} .
\]
We call \(\underline{\int}\) the lower Darboux integral and \(\overline{\int}\) the upper Darboux integral. To avoid worrying about the variable of integration, we often simply write \(\underline{\int_a^b} f := \underline{\int_a^b} f(x) \text{--dx \&} \quad \overline{\int_a^b} f := \overline{\int_a^b} f(x) \text{--dx .}\)

If integration is to make sense, then the lower and upper Darboux integrals should be the same number, as we want a single number to call the integral. However, these two integrals may in fact differ for some functions.

Take the Dirichlet function \(f \colon [0,1] \to \mathbb{R}\), where \(f(x) := 1\) if \(x \in \mathbb{Q}\) and \(f(x) := 0\) if \(x \notin \mathbb{Q}\). Then \(\underline{\int_0^1} f = 0 \quad \text{and} \quad \overline{\int_0^1} f = 1\). The reason is that for every \(i\) we have \(m_i = \inf \{ f(x) : x \in [x_{i-1},x_i] \} = 0\) and \(M_i = \sup \{ f(x) : x \in [x_{i-1},x_i] \} = 1\). Thus \(\begin{aligned} L(P,f) = \sum_{i=1}^n 0 \cdot \Delta x_i = 0 , \quad \& \quad U(P,f) = \sum_{i=1}^n 1 \cdot \Delta x_i = \sum_{i=1}^n \Delta x_i = 1 . \end{aligned}\)

The same definition of \(\underline{\int_a^b} f\) and \(\overline{\int_a^b} f\) is used when \(f\) is defined on a larger set \(S\) such that \([a,b] \subset S\). In that case, we use the restriction of \(f\) to \([a,b]\) and we must ensure that the restriction is bounded on \([a,b]\).

To compute the integral we often take a partition \(P\) and make it finer. That is, we cut intervals in the partition into yet smaller pieces.

Let \(P = \{ x_0, x_1, \ldots, x_n \}\) and \(\widetilde{P} = \{ \widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_m \}\) be partitions of \([a,b]\). We say \(\widetilde{P}\) is a refinement of \(P\) if as sets \(P \subset \widetilde{P}\). That is, \(\widetilde{P}\) is a refinement of a partition if it contains all the numbers in \(P\) and perhaps some other numbers in between. For example, \(\{ 0, 0.5, 1, 2 \}\) is a partition of \((0,2)\) and \(\{ 0, 0.2, 0.5, 1, 1.5, 1.75, 2 \}\) is a refinement. The main reason for introducing refinements is the following proposition.

\[\text{[prop:refinement]}\]
Let \(f\colon [a,b] \to \mathbb{R}\) be a bounded function, and let \(P\) be a partition of \([a,b]\). Let \(\widetilde{P}\) be a refinement of \(P\). Then \(L(P,f) \leq L(\widetilde{P},f) \quad \& \quad U(\widetilde{P},f) \leq U(P,f) .\]

The tricky part of this proof is to get the notation correct. Let \(\widetilde{P} := \{ \widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_m \}\) be a refinement of \(P := \{ x_0, x_1, \ldots, x_n \}\). Then \(\Delta x_0 = \widetilde{x}_0 - x_0 = \widetilde{x}_0 - \widetilde{x}_1 = \Delta \widetilde{x}_1 .\) In fact, we can find integers \(k_0 < k_1 < \cdots < k_n\) such that \(x_{j} = \widetilde{x}_{k_j}\) for \(j=0,1,2,\ldots,n\).

Let \(\Delta \widetilde{P} := \{ \widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_m \}\) be a refinement of \(P := \{ x_0, x_1, \ldots, x_n \}\). Then \(x_0 = \widetilde{x}_0\) and \(x_n = \widetilde{x}_m\). In fact, we can find integers \(k_0 < k_1 < \cdots < k_n\) such that \(x_{j} = \widetilde{x}_{k_j}\) for \(j=0,1,2,\ldots,n\).

Let \(\Delta \widetilde{P} := \{ \widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_{m-1} \}\) be a refinement of \(P := \{ x_0, x_1, \ldots, x_n \}\). Then \(x_0 = \widetilde{x}_0\) and \(x_n = \widetilde{x}_m\). In fact, we can find integers \(k_0 < k_1 < \cdots < k_n\) such that \(x_{j} = \widetilde{x}_{k_j}\) for \(j=0,1,2,\ldots,n\).

Let \(\Delta \widetilde{P} := \{ \widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_{m-1} \}\) be a refinement of \(P := \{ x_0, x_1, \ldots, x_n \}\). Then \(x_0 = \widetilde{x}_0\) and \(x_n = \widetilde{x}_m\). In fact, we can find integers \(k_0 < k_1 < \cdots < k_n\) such that \(x_{j} = \widetilde{x}_{k_j}\) for \(j=0,1,2,\ldots,n\).

Let \(\Delta \widetilde{P} := \{ \widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_{m-1} \}\) be a refinement of \(P := \{ x_0, x_1, \ldots, x_n \}\). Then \(x_0 = \widetilde{x}_0\) and \(x_n = \widetilde{x}_m\). In fact, we can find integers \(k_0 < k_1 < \cdots < k_n\) such that \(x_{j} = \widetilde{x}_{k_j}\) for \(j=0,1,2,\ldots,n\).

Let \(\Delta \widetilde{P} := \{ \widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_{m-1} \}\) be a refinement of \(P := \{ x_0, x_1, \ldots, x_n \}\). Then \(x_0 = \widetilde{x}_0\) and \(x_n = \widetilde{x}_m\). In fact, we can find integers \(k_0 < k_1 < \cdots < k_n\) such that \(x_{j} = \widetilde{x}_{k_j}\) for \(j=0,1,2,\ldots,n\).

Let \(\Delta \widetilde{P} := \{ \widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_{m-1} \}\) be a refinement of \(P := \{ x_0, x_1, \ldots, x_n \}\). Then \(x_0 = \widetilde{x}_0\) and \(x_n = \widetilde{x}_m\). In fact, we can find integers \(k_0 < k_1 < \cdots < k_n\) such that \(x_{j} = \widetilde{x}_{k_j}\) for \(j=0,1,2,\ldots,n\).
\sum_{p=k_{j-1}+1}^{k_j} \Delta \tilde{x}_p = \sum_{p=k_{j-1}+1}^{k_j} m_j \Delta \tilde{x}_p \leq \sum_{p=k_{j-1}+1}^{k_j} \widetilde{m}_p \Delta \tilde{x}_p .

So

\[ L(P, f) = \sum_{j=1}^n m_j \Delta x_j \leq \sum_{j=1}^n \sum_{p=k_{j-1}+1}^{k_j} \widetilde{m}_p \Delta \tilde{x}_p = \sum_{j=1}^m \widetilde{m}_j \Delta \tilde{x}_j = L(\tilde{P}, f). \]

The proof of \( U(\tilde{P}, f) \leq U(P, f) \) is left as an exercise.

Armed with refinements we prove the following. The key point of this next proposition is that the lower Darboux integral is less than or equal to the upper Darboux integral.

\[ \int \text{Intbound:prop} \]

Let \( f \colon [a, b] \to \mathbb{R} \) be a bounded function. Let \( m, M \in \mathbb{R} \) be such that for all \( x \in [a, b] \) we have \( m \leq f(x) \leq M \). Then

\[ \text{label} \int \text{Intbound:eq} \]

\( m(b-a) \leq \underline{\int}_a^b f \leq \overline{\int}_a^b f \leq M(b-a). \]

By we have for any partition \( P \)

\[ m(b-a) \leq \text{leq } L(P, f) \leq U(P, f) \leq \text{leq } M(b-a). \]

The inequality \( m(b-a) \leq \text{leq } L(P, f) \) implies \( \text{leq } \text{underline } \int_a^b f \). Also \( U(P, f) \leq \text{leq } M(b-a) \) implies \( \text{overline } \int_a^b f \leq \text{leq } M(b-a) \).

The key point of this proposition is the middle inequality in \( \int \text{Intbound:eq} \). Let \( (P_1, P_2) \) be partitions of \([a, b]\). Define

\( \text{label } \int \text{Intbound:prop} \)

\( \text{wideltilde } \{ P \} := P_1 \cup \text{cup } P_2 \). \text{The set } \{ \text{wideltilde } \{ P \} \} \text{ is a partition of } \{[a, b]\}. \text{Furthermore, } \{ \text{wideltilde } \{ P \} \} \text{ is a refinement of } \{P_1\} \text{ and it is also a refinement of } \{P_2\}. \text{By we have } \text{leq } L(P_1, f) \leq \text{leq } L(\text{wideltilde } \{ P \}, f) \) and

\[ \text{leq } U(\text{wideltilde } \{ P \}, f) \leq \text{leq } U(P_2, f) . \]

In other words, for two arbitrary partitions \( \{P_1\} \) and \( \{P_2\} \) we have \( \text{leq } L(P_1, f) \leq \text{leq } U(\text{wideltilde } \{ P \}, f) \). Now we recall . Taking the supremum and infimum over all partitions we get

\[ \text{sup } \{ L(P, f) : \text{SPS a partition of } S[a,b] \} \leq \text{leq } \text{inf } \{ U(P, f) : \text{SPS a partition of } S[a,b] \} . \]

In other words \( \text{underline } \int_a^b f \leq \text{overline } \int_a^b f \).

### Riemann integral

We can finally define the Riemann integral. However, the Riemann integral is only defined on a certain class of functions, called the Riemann integrable functions.

Let \( f \colon [a, b] \to \mathbb{R} \) be a bounded function such that \( \text{underline } \int_a^b f(x) \text{dx} = \text{overline } \int_a^b f(x) \text{dx} \) \text{.} \text{Then } \text{label } \int \text{Intbound:prop} \)

\( f \) is said to be \text{Riemann integrable}. \text{The set of Riemann integrable functions on } \{[a, b]\} \text{ is denoted by } \{ \text{mathcal } \{ R \} \}[a,b]. \text{When } \{ f \in \text{mathcal } \{ R \} [a,b] \} \text{ we define } \{ \text{int}_a^b f(x) \text{dx} := \text{underline } \{ \text{int}_a^b f(x) \text{dx} = \text{overline } \{ \text{int}_a^b f(x) \text{dx} \} \text{. As before, we often simply write } \{ \text{int}_a^b f := \text{int}_a^b f(x) \text{dx} \} \text{. The number } \{ \text{int}_a^b f \} \text{ is called the Riemann integral of } \{ f \}, \text{ or sometimes simply the integral of } \{ f \}. \text{By definition, any Riemann integrable function is bounded. By appealing to we immediately obtain the following proposition.} \text{\label } \int \text{Intbound:prop} \)

Let \( f \colon [a, b] \to \mathbb{R} \) be a Riemann integrable function. Let \( m, M \in \mathbb{R} \) be such that \( m \leq f(x) \leq M \) for all \( x \in [a, b] \). Then \( m(b-a) \leq \text{leq } \text{int}_a^b f \leq \text{leq } M(b-a) . \)

Often we use a weaker form of this proposition. That is, if \( \text{left\lvert } f(x) \text{right\rvert } \leq M \) for all \( x \in [a, b] \), then

\[ \text{left\lvert } \text{int}_a^b f \text{right\rvert } \leq M(b-a) . \]
We integrate constant functions using \( f(x) := c \) for some constant \( c \), then we take \( m = M = c \). In inequality \([\text{intulbound:eq}]\) all the inequalities must be equalities. Thus \( f \) is integrable on \([a,b] \) and \( \int_a^b f = c(b-a) \).

Let \( f \colon [0,2] \to \mathbb{R} \) be defined by \( f(x) := \begin{cases} 1 & \text{if } x < 1, \\ \frac{1}{2} & \text{if } x = 1, \\ 0 & \text{if } x > 1. \end{cases} \) We claim \( f \) is Riemann integrable and \( \int_0^2 f = 1 \).

Proof: Let \( 0 < \epsilon < 1 \) be arbitrary. Let \( P := \{0, 1-\epsilon, 1+\epsilon, 2\} \) be a partition. We use the notation from the definition of the Darboux sums. Then \( \{\text{begin[aligned]} \ m_1 \ &= \ \inf \{ f(x) : x \in [0,1-\epsilon] \} = 1, \\ M_1 \ &= \ \sup \{ f(x) : x \in [0,1-\epsilon] \} = 1 \}, \quad \{\text{end[aligned]}\} \). Furthermore, \( \{\text{aligned}\} \) \( \Delta x_1 = 1-\epsilon \), \( \Delta x_2 = 2\epsilon \) and \( \Delta x_3 = 1-\epsilon \). We compute \( L(P,f) = \sum_{i=1}^3 m_i \Delta x_i = 1 \cdot (1-\epsilon) + 0 \cdot 2\epsilon + 0 \cdot (1-\epsilon) = 1-\epsilon, \) \quad \( U(P,f) = \sum_{i=1}^3 M_i \Delta x_i = 1 \cdot (1-\epsilon) + 1 \cdot 2\epsilon + 0 \cdot (1-\epsilon) = 1+\epsilon \). Thus, \( \overline{\int_0^2} f - \underline{\int_0^2} f \leq U(P,f) - L(P,f) = 2 \epsilon \). By we have \( \overline{\int_0^2} f \leq \underline{\int_0^2} f \). As \( \epsilon \) was arbitrary we see \( \overline{\int_0^2} f = \underline{\int_0^2} f \). So \( f \) is Riemann integrable. Finally, \( \int_0^2 f = \underline{\int_0^2} f = 1 \). Hence, \( \big| \int_0^2 f - 1 \big| \leq \epsilon \). As \( \epsilon \) was arbitrary, we have \( \int_0^2 f = 1 \).

It may be worthwhile to extract part of the technique of the example into a proposition.

Let \( f \colon [a,b] \to \mathbb{R} \) be a bounded function. Then \( f \) is Riemann integrable if for every \( \epsilon > 0 \), there exists a partition \( P \) such that \( U(P,f) - L(P,f) < \epsilon \).

If for every \( \epsilon > 0 \), a \( P \) exists we have: \( \overline{\int_a^b} f = \underline{\int_a^b} f \). Therefore, \( \overline{\int_a^b} f = \underline{\int_a^b} f \), and \( f \) is integrable.

Let us show \( \frac{1}{1+x} \) is integrable on \([0,b]\) for any \( b > 0 \). We will see later that all continuous functions are integrable, but let us demonstrate how we do it directly.

Let \( 0 < \epsilon < 1 \) be given. Take \( n \in \{\text{mathbb\{N\}}\} \) and pick \( x_j := \nicefrac{jb}{n} \), to form the partition \( P := \{x_0,x_1,\ldots,x_n\} \) of \([0,b]\). We have \( \Delta x_j = \nicefrac{jb}{n} \) for all \( j \). As \( f \) is decreasing, for any subinterval \( \{x_{j-1},x_j\} \) we obtain \( m_j = \inf \{ f(x) : x \in [x_{j-1},x_j] \} = \frac{1}{1+x_j}, \) \quad \( M_j = \sup \{ f(x) : x \in [x_{j-1},x_j] \} = \frac{1}{1+x_{j-1}} \). Then we have \( \overline{\int_a^b} f - \underline{\int_a^b} f = \sum_{j=1}^n \Delta x_j (M_j-m_j) = \sum_{j=1}^n \left( \frac{1}{1+x_j} - \frac{1}{1+x_{j-1}} \right) = \frac{b}{n(b+1)} \). The sum telescopes, the terms successively cancel each other, something we have seen before. Picking \( n \) to be such that \( \frac{b}{n(b+1)} < \epsilon \) the proposition is satisfied and the function is integrable.
More notation

When \( f : S \to \mathbb{R} \) is defined on a larger set \( S \) and \( [a,b] \subset S \), we say \( f \) is Riemann integrable on \( [a,b] \) if the restriction of \( f \) to \( [a,b] \) is Riemann integrable. In this case, we say \( f \) is \( \mathcal{R}[a,b] \), and we write \( \int_a^b f \) to mean the Riemann integral of the restriction of \( f \) to \( [a,b] \).

It is useful to define the integral \( \int_a^b f \) even if \( a \not< b \). Suppose \( b < a \) and \( f \in \mathcal{R}[b,a] \), then define \( \int_a^b f := - \int_b^a f \). For any function \( f \) we define \( \int_a^a f := 0 \).

At times, the variable \( x \) may already have some other meaning. When we need to write down the variable of integration, we may simply use a different letter. For example, \( \int_a^b f(s) \, ds := \int_a^b f(x) \, dx \).

Exercises

Let \( f : [0,1] \to \mathbb{R} \) be defined by \( f(x) := x^3 \) and let \( P := \{ 0, 0.1, 0.4, 1 \} \). Compute \( L(P,f) \) and \( U(P,f) \).

Let \( f : [0,1] \to \mathbb{R} \) be defined by \( f(x) := x \). Show that \( f \in \mathcal{R}[0,1] \) and compute \( \int_0^1 f \) using the definition of the integral (but feel free to use the propositions of this section).

Let \( f : [a,b] \to \mathbb{R} \) be a bounded function. Suppose there exists a sequence of partitions \( \{ P_k \} \) of \( [a,b] \) such that \( \lim_{k \to \infty} (U(P_k,f) - L(P_k,f)) = 0 \). Show that \( f \) is Riemann integrable and that \( \int_a^b f = \lim_{k \to \infty} U(P_k,f) = \lim_{k \to \infty} L(P_k,f) \).

Finish the proof of .

Suppose \( f : [-1,1] \to \mathbb{R} \) is defined as \( f(x) := \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases} \) Prove that \( f \) is Riemann integrable on \( [-1,1] \) and compute \( \int_{-1}^1 f \) using the definition of the integral (but feel free to use the propositions of this section).

Let \( c \in (a,b) \) and define \( f : [a,b] \to \mathbb{R} \) as \( f(x) := \begin{cases} d & \text{if } x = c, \\ 0 & \text{if } x \not= c. \end{cases} \) Prove that \( f \) is Riemann integrable on \( [a,b] \) and compute \( \int_a^b f \) using the definition of the integral (but feel free to use the propositions of this section).

[exercise:taggedpartition] Suppose \( f : [a,b] \to \mathbb{R} \) is Riemann integrable. Let \( \epsilon > 0 \) be given. Then show that there exists a partition \( P = \{ x_0, x_1, \ldots, x_n \} \) such that if we pick any set of numbers \( \{ c_1, c_2, \ldots, c_n \} \) with \( c_k \in [x_{k-1},x_k] \) for all \( k \), then \( \left| \sum_{k=1}^n f(c_k) \Delta x_k - \int_a^b f \right| < \epsilon \).

Let \( f : [a,b] \to \mathbb{R} \) be a Riemann integrable function. Let \( \alpha > 0 \) and let \( \beta \in \mathbb{R} \). Then define \( g(x) := f(\alpha x + \beta) \) on the interval \( I = [\frac{a-\beta}{\alpha}, \frac{b-\beta}{\alpha}] \). Show that \( g \) is Riemann integrable on \( I \).
Suppose \( f \colon [0,1] \to \mathbb{R} \) and \( g \colon [0,1] \to \mathbb{R} \) are such that for all \( x \in (0,1] \) we have \( f(x) = g(x) \). Suppose \( f \) is Riemann integrable. Prove \( g \) is Riemann integrable and \( \int_{0}^{1} f = \int_{0}^{1} g \).

Let \( f \colon [0,1] \to \mathbb{R} \) be a bounded function. Let \( P_n = [x_0,x_1,\ldots,x_n] \) be a uniform partition of \([0,1] \), that is, \( x_j := \text{nicefrac}(j,n) \). Is \( \{ L(P_n,f) \} \) always monotone? Yes/No: Prove or find a counterexample.

For a bounded function \( f \colon [0,1] \to \mathbb{R} \) let \( R_n := \text{nicefrac}(\{1\}\{n\}) \text{sum}_{j=1}^{n} f(\text{nicefrac}(j,n)) \) (the uniform right hand rule). a) If \( f \) is Riemann integrable show \( \int_{0}^{1} f = \lim R_n \). b) Find an \( f \) that is not Riemann integrable, but \( \lim R_n \) exists.

[exercise:riemannintdarboux] Generalize the previous exercise. Show that \( f \in \mathcal{R}([a,b]) \) if and only if there exists an \( I \in \mathbb{R} \) such that for every \( \epsilon > 0 \) there exists a \( \delta > 0 \) such that if \( P \) is a partition with \( (\Delta x_i < \delta) \) for all \( i \), then \( \left| L(P,f) - I \right| < \epsilon \) and \( \left| U(P,f) - I \right| < \epsilon \). If \( f \in \mathcal{R}([a,b]) \), then \( I = \int_{a}^{b} f \).

[exercise:riemannintdarboux2] Using and the idea of the proof in , show that Darboux integral is the same as the standard definition of Riemann integral, which you have most likely seen in calculus. That is, show that \( f \in \mathcal{R}([a,b]) \) if and only if there exists an \( I \in \mathbb{R} \) such that for every \( \epsilon > 0 \) there exists a \( \delta > 0 \) such that if \( P = [x_0,x_1,\ldots,x_n] \) is a partition with \( (\Delta x_i < \delta) \) for all \( i \), then \( \left| L(P,f) - I \right| < \epsilon \) and \( \left| U(P,f) - I \right| < \epsilon \) for any set \( \{ c_1,c_2,\ldots,c_n \} \) with \( c_i \in [x_{i-1},x_i] \). If \( f \in \mathcal{R}([a,b]) \), then \( I = \int_{a}^{b} f \).

Find an example of functions \( f \colon [0,1] \to \mathbb{R} \) which is Riemann integrable, and \( g \colon [0,1] \to \mathbb{R} \) which is one-to-one and onto, such that the composition \( f \circ g \) is not Riemann integrable.

---

**Properties of the integral**

*Note: 2 lectures, integrability of functions with discontinuities can safely be skipped*

**Additivity**

The next result we prove is usually referred to as the additive property of the integral. First we prove the additivity property for the lower and upper Darboux integrals.

[lemma:darbouxadd] Suppose \( a < b < c \) and \( f \colon [a,c] \to \mathbb{R} \) is a bounded function. Then
\[
\underline{\int_{a}^{c}} f = \underline{\int_{a}^{b}} f + \underline{\int_{b}^{c}} f \quad \text{and} \quad \overline{\int_{a}^{c}} f = \overline{\int_{a}^{b}} f + \overline{\int_{b}^{c}} f.
\]

If we have partitions \( P_1 = [x_0,x_1,\ldots,x_k] \) of \([a,b]\) and \( P_2 = [x_k,x_{k+1},\ldots,x_n] \) of \([b,c]\), then the set \( P := \text{cup} P_2 = [x_0,x_1,\ldots,x_n] \) is a partition of \([a,c]\). Then \( \int(P,f) = \sum_{j=1}^{n} m_j \Delta x_j = \sum_{j=1}^{k} m_j \Delta x_j + \sum_{j=k+1}^{n} m_j \Delta x_j = L(P_1,f) + L(P_2,f) \). When we take the
supremum of the right hand side over all \( (P_1) \) and \( (P_2) \), we are taking a supremum of the left hand side over all partitions \( (P) \) of \( ([a,c]) \) that contain \( (b) \). If \( (Q) \) is any partition of \( ([a,c]) \) and \( (P = Q \cup \{ b \}) \), then \( (P) \) is a refinement of \( (Q) \) and so \( (L(Q,f) \leq L(P,f)) \). Therefore, taking a supremum only over the \( (P) \) that contain \( (b) \) is sufficient to find the supremum of \( (L(P,f)) \) over all partitions \( (P) \), see . Finally recall to compute \[ \int_a^c f = \sup \{ L(P,f) : \text{P a partition of } [a,c] \} = \sup \{ L(P_1,f) + L(P_2,f) : \text{P_1 a partition of } [a,b], P_2 a partition of } [b,c] \} \]

Similarly, for \( (P), (P_1), \) and \( (P_2) \) as above we obtain \[ U(P,f) = \sum_{i=1}^n M_i \Delta x_i = \sum_{i=1}^k M_i \Delta x_i + \sum_{i=k+1}^n M_i \Delta x_i = U(P_1,f) + U(P_2,f) .\]

Let \( (a < b < c) \). A function \( f \colon [a,c] \to \mathbb{R} \) is Riemann integrable if and only if \( (f) \) is Riemann integrable on \( ([a,b]) \) and \( ([b,c]) \). If \( (f) \) is Riemann integrable, then \[ \int_a^c f = \int_a^b f + \int_b^c f .\]

Suppose \( (f) \in \mathcal{R}([a,b]) \), then \[ \int_a^c f = \underline{\int_a^b} f + \underline{\int_b^c} f = \overline{\int_a^b} f + \overline{\int_b^c} f .\]

Now assume the restrictions of \( (f) \) to \( ([a,b]) \) and to \( ([b,c]) \) are Riemann integrable. We again apply the lemma to get \[ \int_a^c f = \underline{\int_a^b} f + \underline{\int_b^c} f = \overline{\int_a^b} f + \overline{\int_b^c} f .\]

An easy consequence of the additivity is the following corollary. We leave the details to the reader as an exercise.

**Linearity and monotonicity**

Let \( (f) \) and \( (g) \) be in \( \mathcal{R}([a,b]) \) and \( \alpha \in \mathbb{R} \). Then \[ \int_a^b (f + g) = \int_a^b f + \int_a^b g ,\]

i. \( \alpha \int_a^b f = \int_a^b \alpha f ,\]

ii. \( (f \geq g) \) on \( [a,b] \) implies \( \int_a^b f \geq \int_a^b g .\]

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
ii. \((f+g)\) is in \((\text{mathcal \{R\}} \{[a,b]\}) \) and \((\text{int}_a^b (f(x)+g(x)) \, \text{bigr}) \sim dx = \text{int}_a^b f(x) \sim dx + \text{int}_a^b g(x) \sim dx \).

Let us prove the first item. First suppose \((\alpha \geq 0)\). Let \((\{m_i := \inf \{ f(x) : x \in [x_{i-1},x_i] \}\})\) as usual. Since \((\alpha \geq 0)\) is nonnegative, we can move the multiplication by \((\alpha)\) past the infimum, \((\inf \{ \alpha f(x) : x \in [x_{i-1},x_i] \} = \alpha \inf \{ f(x) : x \in [x_{i-1},x_i] \} = \alpha m_i \}).\) Therefore \(\text{L}(P,\alpha f) = \sum_{i=1}^n \alpha m_i \Delta x_i = \alpha \sum_{i=1}^n m_i \Delta x_i = \alpha \text{L}(P,f).\) Similarly \(\text{U}(P,\alpha f) = \alpha \text{U}(P,f).\) Again, as \((\alpha \geq 0)\) we may move multiplication by \((\alpha)\) past the supremum. Hence, \(\underline{\int_a^b} \alpha f(x) \sim dx = \alpha \underline{\int_a^b} f(x) \sim dx, \) and \(\overline{\int_a^b} \alpha f(x) \sim dx = \alpha \overline{\int_a^b} f(x) \sim dx.\) The conclusion now follows for \((\alpha \geq 0)\).

To finish the proof of the first item, we need to show that \((-f)\) is Riemann integrable and \((\text{int}_a^b - f(x) \sim dx = -\text{int}_a^b f(x) \sim dx).\) The proof of this fact is left as an exercise.

The proof of the second item in the proposition is also left as an exercise. It is not as trivial as it may appear at first glance.

We should note that the second item in the proposition does not hold with equality for the Darboux integrals. For arbitrary bounded functions \((f)\) and \((g)\) we only obtain \(\overline{\int_a^b} (f(x)+g(x)) \leq \overline{\int_a^b} f(x) + \overline{\int_a^b} g(x)\) and \(\underline{\int_a^b} (f(x)+g(x)) \geq \underline{\int_a^b} f(x) + \underline{\int_a^b} g(x).\)

Let \(f\) and \((g)\) be in \((\text{mathcal \{R\}} \{[a,b]\})\) and let \((f(x) \sim dx = \text{int}_a^b f(x) \sim dx).\) The proof of this fact is left as an exercise.

**Continuous functions**

Let us show that continuous functions are Riemann integrable. In fact we will show we can even allow some discontinuities.

We start with a function continuous on the whole closed interval \(([a,b]).\)

\[\text{lemma:contint} \text{ If } (f: \text{colon} \{[a,b]\} \to \text{mathbb \{R\}}) \text{ is a continuous function, then } (f: \text{in} \ \text{mathcal \{R\}} \{[a,b]\}).\]

As \((f)\) is continuous on a closed bounded interval, it is uniformly continuous. Let \((\epsilon > 0)\) be given. Find a \((\delta > 0)\) such that \((\text{left} \{x-y\} \right\{x-y\} \epsilon < \delta)\) implies \((\text{left} \{f(x)-f(y)\} \right\{f(x)-f(y)\} \epsilon < \frac{1}{n} \epsilon).\)
Let \( P = \{ x_0, x_1, \ldots, x_n \} \) be a partition of \([a,b]\) such that \( \Delta x_i < \delta \) for all \( i = 1, 2, \ldots, n \). For example, take \( \{n\} \) such that \( \frac{b-a}{n} < \delta \) and let \( x_i := \frac{i}{n}(b-a) + a \). Then for all \( x, y \in [x_{i-1},x_i] \) we have \( \left| x - y \right| < \delta \) and so \( \int (x-y) f(x) \delta \). As \( f \) is continuous on \([x_{i-1},x_i]\), it attains a maximum and a minimum on this interval. Let \( x \) be a point where \( f \) attains the maximum and \( y \) be a point where \( f \) attains the minimum. Then \( \int f(x) \delta \). As \( \epsilon > 0 \) was arbitrary, \( \int_a^b f = \int_a^b f \), and \( f \) is Riemann integrable on \([a,b]\).

The second lemma says that we need the function to only be “Riemann integrable inside the interval,” as long as it is bounded. It also tells us how to compute the integral.

[lemma:boundedimpriemann] Let \( f \colon [a,b] \to \mathbb{R} \) be a bounded function that is Riemann integrable on \([a',b']\) for all \( a',b' \) such that \( a < a' < b' < b \). Then \( \int_{a'}^{b'} f \). Furthermore, if \( a < a_n < b_n < b \) are such that \( \lim \Delta a_n = a \) and \( \lim b_n = b \), then \( \int_{a_n}^{b_n} f = \lim_{n \to \infty} \int_{a_n}^{b_n} f \).

Let \( M > 0 \) be a real number such that \( \left| f(x) \right| < M \). Pick two sequences of numbers \( a < a_n < b_n < b \) such that \( \lim a_n = a \) and \( \lim b_n = b \). Note \( \lim (b-a) \left( b_n-a_n \right) = \lim (b-a) \left( b_n-a_n \right) \). Therefore the sequence of numbers \( \left| \int_{a_n}^{b_n} f \right| \sum_{i=1}^{n} \int_{a_{n_i}}^{b_{n_i}} f \) is bounded and has a convergent subsequence indexed by \( k \). Let us call \( L \) the limit of the subsequence \( \left| \int_{a_n}^{b_n} f \right| \sum_{i=1}^{k} \). It says that the lower and upper integral are additive and the hypothesis says that \( \int_a^b f \) is integrable on \([a_n,b_n]\).

Therefore \( \int_a^b f = \lim_{n \to \infty} \int_{a_n}^{b_n} f + \int_{a_{n_k}}^{b_{n_k}} f \). We take the limit as \( k \) goes to \( \infty \) on the right-hand side, \( \int_{a_n}^{b_n} f \left( b_n-a_n \right) \). We use additivity of the upper integral, \( \overline{\int_{a_n}^{b_n} f} = \overline{\int_{a_n}^{b_n} f} + \int_{a_{n_k}}^{b_{n_k}} f \). We take the limit as \( k \) goes to \( \infty \) on the right-hand side, \( \int_{a_n}^{b_n} f \left( b_n-a_n \right) \). We prove the final statement of the lemma we use. We have shown that every convergent subsequence \( \left( \int_{a_n}^{b_n} f \right) \) converges to \( L = \int_{a_n}^{b_n} f \). Therefore, the sequence \( \left( \int_{a_n}^{b_n} f \right) \) is convergent and converges to \( L \).

We say a function \( f \colon [a,b] \to \mathbb{R} \) has \( \text{finitely many discontinuities} \) if there exists a finite set \( S := \{ x_1, x_2, \ldots, x_n \} \subset [a,b] \), and \( f \) is continuous at all points of \( f \).
Let \( f \colon [a,b] \to \mathbb{R} \) be a bounded function with finitely many discontinuities. Then \( f \in \mathcal{R}[a,b] \).

We divide the interval into finitely many intervals \([(a_i,b_i)]\) so that \( f \) is continuous on the interior \((a_i,b_i)\). If \( f \) is continuous on \((a_i,b_i)\), then it is continuous and hence integrable on \([(c_i,d_i)]\) whenever \(a_i < c_i < d_i < b_i\). By the restriction of \( f \) to \([(a_i,b_i)]\) is integrable. By additivity of the integral (and ) \( f \) is integrable on the union of the intervals.

Sometimes it is convenient (or necessary) to change certain values of a function and then integrate. The next result says that if we change the values only at finitely many points, the integral does not change.

Let \( f \colon [a,b] \to \mathbb{R} \) be Riemann integrable. Let \( g \colon [a,b] \to \mathbb{R} \) be a function such that \( f(x) = g(x) \) for all \( x \in [a,b] \setminus S \), where \( S \) is a finite set. Then \( g \) is a Riemann integrable function and \( \int_a^b g = \int_a^b f \).

Using additivity of the integral, we split up the interval \([a,b]\) into smaller intervals such that \( f(x) = g(x) \) holds for all \( x \) except at the endpoints (details are left to the reader).

Therefore, without loss of generality suppose \( f(x) = g(x) \) for all \( x \in (a,b) \). The proof follows by , and is left as an exercise.

**Exercises**

Let \( f \) be in \( \mathcal{R}[a,b] \). Prove that \( -f \) is in \( \mathcal{R}[a,b] \) and \( \int_a^b - f(x) ~dx = - \int_a^b f(x) ~dx \).

Let \( f \) and \( g \) be in \( \mathcal{R}[a,b] \). Prove that \( f+g \) is in \( \mathcal{R}[a,b] \) and \( \int_a^b (f(x)+g(x)) ~dx = \int_a^b f(x) ~dx + \int_a^b g(x) ~dx \). Hint: Use to find a single partition \( P \) such that \( U(P,f)+L(P,f) \) and \( U(P,g)+L(P,g) \) are close to \( \epsilon \).

Let \( f \colon [a,b] \to \mathbb{R} \) be Riemann integrable. Let \( g \colon [a,b] \to \mathbb{R} \) be a function such that \( f(x) = g(x) \) for all \( x \in (a,b) \). Prove that \( g \) is Riemann integrable and that \( \int_a^b g = \int_a^b f \).

Prove the **mean value theorem for integrals**. That is, prove that if \( f \colon [a,b] \to \mathbb{R} \) is continuous, then there exists a \( c \in [a,b] \) such that \( \int_a^b f = f(c)(b-a) \).

If \( f \colon [a,b] \to \mathbb{R} \) is a continuous function such that \( f(x) \geq 0 \) for all \( x \in [a,b] \) and \( \int_a^b f = 0 \). Prove that \( f(x) = 0 \) for all \( x \in [a,b] \).

If \( f \colon [a,b] \to \mathbb{R} \) is a continuous function for all \( x \in [a,b] \) and \( \int_a^b f = 0 \). Prove that there exists a \( c \in [a,b] \) such that \( f(c) = 0 \) (Compare with the previous exercise).

If \( f \colon [a,b] \to \mathbb{R} \) and \( g \colon [a,b] \to \mathbb{R} \) are continuous functions such that \( \int_a^b f = \int_a^b g \). Then show that there exists a \( c \in [a,b] \) such that \( f(c) = g(c) \).
Let \( f \in \mathcal{R}[\alpha, \beta] \). Let \( \alpha, \beta, \gamma \) be arbitrary numbers in \([a, b]\) (not necessarily ordered in any way). Prove \( \int_\alpha^\gamma f = \int_\alpha^\beta f + \int_\beta^\gamma f \). Recall what \( \int_a^b f \) means if \( b \leq a \).

Prove .

**Exercise**: Suppose \( f : [a, b] \to \mathbb{R} \) is bounded and has finitely many discontinuities. Show that as a function of \( x \) the expression \( \left| \left( \int \left| \left( \int_a^b f(x) \right) \right| dx \right) \right| \) is bounded with finitely many discontinuities and is thus Riemann integrable. Then show \( \| \int_a^b f(x) \| dx \leq \int_a^b \| f(x) \| dx \).

Show that the Thomae or popcorn function (see ) is Riemann integrable. Therefore, there exists a function discontinuous at all rational numbers (a dense set) that is Riemann integrable.

In particular, define \( f : [0, 1] \to \mathbb{R} \) by \( f(x) = \begin{cases} \frac{1}{n} & \text{if } x = \frac{m}{n} \text{ where } m, n \in \mathbb{N} \text{ and } m \text{ and } n \text{ have no common divisors,} \\ 0 & \text{otherwise}. \end{cases} \)

Show \( \int_0^1 f = 0 \).

If \( I \subset \mathbb{R} \) is a bounded interval, then the function \( \varphi_I(x) := \begin{cases} 1 & \text{if } x \in I, \\ 0 & \text{otherwise}, \end{cases} \) is called an elementary step function.

Let \( I \) be an arbitrary bounded interval (you should consider all types of intervals: closed, open, half-open) and \( (a < b) \), then using only the definition of the integral show that the elementary step function \( \varphi_I(x) \) is integrable on \([a, b]\), and find the integral in terms of \( (a), (b) \), and the endpoints of \( I \).

When a function \( f \) can be written as \( f(x) = \sum_{k=1}^n \alpha_k \varphi_{I_k}(x) \) for some real numbers \( \alpha_1, \alpha_2, \ldots, \alpha_n \) and some bounded intervals \( I_1, I_2, \ldots, I_n \), then \( f \) is called a step function.

Using the previous exercise, show that a step function is integrable on any interval \([a, b]\)). Furthermore, find the integral in terms of \( (a), (b) \), the endpoints of \( I_k \) and the \( \alpha_k \).

**Exercise**: Let \( f : [a, b] \to \mathbb{R} \) be increasing. a) Show that \( f \) is Riemann integrable. Hint: Use a uniform partition; each subinterval of same length. b) Use part a to show that a decreasing function is Riemann integrable. c) Suppose \( (h = f-g) \) where \( f \) and \( g \) are increasing functions on \([a, b]\). Show that \( h \) is Riemann integrable.

**Exercise**: Suppose \( f \in \mathcal{R}[\alpha, \beta] \), then the function that takes \( (x) \) to \( \left| \left( \int_a^b f(x) \right) \right| 

**Exercise**: Suppose \( f, g \) are bounded. a) Show \( \overline{\left( \int_a^b f(x) \right)} \) and \( \underline{\left( \int_a^b g(x) \right)} \) where the inequality is strict. Hint: \( (f) \) and \( (g) \) should not be Riemann integrable.

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
In this chapter we discuss and prove the fundamental theorem of calculus. The entirety of integral calculus is built upon this theorem, ergo the name. The theorem relates the seemingly unrelated concepts of integral and derivative. It tells us how to compute the antiderivative of a function using the integral and vice-versa.

**First form of the theorem**

Let \( F \colon [a,b] \to \mathbb{R} \) be a continuous function, differentiable on \((a,b))\). Let \( f \in \mathcal{R}[a,b] \) be such that \( f(x) = F'(x) \) for \( x \in (a,b) \). Then \( \int_a^b f = F(b)-F(a) \).

It is not hard to generalize the theorem to allow a finite number of points in \([a,b]\) where \( F \) is not differentiable, as long as it is continuous. This generalization is left as an exercise.

Let \( P = \{ x_0, x_1, \ldots, x_n \} \) be a partition of \([a,b]\). For each interval \([x_{i-1},x_i]\), use the to find a \( c_i \in (x_{i-1},x_i) \) such that \( f(c_i) \Delta x_i = F'(c_i) (x_i - x_{i-1}) = F(x_i) - F(x_{i-1}) \). Using the notation from the definition of the integral, we have \( m_i \leq f(c_i) \leq M_i \) and so \( m_i \Delta x_i \leq F(x_i) - F(x_{i-1}) \leq M_i \Delta x_i \). We sum over \( i = 1,2, \ldots, n \) to get \( m_i \Delta x_i \leq \sum_{i=1}^n \bigl(F(x_i) - F(x_{i-1})\bigr) \leq M_i \Delta x_i \). We sum over \( i = 1,2, \ldots, n \) to get \( m_i \Delta x_i \leq \sum_{i=1}^n \bigl(F(x_i) - F(x_{i-1})\bigr) \leq M_i \Delta x_i \). The inequalities must be equalities and we are done.

The theorem is used to compute integrals. Suppose we know that the function \( f(x) \) is a derivative of some other function \( F(x) \), then we can find an explicit expression for \( \int_a^b f \).

Suppose we are trying to compute \( \int_0^1 x^2 \, dx \). We notice \( x^2 \) is the derivative of \( \frac{x^3}{3} \). We use the fundamental theorem to write \( \int_0^1 x^2 \, dx = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3} \).

**Second form of the theorem**

The second form of the fundamental theorem gives us a way to solve the differential equation \( F'(x) = f(x) \), where \( f \) is a known function and we are trying to find an \( F \) that satisfies the equation.

Let \( f \colon [a,b] \to \mathbb{R} \) be a Riemann integrable function. Define \( F(x) := \int_a^x f \). First, \( F \) is continuous on \([a,b]\). Second, if \( f \) is continuous at \( c \in [a,b] \), then \( F \) is differentiable at \( c \) and \( F'(c) = f(c) \).

As \( f \) is bounded, there is an \( M > 0 \) such that \( \|f\| \leq M \) for all \( x \in [a,b] \). Suppose \( x,y \in [a,b] \) with \( x > y \). Then \( \|f\| \leq \|F(x)-F(y)\| \). The inequalities must be equalities and we are done.
By symmetry, the same also holds if \( x < y \). So \( F \) is Lipschitz continuous and hence continuous.

Now suppose \( \epsilon \) is continuous at \( c \). Let \( \epsilon > 0 \) be given. Let \( \delta > 0 \) be such that for \( x \in [a,b] \) \( |x-c| < \delta \) implies \( |f(x) - f(c)| < \epsilon \). In particular, for such \( x \) we have \( |f(c)| + \epsilon \leq |f(x)| \leq |f(c)| + \epsilon \). Thus if \( x > c \), then
\[
|f(c)| - \epsilon \leq f(x) \leq f(c) + \epsilon.
\]
When \( c > x \), then the inequalities are reversed. Therefore, assuming \( c \not= x \) we get

\[
|f(c)| - \epsilon \leq \int_c^x f \leq f(c) + \epsilon.
\]
As \( \int_c^c f \leq \epsilon \), as \( f \) is continuous at \( c \), we can write
\[
|f(c)| - \epsilon \leq \frac{\int_c^x f}{x-c} \leq f(c) + \epsilon.
\]

We have
\[
\frac{F(x) - F(c)}{x-c} = \frac{\int_a^x f - \int_a^c f}{x-c} = \frac{\int_c^c f}{x-c}.
\]
Hence
\[
|\frac{F(x) - F(c)}{x-c} - f(c)| \leq \epsilon.
\]
The result follows. It is left to the reader to see why is it OK that we just have a non-strict inequality.

Of course, if \( f \) is continuous on \([a,b]\), then it is automatically Riemann integrable, \( F \) is differentiable on all of \([a,b]\) and \( F'(x) = f(x) \) for all \( x \in [a,b] \).

The second form of the fundamental theorem of calculus still holds if we let \( d \in [a,b] \) and define \( F(x) := \int_d^x f \). That is, we can use any point of \([a,b]\) as our base point. The proof is left as an exercise.

Let us look at what a simple discontinuity can do. Take \( f(x) := -1 \) if \( x < 0 \), and \( f(x) := 1 \) if \( x \geq 0 \). Let \( F(x) := \int_0^x f \). It is not difficult to see that \( F(x) = \left\lfloor x \right\rfloor \). Notice that \( f \) is discontinuous at \( 0 \) and \( F \) is not differentiable at \( 0 \). However, the converse does not hold. Let us do another quick example. Let \( g(x) := 0 \) if \( x \not= 0 \), and \( g(0) = 1 \). Letting \( G(x) := \int_0^x g \), we find that \( G(x) = 0 \) for all \( x \). So \( g \) is discontinuous at \( 0 \), but \( G(0) \) exists and is equal to 0.

A common misunderstanding of the integral for calculus students is to think of integrals whose solution cannot be given in closed-form as somehow deficient. This is not the case. Most integrals we write down are not computable in closed-form.

Even some integrals that we consider in closed-form are not really such. For example, how does a computer find the value of \( \ln x \)? One way to do it is to note that we define the natural log as the antiderivative of \( \frac{1}{x} \) such that \( \ln 1 = 0 \). Therefore, \( \ln x = \int_1^x \frac{1}{s} \, ds \) Then we can numerically approximate the integral. Morally, we did not really “simplify” \( \ln x \) by writing down \( \int_1^x \frac{1}{s} \, ds \). We simply gave the integral a name. If we require numerical answers, it is possible we end up doing the calculation by approximating an integral anyway.

Another common function defined by an integral that cannot be evaluated symbolically is the erf function, defined as
\[
\text{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} \, ds.
\]
This function comes up often in applied mathematics. It is simply the antiderivative of \( e^{-x^2} \) that is zero at zero. The second form of the fundamental theorem tells us that we can write the function as an integral. If we wish to compute any particular value, we numerically approximate the integral.

**Change of variables**

A theorem often used in calculus to solve integrals is the change of variables theorem. Let us prove it now. Recall a function is continuously differentiable if it is differentiable and the derivative is continuous.
Let $g \colon [a, b] \to \mathbb{R}$ be a continuously differentiable function. If $(g([a, b]) \subset [c, d])$ and $f \colon [c, d] \to \mathbb{R}$ is continuous, then $\int_{a}^{b} f(g(x)) g'(x) \, dx = \int_{g(a)}^{g(b)} f(s) \, ds$.

As $g'$, $g''$, and $f$ are continuous, we know $f(g(x)) g'(x)$ is a continuous function on $[a, b]$, therefore it is Riemann integrable.

Define $F(y) := \int_{g(a)}^{y} f(s) \, ds$. By the second form of the fundamental theorem of calculus (using below) $F(y)$ is a differentiable function and $F'(y) = f(y)$. We apply the chain rule and write $g(x) = f(b)$, $g(x) = f(a)$, $g(x) = f(b)$, $g(x) = f(a)$. We note that $(F(b) - F(a)) = 0$ and we use the first form of the fundamental theorem to obtain $\int_{a}^{b} f(g(x)) g'(x) \, dx = F(b) - F(a) = \int_{a}^{b} \left( F \circ g \right)'(x) \, dx = \int_{a}^{b} f(g(x)) g'(x) \, dx$. %FIXME \qedhere

The change of variables theorem is often used to solve integrals by changing them to integrals that we know or that we can solve using the fundamental theorem of calculus.

From an exercise, we know that the derivative of $\sin(x)$ is $\cos(x)$. Therefore we solve $\int_0^{\sqrt{\pi}} x \cos(x^2) \, dx = \frac{1}{2} \int_0^{\pi} \cos(s) \, ds = \frac{\sin(\pi) - \sin(0)}{2} = 0$.

However, beware that we must satisfy the hypotheses of the theorem. The following example demonstrates why we should not just move symbols around mindlessly. We must be careful that those symbols really make sense.

Suppose we write down $\int_{-1}^{1} \frac{\ln |x|}{x} \, dx$. It may be tempting to take $g(x) := \ln |x|$, then take $g'(x) = \frac{1}{x}$ and try to write $\int_{0}^{0} s \, ds = \int_{0}^{0} \frac{\ln |x|}{x} \, dx$. This “solution” is incorrect, and it does not say that we can solve the given integral. First problem is that $\frac{\ln |x|}{x}$ is not continuous on $[-1, 1]$. Second, $\frac{\ln |x|}{x}$ is not even Riemann integrable on $[-1, 1]$ (it is unbounded). The integral we wrote down simply does not make sense. Finally, $\frac{\ln |x|}{x}$ is not continuous on $[-1, 1]$ either.

**Exercises**

Compute $\frac{d}{dx} \left( \int_{-x}^{x} e^{s^2} \, ds \right)$.

Compute $\frac{d}{dx} \left( \int_{0}^{x^2} \sin(s^2) \, ds \right)$.

Suppose $F \colon [a, b] \to \mathbb{R}$ is continuous and differentiable on $[a, b] \setminus S$, where $S$ is a finite set. Suppose there exists an $f \in \mathcal{R}[a, b]$ such that $f(x) = F'(x)$ for $x \in [a, b] \setminus S$. Show that $\int_{a}^{b} f = F(b) - F(a)$.

Let $F \colon [a, b] \to \mathbb{R}$ be a continuous function. Let $c \in [a, b])$ be arbitrary. Define $F(x) := \int_{c}^{x} f$. Prove that $F$ is differentiable and that $F(x) = f(x)$ for all $x \in [a, b])$.

Prove integration by parts. That is, suppose $F$ and $G$ are continuously differentiable functions on $[a, b]$. Then prove $\int \left[ \int_{a}^{b} F(x)G'(x) \, dx = F(b)G(b) - F(a)G(a) - \int_{a}^{b} F'(x)G(x) \, dx \right]$. 

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
Suppose \( (F) \) and \( (G) \) are continuously\(^{23} \) differentiable functions defined on \( ([a,b]) \) such that \( (F'(x) = G'(x)) \) for all \( (x \in [a,b]) \). Using the fundamental theorem of calculus, show that \( (F) \) and \( (G) \) differ by a constant. That is, show that there exists a \( (C \in \{\mathbb{R}\}) \) such that \( (F(x)-G(x) = C) \).

The next exercise shows how we can use the integral to “smooth out” a non-differentiable function.

[exercise:smoothingout] Let \( (f : [a,b] \to \{\mathbb{mathbb{R}}\}) \) be a continuous function. Let \( (\epsilon > 0) \) be a constant. For \( (x \in [a+\epsilon,b-\epsilon]) \), define \( (g(x) := \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} f) \). a) Show that \( (g) \) is differentiable and find the derivative. b) Let \( (f) \) be differentiable and fix \( (x \in (a,b)) \) (let \( (\epsilon) \) be small enough). What happens to \( (g'(x)) \) as \( (\epsilon) \) gets smaller? c) Find \( (g) \) for \( (f(x) := \left| x \right|) \), \( (\epsilon = 1) \) (you can assume \( ([a,b]) \) is large enough).

Suppose \( (f : [a,b] \to \{\mathbb{mathbb{R}}\}) \) is continuous and \( (\int_a^x f = \int_x^b f) \) for all \( (x \in [a,b]) \). Show that \( (f(x) = 0) \) for all \( (x \in [a,b]) \).

Suppose \( (f : [a,b] \to \{\mathbb{mathbb{R}}\}) \) is continuous and \( (\int_{-a}^a f = 0) \) for all rational \( (x) \) in \( ([a,b]) \). Show that \( (f(x) = 0) \) for all \( (x \in [a,b]) \).

A function \( (f) \) is an odd function if \( (f(x) = -f(-x)) \), and \( (f) \) is an even function if \( (f(x) = f(-x)) \). Let \( (a > 0) \). Assume \( (f) \) is continuous. Prove: a) If \( (f) \) is odd, then \( (\int_{-a}^a f = 0) \). b) If \( (f) \) is even, then \( (\int_{-a}^a f = 2 \int_0^a f) \).

a) Show that \( (f(x) := \sin(\nicefrac{1}{x})) \) is integrable on any interval (you can define \( (f(0)) \) to be anything). b) Compute \( (\int_{-1}^1 \sin(\nicefrac{1}{x}) \, dx) \). (Mind the discontinuity)

§6 a) Suppose \( (f : [a,b] \to \{\mathbb{mathbb{R}}\}) \) is increasing, by , \( (f) \) is Riemann integrable. Suppose \( (f) \) has a discontinuity at \( (c \in (a,b)) \), show that \( (F(x) := \int_a^x f) \) is not differentiable at \( (c) \). b) In , you have constructed an increasing function \( (f : [0,1] \to \{\mathbb{mathbb{R}}\}) \) that is discontinuous at every \( (x \in [0,1] \cap \{\mathbb{Q}\}) \). Use this \( (f) \) to construct a function \( (F(x)) \) that is continuous on \( ([0,1]) \), but not differentiable at every \( (x \in [0,1] \cap \{\mathbb{Q}\}) \).

The logarithm and the exponential

Note: 1 lecture (optional, requires the optional sections , )

We now have all that is required to finally properly define the exponential and the logarithm that you know from calculus so well. First recall that we have a good idea of what \( (x^n) \) means as long as \( (n) \) is a positive integer. Simply, \( (x^n := \underbrace{x \cdot x \cdot x \cdots x}_{\text{\(n\) times}}) \). It makes sense to define \( (x^0 := 1) \). For negative integers we define \( (x^{-n} := \nicefrac{1}{x^n}) \). If \( (x > 0) \), we mentioned before that \( (x^{1/n}) \) is defined as the unique positive \( (n) \)th root. Finally for any rational number \( (\nicefrac{1}{m}) \), we define \( (x^{\nicefrac{1}{m}} := \{\text{bigl}(x^{\{1/m\}}) \} \). However, what do we mean by \( (\sqrt[2]{2}^{\sqrt[2]{2}}) \)? Or \( (x^y) \) in general? In particular, what is \( (e^x) \) for all \( (x) \)? And how do we solve \( (y = e^x) \)? This section answers these questions and more.
The logarithm

It is convenient to start with the logarithm. Let us show that a unique function with the right properties exists, and only then will we call it the logarithm.

There exists a unique function \(L : (0, \infty) \to \mathbb{R}\) such that

i. \([\text{it:log:i}]\) \(L(1) = 0\).

ii. \([\text{it:log:ii}]\) \(L(x)\) is differentiable and \(L'(x) = \nicefrac{1}{x}\).

iii. \([\text{it:log:iii}]\) \(L(x)\) is strictly increasing, bijective, and \(\lim_{x \to 0} L(x) = -\infty\) and \(\lim_{x \to \infty} L(x) = \infty\).

iv. \([\text{it:log:iv}]\) \(L(xy) = L(x) + L(y)\) for all \(x, y \in (0, \infty)\).

v. \([\text{it:log:v}]\) If \(q\) is a rational number and \(x > 0\), then \(L(x^q) = q L(x)\).

To prove existence, let us define a candidate and show it satisfies all the properties. Define \(L(x) := \int_1^x \frac{1}{t} \, dt\).

Obviously \([\text{it:log:i}]\) holds. Property \([\text{it:log:ii}]\) holds via the fundamental theorem of calculus.

To prove property \([\text{it:log:iv}]\), we change variables \(u=yt\) to obtain \(L(x) = \int_1^x \frac{1}{t} \, dt = \int_y^{xy} \frac{1}{u} \, du = \int_1^{xy} \frac{1}{u} \, du - \int_1^{y} \frac{1}{u} \, du = L(xy) - L(y)\).

Property \([\text{it:log:ii}]\) together with the fact that \(L(x) = \nicefrac{1}{x}\) for \(x > 0\), implies that \(L(x)\) is strictly increasing and hence one-to-one. Let us show \(\text{im}(L)\) is onto. As \(\nicefrac{1}{t} \geq \frac{1}{2}\) when \(t \in [1,2]\), \(L(2) = \int_1^2 \frac{1}{t} \, dt \geq \frac{1}{2}\). By induction, \([\text{it:log:iv}]\) implies that for \(n \in \mathbb{N}\) \(L(2^n) = L(2) + L(2) + \cdots + L(2) = n L(2)\). Given any \(y > 0\), by the of the real numbers (notice \(L(2) > 0\)), there is an \(n \in \mathbb{N}\) such that \(L(2^n) > y\). By the there is an \(x_1 \in (1,2^n)\) such that \(L(x_1) = y\). We get \(L(0, \infty)\) is in the image of \(L\). As \(L(x)\) is increasing, \(L(x) > y\) for all \(x > 2^n\), so \(\lim_{x \to \infty} L(x) = \infty\). Next \(0 = L(\nicefrac{1}{x})\) and so \(\lim_{x \to 0} L(x) = -L(\nicefrac{1}{x})\). Using \(x=2^n\), we obtain as above that \(L\) achieves all negative numbers. And \(\lim_{x \to 0} L(x) = \lim_{x \to 0} -L(\nicefrac{1}{x}) = -\infty\). In the limits, note that only \(x > 0\) are in the domain of \(L\).

Let us now prove \([\text{it:log:v}]\). As above, \([\text{it:log:iv}]\) implies for \(n \in \mathbb{N}\) we have \(L(x^n) = n L(x)\). We already saw that \(L(x) = -L(\nicefrac{1}{x})\) so \(L(x^{-n}) = -L(x^n) = -n L(x)\). Then for \(m \in \mathbb{N}\) \(L(x^m) = L(\text{Bigl({(x^{1/m})}^m\Bigr)} = m L(x^{1/m})\). Putting everything together for \(n, m \in \mathbb{N}\) we have \(L(x^n) = n L(x^{1/m}) = (\nicefrac{n}{m}) L(x)\).

Finally for uniqueness, let us use properties \([\text{it:log:i}]\) and \([\text{it:log:ii}]\). Via the fundamental theorem of calculus \(L(x) = \int_1^x \frac{1}{t} \, dt\) is the unique function such that \(L(1) = 0\) and \(L'(x) = \frac{1}{x}\).
Having proved that there is a unique function with these properties we simply define the \textit{logarithm} or sometimes called the \textit{natural logarithm}: \(\ln(x) := L(x)\). Often mathematicians write \(\log(x)\) instead of \(\ln(x)\), which is more familiar to calculus students.

\textbf{The exponential}

Just as with the logarithm we define the exponential via a list of properties.

There exists a unique function \(E \colon \mathbb{R} \to (0, \infty)\) such that

1. \(E(0) = 1\).
2. \(E\) is differentiable and \(E'(x) = E(x)\).
3. \(E\) is strictly increasing, bijective, and \(\lim_{x \to -\infty} E(x) = 0\) and \(\lim_{x \to \infty} E(x) = \infty\).
4. \(E(x+y) = E(x)E(y)\) for all \(x, y \in \mathbb{R}\).
5. If \(q \in \mathbb{Q}\), then \(E(qx) = {E(x)}^q\).

Again, let us prove existence of such a function by defining a candidate, and prove that it satisfies all the properties. The \(L\) defined above is invertible. Let \(E\) be the inverse function of \(L\). Property \(E(0) = 1\) is immediate.

Property \(E\) follows via the inverse function theorem, in particular : \(E\) satisfies all the hypotheses of the lemma, and hence \(E'(x) = \frac{1}{L'(E(x))} = E(x)\).

Let us look at property \(E\). The function \(E\) is strictly increasing since \(E(x) > 0\) and \(E'(x) = E(x) > 0\). As \(E\) is the inverse of \(L\), it must also be bijective. To find the limits, we use that \(E\) is strictly increasing and onto \((0, \infty)\).

For every \(M > 0\), there is an \(x_0\) such that \(E(x_0) = M\) and \(E(x) \geq M\) for all \(x \geq x_0\). Similarly for every \(\epsilon > 0\), there is an \(x_0\) such that \(E(x_0) = \epsilon\) and \(E(x) < \epsilon\) for all \(x < x_0\). Therefore, \(\lim_{x \to -\infty} E(x) = 0\) and \(\lim_{x \to \infty} E(x) = \infty\).

To prove property \(E\), we use the corresponding property for the logarithm. Take \(x, y \in \mathbb{R}\). As \(L\) is bijective, find \(a\) and \(b\) such that \(x = L(a)\) and \(y = L(b)\). Then \(E(x+y) = E(L(ab)) = E\bigl(L(ab)\bigr) = ab = E(x)E(y)\).

Property \(E\) also follows from the corresponding property of \(L\). Given \(x \in \mathbb{R}\), let \(a\) be such that \(x = L(a)\) and \(E(qx) = E\bigl(qL(a)\bigr)\). Then \(E(qx) = a^q = E(x)^q\).

Finally, uniqueness follows from \(E\) and \(F\) be two functions satisfying \(E\) and \(F\). Therefore by , \(F(x)E(-x) = 0\) for all \(x \in \mathbb{R}\). Doing the computation with \(F = E\), we obtain \(E(x)E(-x) = 1\).
Then \[0 = 1 - F(x)E(-x) - E(x)E(-x) = \bigl(F(x) - E(x)\bigr) E(-x) .\] Since \(\langle E(x)E(-x) = 1\rangle\), then \(\langle E(-x) \not= 0\rangle\) for all \(\langle x\rangle\). So \(\langle F(x)E(-x) = 0\rangle\) for all \(\langle x\rangle\), and we are done.

Having proved \(\langle E\rangle\) is unique, we define the exponential function as \[
\exp(x) := E(x) .
\]

We can now make sense of exponentiation \(\langle x^y\rangle\) for arbitrary numbers when \(\langle x > 0\rangle\). First suppose \(\langle y \in \{\text{mathbb Q}\}\rangle\). Then \(\langle x^y = \exp\bigl(\ln(x^y)\bigr)\rangle\) \(\langle \exp\bigl(y\ln(x)\bigr)\rangle\). Therefore when \(\langle x > 0\rangle\) and \(\langle y\rangle\) is irrational let us define \(\langle x^y := \exp\bigl(y\ln(x)\bigr)\rangle\). As \(\langle \exp\rangle\) is continuous then \(\langle x^y\rangle\) is a continuous function of \(\langle y\rangle\). Therefore, we would obtain the same result had we taken a sequence of rational numbers \(\langle \{ y_n \} \rangle\) approaching \(\langle y\rangle\) and defined \(\langle x^y = \lim, x^{\langle y_n \}}\rangle\).

Define the number \(\langle e\rangle\) as \(\langle e := \exp(1) \rangle\). The number \(\langle e\rangle\) is sometimes called Euler’s number or the base of the natural logarithm. We notice \(\langle e^x = \exp(x \ln(e) \bigr)\rangle\) \(\langle \exp(x) \rangle\). We have justified the notation \(\langle e^x\rangle\) for \(\langle \exp(x)\rangle\).

Finally, let us extend properties of logarithm and exponential to irrational powers. The proof is immediate.

Let \(\langle x, y \in \{\text{mathbb R}\} \rangle\).

i. \(\langle \exp(xy) = (\exp(x)\bigr)^y\rangle\).

ii. If \(\langle x > 0\rangle\) then \(\langle \ln(x^y) = y \ln (x)\rangle\).

**Exercises**

Let \(\langle y\rangle\) be any real number and \(\langle b > 0\rangle\). Define \(\langle f: \{\text{mathbb R}\} \to \{\text{mathbb R}\} \rangle\) and \(\langle g: \{\text{mathbb R}\} \to \{\text{mathbb R}\} \rangle\) as, \(\langle f(x) := x^y \rangle\) and \(\langle g(x) := b^x\rangle\). Show that \(\langle f\rangle\) and \(\langle g\rangle\) are differentiable and find their derivative.

Let \(\langle b > 0\rangle\) be given.

a) Show that for every \(\langle y > 0\rangle\), there exists a unique number \(\langle x\rangle\) such that \(\langle y = b^x\rangle\). Define the logarithm base \(\langle b\rangle\), \(\langle \log_b: \{\text{mathbb R}\} \to \{\text{mathbb R}\} \rangle\), by \(\langle \log_b(y) := x\rangle\).

b) Show that \(\langle \log_b(x) = \frac{\ln(x)}{\ln(b)}\rangle\).

c) Prove that if \(\langle c > 0\rangle\), then \(\langle \log_b(x) = \frac{\log_c(x)}{\log_c(b)}\rangle\).

d) Prove \(\langle \log_b(xy) = \log_b(x) + \log_b(y)\rangle\), and \(\langle \log_b(x^y) = y \log_b(x)\rangle\).

§3] Use to study the remainder term and show that for all \(\langle x \in \{\text{mathbb R}\} \rangle\), \[\langle e^x = \sum_n=0^\infty \frac{x^n}{n!} \rangle\]. Hint: Do not differentiate the series term by term (unless you would prove that it works).

Use the geometric sum formula to show (for \(\langle t\not= -1\rangle\)) \[\langle 1+t+t^2+\cdots+(-1)^n t^n = \frac{1-t^n}{1+t}\rangle\].

ParseError: EOF expected (click for details)
Using this fact show that
\[
\ln (1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}x^n}{n}
\]
for all \(x \in (-1,1]\) (note that \(x=1\) is included). Finally, find the limit of the alternating harmonic series
\[
\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}
\]

\[
\sum_{k=2}^{n} \frac{1}{k} \leq \ln (n) \leq \sum_{k=1}^{n-1} \frac{1}{k}
\]
a) Prove that for \(n \in \mathbb{N}\) we have
\[
\sum_{k=1}^{n-1} \frac{1}{k} \leq \ln (n) \leq \sum_{k=1}^{n-1} \frac{1}{k}
\]
b) Prove that the limit
\[
\gamma := \lim_{n \to \infty} \left( \sum_{k=1}^{n} \frac{1}{k} - \ln (n) \right)
\]
exists. This constant is known as the Euler-Mascheroni constant. It is not known if this constant is rational or not, it is approximately \(\gamma \approx 0.5772\).

\[
\lim_{x \to \infty} \frac{\ln(x)}{x} = 0
\]
Show that \(e^x\) is convex, in other words, show that if \(a \leq x \leq b\) then
\[
e^x \leq a \frac{b-x}{b-a} + e^b \frac{x-a}{b-a}\]

\[
\lim_{n \to \infty} (1 + \frac{1}{n})^n = e
\]
\[
\lim_{n \to \infty} n^{1/n} = e
\]

**Improper integrals**

_Note: 2–3 lectures (optional section, can safely be skipped, requires the optional )_

Often it is necessary to integrate over the entire real line, or a infinite interval of the form \([a,\infty\)) or \((\infty,b]\). Also, we may wish to integrate functions defined on a finite interval \((a,b)\) but not bounded. Such functions are not Riemann integrable, but we may want to write down the integral anyway in the spirit of \(\int\). These integrals are called improper integrals, and are limits of integrals rather than integrals themselves.

Suppose \(f:\colon [a,b) \to \mathbb{R}\) is a function (not necessarily bounded) that is Riemann integrable on \([a,c]\) for all \(c < b\). We define \(\int_a^b f := \lim_{c \to b^-} \int_a^c f\). if the limit exists.
Suppose \( f: [a, \infty) \to \mathbb{R} \) is a function such that \( f \) is Riemann integrable on \([a,c]\) for all \( c < \infty \). We define \( \int_a^\infty f := \lim_{c \to \infty} \int_a^c f \) if the limit exists.

If the limit exists, we say the improper integral **converges**. If the limit does not exist, we say the improper integral **diverges**.

We similarly define improper integrals for the left hand endpoint, we leave this to the reader.

For a finite endpoint \( b \), using we see that if \( f \) is bounded, then we have defined nothing new. What is new is that we can apply this definition to unbounded functions. The following set of examples is so useful that we state it as a proposition.

The improper integral \( \int_1^\infty \frac{1}{x^p} \, dx \) converges to \( \frac{1}{p-1} \) if \( p > 1 \) and diverges if \( 0 < p \leq 1 \).

The improper integral \( \int_0^1 \frac{1}{x^p} \, dx \) converges to \( \frac{1}{1-p} \) if \( 0 < p < 1 \) and diverges if \( p \geq 1 \).

The proof follows by application of the fundamental theorem of calculus. Let us do the proof for \( p > 1 \) for the infinite right endpoint, and we leave the rest to the reader. Hint: You should handle \( p=1 \) separately.

Suppose \( p > 1 \). Then \( \int_1^b \frac{1}{x^p} \, dx = \int_1^b x^{-p} \, dx = \frac{b^{-p+1}}{-p+1} - \frac{1^{-p+1}}{-p+1} = - \frac{1}{(p-1)b^{p-1}} + \frac{1}{p-1} \). As \( p > 1 \), then \( p-1 > 0 \). Taking the limit as \( b \to \infty \) we obtain that \( \int_1^\infty \frac{1}{x^p} \, dx \) goes to 0, and the result follows.

We state the following proposition for just one type of improper integral, though the proof is straightforward and the same for other types of improper integrals.

The proof follows by application of the fundamental theorem of calculus. Let us do the proof for \( p > 1 \) for the infinite right endpoint, and we leave the rest to the reader. Hint: You should handle \( p=1 \) separately.

Suppose \( p > 1 \). Then \( \int_1^b \frac{1}{x^p} \, dx = \int_1^b x^{-p} \, dx = \frac{b^{-p+1}}{-p+1} - \frac{1^{-p+1}}{-p+1} = - \frac{1}{(p-1)b^{p-1}} + \frac{1}{p-1} \). As \( p > 1 \), then \( p-1 > 0 \). Taking the limit as \( b \to \infty \) we obtain that \( \int_1^\infty \frac{1}{x^p} \, dx \) goes to 0, and the result follows.

We state the following proposition for just one type of improper integral, though the proof is straightforward and the same for other types of improper integrals.

The proof follows by application of the fundamental theorem of calculus. Let us do the proof for \( p > 1 \) for the infinite right endpoint, and we leave the rest to the reader. Hint: You should handle \( p=1 \) separately.

Suppose \( p > 1 \). Then \( \int_1^b \frac{1}{x^p} \, dx = \int_1^b x^{-p} \, dx = \frac{b^{-p+1}}{-p+1} - \frac{1^{-p+1}}{-p+1} = - \frac{1}{(p-1)b^{p-1}} + \frac{1}{p-1} \). As \( p > 1 \), then \( p-1 > 0 \). Taking the limit as \( b \to \infty \) we obtain that \( \int_1^\infty \frac{1}{x^p} \, dx \) goes to 0, and the result follows.

We state the following proposition for just one type of improper integral, though the proof is straightforward and the same for other types of improper integrals.

The proof follows by application of the fundamental theorem of calculus. Let us do the proof for \( p > 1 \) for the infinite right endpoint, and we leave the rest to the reader. Hint: You should handle \( p=1 \) separately.

Nonnegative functions are easier to work with as the following proposition demonstrates. The exercises will show that this proposition holds only for nonnegative functions. Analogues of this proposition exist for all the other types of improper limits are left to the student.

Suppose \( f: [a, \infty) \to \mathbb{R} \) is nonnegative \( (f(x) \geq 0) \) for all \( x \) and such that \( f \) is Riemann integrable on \([a,b]\) for all \( b > a \).

i. \( \int_a^\infty f = \sup \{ \int_a^x f : x \geq a \} \)

ii. Suppose \( \{ x_n \} \) is a sequence with \( \lim_{n \to \infty} x_n = \infty \). Then \( \int_a^\infty f \) converges if and only if \( \lim_{n \to \infty} \int_a^{x_n} f \) exists, in which case \( \int_a^\infty f = \lim_{n \to \infty} \int_a^{x_n} f \).
In the first item we allow for the value of $\infty$ in the supremum indicating that the integral diverges to infinity.

Let us start with the first item. Notice that as \( f \) is nonnegative, then \( \int_a^\infty f \) is increasing as a function of \( x \). If the supremum is infinite, then for every \( M \in \{ \mathbb{R} \} \) we find \( \{ N \} \) such that \( \int_a^N f \geq M \). As \( \int_a^\infty f \) is increasing then \( \int_a^\infty f \geq M \) for all \( x \geq N \). So \( \int_a^\infty \) diverges to infinity.

Next suppose the supremum is finite, say \( (A = \sup \{ \int_a^x f : x \geq a \}) \). For every \( \epsilon > 0 \), we find an \( \{ N \} \) such that \( (A - \epsilon \leq \int_a^N f < A + \epsilon) \). As \( \int_a^\infty f \) is increasing, then \( \int_a^\infty f < \epsilon \) for all \( x \geq N \) and hence \( \int_a^\infty f \) converges to \( (A) \).

Let us look at the second item. If \( \int_a^\infty f \) converges then every sequence \( \{ \{ x_n \} \} \) going to infinity works. The trick is proving the other direction. Suppose \( \{ \{ x_n \} \} \) is such that \( \lim x_n = \infty \) and \( \lim \int_a^{x_n} f = A \) converges. Given \( \epsilon > 0 \), pick \( \{ M \} \) such that for all \( n \geq M \) we have \( \int_a^M f \leq A + \epsilon \). Because \( \int_a^\infty f \) is increasing as a function of \( x \), we have that for all \( x \geq M \) we have \( \int_a^x f \leq A + \epsilon \).

Let \( f \colon [a, \infty) \to \mathbb{R} \) and \( g \colon [a, \infty) \to \mathbb{R} \) be functions that are Riemann integrable on \([a,b]\) for all \( b > a \). Suppose that for all \( x \geq a \) we have \( \| f(x) \| \leq g(x) \). We need to show that the limit is unique. Suppose \( \{ x_n \} \) is a sequence converging to infinity such that \( \{ \int_a^{x_n} f \} \) converges to \( L_1 \), and \( \{ y_n \} \) is a sequence converging to infinity such that \( \{ \int_a^{y_n} f \} \) converges to \( L_2 \). Then there must be some \( \{ n \} \) such that \( \int_a^{x_n} f = L_1 \) and \( \int_a^{y_n} f = L_2 \). We can also suppose \( \{ x_n \} \geq B \) and \( \{ y_n \} \geq B \). Then \( \int_a^{x_n} f = L_1 \) and \( \int_a^{y_n} f = L_2 \).
As \( \epsilon > 0 \) was arbitrary, \( (L_1 = L_2) \), and hence \( (\int_a^\infty f) \) converges. Above we have shown that \( (\int_{-\infty} a \, g) \) for all \( (c > a) \). By taking the limit \( (c \to \infty) \), the first item is proved.

The second item is simply a contrapositive of the first item.

The improper integral \( \left[ \int_0^\infty \frac{\sin(x^2)(x+2)}{x^3+1} \, dx \right] \) converges.

Proof: First observe we simply need to show that the integral converges when going from 1 to infinity. For \( (x \geq 1) \) we obtain \( \left( \text{leftIntegral} \left[ \frac{\sin(x^2)(x+2)}{x^3+1} \right] \right) \) which is not true that you can split the improper integral in two; you cannot split the limit. However, if you succumb to the temptation to write \( \left( \text{leftIntegral} \left[ \frac{\sin(x^2)(x+2)}{x^3+1} \right] \right) \) and try to integrate each part separately, you will not succeed. It is not true that you can split the improper integral in two; you cannot split the limit. \( \left( \text{begin{split} \int_{2}^{\infty} \frac{2}{x^2 - 1} \, dx \end{split} \right) \) converges via the comparison test again using \( \left( \text{frac} \{1\} \{x^2\} \right) \). Suppose \( (f \colon [a,b] \to \mathbb{R}) \) is a function that is Riemann integrable on \( ([c,d]) \) for all \( (c,d) \) such that \( (a < c < d < b) \), then we define \( (\int_a^b f := \lim_{c \to a^+} \lim_{d \to b^-} \int_c^d f) \) if the limits exist.

We similarly define improper integrals with one infinite and one finite improper endpoint, we leave this to the reader.

One ought to always be careful about double limits. The definition given above says that we first take the limit as \( (d) \) goes to \( (b) \) or \( (\infty) \) for a fixed \( (c) \), and then we take the limit in \( (c) \). We will have to prove that in this case it does not matter which limit we compute first.

Let us see an example: \( \left( \int_{-\infty}^\infty \frac{1}{1+x^2} \, dx \right) \) converges of which we can apply the comparison test appropriately with \( \left( \text{nicefrac} \{1\} \{x\} \right) \). We get \( (\infty - \infty) \).

Now let us suppose that we need to take limits at both endpoints.

Suppose \( (f \colon \mathbb{R} \to \mathbb{R}) \) is a function such that \( (f) \) is Riemann integrable on all finite intervals \( ([a,b]) \). Then we define \( (\int_{-\infty}^\infty f := \lim_{c \to -\infty} \lim_{d \to \infty} \int_c^d f) \) if the limits exist.

If \( (f \colon \mathbb{R} \to \mathbb{R}) \) is a function integrable on every interval. Then \( (\lim_{a \to -\infty} \lim_{b \to \infty} \int_a^b f) \) in which case the two expressions are equal. If either of the expressions converges then the improper integral converges and \( (\lim_{a \to \infty} \int_{-\infty}^a f = \int_{-\infty}^\infty f) \).
Without loss of generality assume \(a < 0\) and \(b > 0\). Suppose the first expression converges. Then
\[
\lim_{a \to -\infty} \lim_{b \to \infty} \int_a^b f = \lim_{b \to \infty} \lim_{a \to -\infty} \left( \int_a^0 f + \int_0^b f \right) = \lim_{b \to \infty} \left( \lim_{a \to -\infty} \int_a^0 f \right) = \lim_{a \to -\infty} \int_a^0 f + \lim_{b \to \infty} \int_0^b f = \int_{-\infty}^0 f + \int_0^\infty f .
\]

Similar computation shows the other direction. Therefore, if either expression converges then the improper integral converges and
\[
\int_{-\infty}^\infty f = \lim_{a \to -\infty} \lim_{b \to \infty} \int_a^b f = \left( \lim_{a \to -\infty} \int_a^0 f \right) + \left( \lim_{b \to \infty} \int_0^b f \right) = \left( \lim_{a \to \infty} \int_{-a}^0 f \right) + \left( \lim_{a \to \infty} \int_0^a f \right) = \int_{-\infty}^\infty f .
\]

On the other hand, you must be careful to take the limits independently before you know convergence. Let \(f(x) = \frac{x}{\vert x \vert}\) for \(x \neq 0\) and \(f(0) = 0\). If \(a < 0\) and \(b > 0\), then
\[
\int_a^b f = \int_a^0 f + \int_0^b f = a + b .
\]
For any fixed \((a < 0)\) the limit as \((b \to \infty)\) is infinite, so even the first limit does not exist, and hence the improper integral \(\int_{-\infty}^\infty f\) does not converge. On the other hand if \((a > 0)\), then \(\int_{-a}^a f = 0\). Therefore, \(\int_{-\infty}^\infty f = 0\).

An example to keep in mind for improper integrals is the so-called sinc function. This function comes up quite often in both pure and applied mathematics. Define
\[
\operatorname{sinc}(x) = \begin{cases} \frac{\sin(x)}{x} & \text{if } x \neq 0 , \\ 0 & \text{if } x = 0 . \end{cases}
\]
It is not difficult to show that the sinc function is continuous at zero, but that is not important right now. What is important is that
\[
\int_{-\infty}^\infty \operatorname{sinc}(x) ~dx = \pi ,
\]
and
\[
\int_{-\infty}^\infty \left\vert \operatorname{sinc}(x) \right\vert ~dx = \infty .
\]
In particular, the fact that the integral converges must be done directly rather than using comparison test.

We will not prove the first statement exactly. Let us simply prove that the integral of the sinc function converges, but we will not worry about the exact limit. Because \(\left\vert \int_{2\pi}^\infty \frac{\sin(x)}{x} ~dx \right\vert \leq \frac{1}{\pi n(n+1)}\), it is enough to show that \(\int_{2\pi}^\infty \frac{\sin(x)}{x} ~dx\) converges. We also avoid \((x=0)\) this way to make our life simpler.

For any \(n \in \mathbb{N}\), we have that for \((x \in [2\pi n, 2\pi (n+1)]\)
\[
\frac{1}{2\pi n} \leq \frac{\sin(x)}{x} \leq \frac{1}{2\pi (n+1)} ,
\]
as \((n \to \infty)\), \(x \to 0\). On \(x \in [2\pi (n+1), 2\pi (n+2)]\)
\[
\frac{1}{2\pi (n+1)} \leq \frac{\sin(x)}{x} \leq \frac{1}{2\pi n} ,
\]
as \((n \to \infty)\), \(x \to 0\).

Via the fundamental theorem of calculus,
\[
\int_{2\pi}^M \frac{\sin(x)}{x} ~dx = \int_{2\pi}^{2\pi n} \frac{\sin(x)}{x} ~dx + \int_{2\pi n}^{2\pi (n+1)} \frac{\sin(x)}{x} ~dx .
\]
Similarly
\[
\int_{2\pi}^{2\pi (n+1)} \frac{\sin(x)}{x} ~dx = \int_{2\pi (n+1)}^{2\pi (n+2)} \frac{\sin(x)}{x} ~dx .
\]
Putting the two together we have
\[
\int_{2\pi}^{2\pi (n+1)} \frac{\sin(x)}{x} ~dx = \left\{ \begin{array}{l l} 0 & \text{if } n = 0 , \\ \int_{2\pi (n+1)}^{2\pi (n+2)} \frac{\sin(x)}{x} ~dx & \text{if } n \geq 0 . \end{array} \right.
\]

Via the fundamental theorem of calculus,
\[
\int_{2\pi}^M \frac{\sin(x)}{x} ~dx = \int_{2\pi}^{2\pi n} \frac{\sin(x)}{x} ~dx + \int_{2\pi n}^{2\pi (n+1)} \frac{\sin(x)}{x} ~dx .
\]
Similarly
\[
\int_{2\pi}^{2\pi (n+1)} \frac{\sin(x)}{x} ~dx = \int_{2\pi (n+1)}^{2\pi (n+2)} \frac{\sin(x)}{x} ~dx .
\]
Putting the two together we have
\[
\int_{2\pi}^M \frac{\sin(x)}{x} ~dx = \left\{ \begin{array}{l l} 0 & \text{if } n = 0 , \\ \int_{2\pi (n+1)}^{2\pi (n+2)} \frac{\sin(x)}{x} ~dx & \text{if } n \geq 0 . \end{array} \right.
\]
Let \((M > 2\pi)\) be arbitrary, and let \(k \in \mathbb{N}\) be the largest integer such that \((2k\pi < M)\). Then
\[
\int_{2\pi k}^M \frac{\sin(x)}{x} ~dx = \int_{2\pi k}^M \frac{\sin(x)}{x} ~dx = \int_{2\pi (k+1)}^M \frac{\sin(x)}{x} ~dx .
\]
For \((x \in [2k\pi, M)\) we have \(\left\{ \begin{array}{l l} 0 & \text{if } n = 0 , \\ \frac{1}{2\pi (n+1)} \leq \frac{\sin(x)}{x} \leq \frac{1}{2\pi n} & \text{if } n \geq 0 . \end{array} \right.\)
\[
\int_{2\pi k}^M \frac{\sin(x)}{x} ~dx = \int_{2\pi (k+1)}^M \frac{\sin(x)}{x} ~dx = \int_{2\pi (k+1)}^M \frac{\sin(x)}{x} ~dx .
\]

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
\[ \frac{M-2k\pi}{2k\pi} \leq \frac{1}{k}. \]

As \( k \) is the largest \( k \) such that \( (2k')/2 \pi \leq M \), this term goes to zero as \( M \) goes to infinity.

Next \[ 0 \leq \int_{2\pi}^{2k\pi} \frac{\sin(x)}{x} \, dx \leq \sum_{n=1}^{k-1} \frac{1}{\pi n(n+1)} \]
and this series converges as \( k \to \infty \).

Putting the two statements together we obtain \[ \int_{2\pi}^{\infty} \frac{\sin(x)}{x} \, dx \leq \sum_{n=1}^{\infty} \frac{1}{\pi n(n+1)} < \infty. \]

The double sided integral of sinc also exists as noted above. We leave the other statement—that the integral of the absolute value of the sinc function diverges—as an exercise.

**Integral test for series**

It can be very useful to apply the fundamental theorem of calculus in proving a series is summable and to estimate its sum.

Suppose \( f \colon [k, \infty) \to \mathbb{R} \) is a decreasing nonnegative function where \( k \in \mathbb{Z} \). Then \[ \sum_{n=k}^{\infty} f(n) \quad \text{converges if and only if} \quad \lim_{b \to \infty} \int_{k}^{b} f \quad \text{converges}. \]

By \( f \) is integrable on every interval \( ([k,b]) \) for all \( b > k \), so the statement of the theorem makes sense without additional hypotheses of integrability.

Let \( \epsilon > 0 \) be given. And suppose \( \sum_{k}^{\infty} f(n) \) converges. Let \( m \in \mathbb{N} \) be such that \( m > \ell \geq k \). Because \( f \) is decreasing we have \( \int_{k}^{\infty} f \leq \sum_{n=\ell}^{m-1} f(n) \). Therefore \[ \int_{k}^{\infty} f \leq \sum_{n=\ell}^{m-1} f(n) \leq \sum_{n=\ell}^{m} f(n) \leq m \int_{k}^{\infty} f. \]

By \( f \) is integrable on every interval \( ([k,b]) \) for all \( b > k \), so the statement of the theorem makes sense without additional hypotheses of integrability.

Conversely suppose \( \sum_{n=1}^{\infty} f(n) \) diverges. As \( f \) is positive then by \( \sum_{n=1}^{\infty} f(n) \) diverges to infinity. Using \( \sum_{n=1}^{\infty} f(n) \) with \( \ell = k \) we find \( \int_{k}^{\infty} f \leq \sum_{n=k}^{\infty} f(n) \). As the left hand side goes to infinity as \( m \to \infty \), so does the right hand side.

Let us show \( \sum_{n=1}^{\infty} \frac{1}{n^2} \) exists and let us estimate its sum to within 0.01. As this series is the \( p \)-series for \( p=2 \), we already know it converges, but we have only very roughly estimated its sum.

Using fundamental theorem of calculus we find that for \( k \in \mathbb{N} \) we have \( \int_{k}^{\infty} \frac{1}{x^2} \, dx = \lim_{x \to \infty} \frac{1}{x} \). In particular, the series must converge. But we also have that \( \int_{k}^{\infty} \frac{1}{x^2} \, dx = \lim_{x \to \infty} \frac{1}{x} \).
\[ \sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{1}{k^2} + \int_k^{\infty} \frac{1}{x^2} \, dx = \frac{1}{k^2} + \frac{1}{k} . \]

Adding the partial sum up to \((k-1)\) we get \[ \frac{1}{k} + \sum_{n=1}^{k-1} \frac{1}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} \leq \frac{1}{k^2} + \frac{1}{k} + \sum_{n=1}^{k-1} \frac{1}{n^2} . \] In other words, \( \frac{1}{k} + \sum_{n=1}^{k-1} \frac{1}{n^2} \) is an estimate for the sum to within \( \frac{1}{k^2} \). Therefore, if we wish to find the sum to within 0.01, we note \( \frac{1}{10} = 0.01 \). We obtain \( 1.6397 \ldots \approx \frac{1}{10} + \sum_{n=1}^{9} \frac{1}{n^2} \) \( \leq \sum_{n=1}^{\infty} \frac{1}{n^2} \leq \frac{1}{100} + \frac{1}{10} + \frac{1}{1} \approx 1.6497 \ldots . \) The actual sum is \( \frac{\pi^2}{6} \approx 1.6449 \ldots \).

### Exercises

Finish the proof of .

Find out for which \( a \in \mathbb{R} \) does \( \sum_{n=1}^{\infty} e^{an} \) converge. When the series converges, find an upper bound for the sum.

a) Estimate \( \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \) correct to within 0.01 using the integral test. b) Compute the limit of the series exactly and compare. Hint: the sum telescopes.

Prove \( \int_{-\infty}^{\infty} |\operatorname{sinc}(x)| \, dx = \infty \). Hint: again, it is enough to show this on just one side.

Can you interpret \( \int_{-1}^{1} \frac{1}{\sqrt{|x|}} \, dx \) as an improper integral? If so, compute its value.

Take \( f \colon [0,1] \to \mathbb{R} \), Riemann integrable on every interval \( [(a,b)] \), and such that there exist \( M \), \( a \), and \( T \), such that \( \int_{[a,b]} f(t) \, dt \leq M e^{\lambda t(a-b)} \) for all \( t \geq T \). Show that the Laplace transform of \( f \) exists. That is, for every \( s > a \) the following integral converges: \( F(s) := \int_{0}^{\infty} f(t) e^{-st} \, dt \).

Let \( f \colon [0,1] \to \mathbb{R} \) be a Riemann integrable function on every interval \( [(a,b)] \), and such that \( \int_{[a,b]} f(t) \, dt < \infty \). Show that the Fourier sine and cosine transforms exist. That is, for every \( \omega \geq 0 \) the following integrals converge: \( F^s(\omega) := \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) \sin(\omega t) \, dt \) and \( F^c(\omega) := \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) \cos(\omega t) \, dt \). Furthermore, show that \( F^s \) and \( F^c \) are bounded functions.

Suppose \( f \colon [0,\infty) \to \mathbb{R} \) is Riemann integrable on every interval \( [(0,b)] \). Show that \( \int_{0}^{\infty} f(t) \, dt \) converges if and only if for every \( \epsilon > 0 \) there exists an \( M \) such that if \( \epsilon < M \) then \( \int_{[a,b]} f(t) \, dt < \epsilon \).

Suppose \( f \colon [0,\infty) \to \mathbb{R} \) is nonnegative and decreasing. a) Show that if \( \int_{0}^{\infty} f(t) \, dt < \infty \), then \( \lim_{x \to \infty} f(x) = 0 \). b) Show that the converse does not hold.

Find an example of an unbounded continuous function \( f \colon [0,\infty) \to \mathbb{R} \) that is nonnegative and such that \( \int_{0}^{\infty} f(t) \, dt < \infty \). Note that this means that \( \lim_{x \to \infty} f(x) \) does not exist; compare previous exercise.
Hint: on each interval \([k, k+1]\), \(k \in \{\text{mathbb\{N\}}\}\), define a function whose integral over this interval is less than say \((2^{-k})\).

Find an example of a function \(f: [0, \infty) \to \{\text{mathbb\{R\}}\}\) integrable on all intervals such that \(\lim_{n \to \infty} \int_0^n f\) converges as a limit of a sequence, but such that \(\int_0^\infty f\) does not exist. Hint: for all \(n \in \{\text{mathbb\{N\}}\}\), divide \([n, n+1]\) into two halves. In one half make the function negative, on the other make the function positive.

Show that if \(f: [1, \infty) \to \{\text{mathbb\{R\}}\}\) is such that \(g(x) := x^2 f(x)\) is a bounded function, then \(\int_1^\infty f\) converges.

It is sometimes desirable to assign a value to integrals that normally cannot be interpreted as even improper integrals, e.g. \(\int_{-1}^1 \frac{1}{x}\,dx\). Suppose \(f: [a, b] \to \{\text{mathbb\{R\}}\}\) is a function and \((a < c < b)\), where \(f\) is Riemann integrable on all intervals \([a, c-\epsilon]\) and \([c+\epsilon, b]\) for all \(\epsilon > 0\). Define the Cauchy principal value of \(\int_a^b f\) as \(p.v.\int_a^b f := \lim_{\epsilon \to 0^+} \left( \int_a^{c-\epsilon} f + \int_{c+\epsilon}^b f \right)\) if the limit exists.

a) Compute \(p.v.\int_{-1}^1 \frac{1}{x}\,dx\).

b) Compute \(\lim_{\epsilon \to 0^+} \left( \int_{-1}^1 \frac{1}{x}\,dx + \int_{1}^-\epsilon \frac{1}{x}\,dx \right)\) and show it is not equal to the principal value.

c) Show that if \(f\) is integrable on \([a, b]\), then \(\int_{-1}^1 f = \int_a^b f\).

d) Find an example of an \(f\) with a singularity at \(c\) as above such that \(\int_{1}^-\epsilon f\) exists, but the improper integrals \(\int_{-1}^c f\) and \(\int_{c}^b f\) diverge.

e) Suppose \(f: [-1, 1] \to \{\text{mathbb\{R\}}\}\) is continuous. Show that \(\int_{-1}^1 f\) exists.

Let \(f: [a, b] \to \{\text{mathbb\{R\}}\}\) and \(g: [a, b] \to \{\text{mathbb\{R\}}\}\) be continuous functions, where \(g(x) = 0\) for all \(x \in (a, b]\) for some interval \([a, b]\).

a) Show that the convolution \(\int \lim_{\to \infty} f\to \infty k(t) g(x-t)\,dt\) is well-defined for all \(x \in \{\text{mathbb\{R\}}\}\).

b) Suppose \(\int_{-1}^1 f\) and \(\int f\) are both integrable. Prove that \(\int_{-1}^1 f\) and \(\int f\) are both integrable.

\[\text{Sequences of Functions}\]

\[\text{Pointwise and uniform convergence}\]

\[\text{Note: 1-1.5 lecture}\]

Up till now when we talked about sequences we always talked about sequences of numbers. However, a very useful concept in analysis is to use a sequence of functions. For example, a solution to some differential equation might be found by finding only approximate solutions. Then the real solution is some sort of limit of those approximate solutions.
When talking about sequences of functions, the tricky part is that there are multiple notions of a limit. Let us describe two common notions of a limit of a sequence of functions.

**Pointwise convergence**

For every \( n \in \{\mathbb{N}\} \) let \( \{f_n\colon S \to \{\mathbb{R}\}\} \) be a function. We say the sequence \( \{f_n\}_{n=1}^{\infty} \) converges pointwise to \( \{f\colon S \to \{\mathbb{R}\}\} \), if for every \( x \in S \) we have \( f(x) = \lim_{n\to\infty} f_n(x) \).

It is common to say that \( \{f_n\colon S \to \{\mathbb{R}\}\} \) converges to \( f \) on \( T \subset \{\mathbb{R}\} \) for some \( f \colon T \to \{\mathbb{R}\} \). In that case we, of course, mean \( f(x) = \lim_{n\to\infty} f_n(x) \) for every \( x \in T \). We simply mean that the restrictions of \( f_n \) to \( T \) converge pointwise to \( f \).

The sequence of functions defined by \( f_n(x) := x^{2n} \) converges to \( f \colon [-1,1] \to \{\mathbb{R}\} \) on \( [-1,1] \), where

\[
\begin{cases}
1 & \text{if } x = -1 \text{ or } x = 1, \\
0 & \text{otherwise.}
\end{cases}
\]

To see this is so, first take \( x \in (-1,1) \). Then \( 0 < x^2 < 1 \). We have seen before that \( \lvert x^{2n} - 0 \rvert = (x^2)^n \to 0 \) as \( n \to \infty \). Therefore \( \lim_{n\to\infty} f(x) = 0 \).

When \( x = 1 \) or \( x = -1 \), then \( x^{2n} = 1 \) for all \( n \) and hence \( \lim f_n(x) = 1 \). We also note that \( \{f_n(x)\} \) does not converge for all other \( x \).

Often, functions are given as a series. In this case, we use the notion of pointwise convergence to find the values of the function.

We write \( \sum_{k=0}^{\infty} x^k \) to denote the limit of the functions \( f_n(x) := \sum_{k=0}^{n} x^k \). When studying series, we have seen that on \( x \in (-1,1) \) the \( f_n(x) \) converge pointwise to \( \sum_{k=0}^{\infty} x^k \).

The subtle point here is that while \( \sum_{k=0}^{\infty} x^k \) is defined for all \( x \not = 1 \), and \( f_n(x) \) are defined for all \( x \) (even at \( x = 1 \)), convergence only happens on \( (-1,1) \).

Therefore, when we write \( f(x) := \sum_{k=0}^{\infty} x^k \) we mean that \( f(x) \) is defined on \( (-1,1) \) and is the pointwise limit of the partial sums.

Let \( f_n(x) := \sin(xn) \). Then \( f_n(x) \) does not converge pointwise to any function on any interval. It may converge at certain points, such as when \( x = 0 \) or \( x = \pi n \). It is left as an exercise that in any interval \( (a,b) \), there exists an \( x \) such that \( f_n(x) \) does not have a limit as \( n \to \infty \).

Before we move to uniform convergence, let us reformulate pointwise convergence in a different way. We leave the proof to the reader, it is a simple application of the definition of convergence of a sequence of real numbers.

**ptwsconv:prop** Let \( \{f_n\colon S \to \{\mathbb{R}\}\} \) and \( \{f\colon S \to \{\mathbb{R}\}\} \) be functions. Then \( \{f_n\} \) converges pointwise to \( f \) if and only if for every \( x \in S \), and every \( \epsilon > 0 \), there exists an \( N \) such that \( \lvert f_n(x) - f(x) \rvert < \epsilon \) for all \( n \geq N \).
The key point here is that \( N \) can depend on \( x \), not just on \( \epsilon \). That is, for each \( x \) we can pick a different \( N \). If we can pick one \( N \) for all \( x \), we have what is called uniform convergence.

**Uniform convergence**

Let \( \{ f_n \colon S \to \mathbb{R} \} \) be functions. We say the sequence \( \{ f_n \} \) **converges uniformly** to \( f \colon S \to \mathbb{R} \), if for every \( \epsilon > 0 \) there exists an \( N \in \mathbb{N} \) such that for all \( n \geq N \) we have
\[
\left| f_n(x) - f(x) \right| < \epsilon \text{ for all } x \in S.
\]

Note that \( N \) now cannot depend on \( x \). Given \( \epsilon > 0 \) we must find an \( N \) that works for all \( x \in S \). Because of this, we see that uniform convergence implies pointwise convergence.

Let \( \{ f_n \} \) be a sequence of functions \( \{ f_n \colon S \to \mathbb{R} \} \). If \( \{ f_n \} \) converges uniformly to \( f \colon S \to \mathbb{R} \), then \( \{ f_n \} \) converges pointwise to \( f \).

The converse does not hold.

The functions \( f_n(x) := x^{2n} \) do not converge uniformly on \((-1,1)\), even though they converge pointwise. To see this, suppose for contradiction that the convergence is uniform. For \( \epsilon := \frac{1}{2} \), there would have to exist an \( N \) such that \( x^{2N} = \left| x^{2N} - 0 \right| < \frac{1}{2} \) for all \( x \in (-1,1) \) (as \( f_n(x) \) converges to 0 on \((-1,1)\)). But that means that for any sequence \( \{ x_k \} \) in \((-1,1)\) such that \( \lim x_k = 1 \) we have \( x_k^{2N} < \frac{1}{2} \) for all \( k \). On the other hand \( x^{2N} \) is a continuous function of \( x \) (it is a polynomial), therefore we obtain a contradiction \( 1 = 1^{2N} = \lim_{k \to \infty} x_k^{2N} \leq \frac{1}{2} \).

However, if we restrict our domain to \((-a,a)\) where \( 0 < a < 1 \), then \( \{ f_n \} \) converges uniformly to 0 on \((-a,a)\).

**Convergence in uniform norm**

For bounded functions there is another more abstract way to think of uniform convergence. To every bounded function we assign a certain nonnegative number (called the uniform norm). This number measures the “distance” of the function from 0. We can then “measure” how far two functions are from each other. We simply translate a statement about uniform convergence into a statement about a certain sequence of real numbers converging to zero.

[def:unifnorm] Let \( \{ f \colon S \to \mathbb{R} \} \) be a bounded function. Define \( \| f \|_u := \sup \left\{ \left| f(x) \right| : x \in S \right\} \). \( \| f \|_u \) is called the **uniform norm**.

To use this notation and this concept, the domain \( S \) must be fixed. Some authors use the notation \( \| f \|_{S} \) to emphasize the dependence on \( S \).
A sequence of bounded functions \( \{f_n\} \colon S \to \mathbb{R} \) converges uniformly to \( f \colon S \to \mathbb{R} \), if and only if
\[
\lim_{n \to \infty} \|f_n - f\|_u = 0.
\]

First suppose \( \lim_{n \to \infty} \|f_n - f\|_u = 0 \). Let \( \epsilon > 0 \) be given. Then there exists an \( N \) such that for all \( n \geq N \) we have \( \|f_n - f\|_u < \epsilon \). Hence \( \|f_n - f\|_u \) is the supremum of \( \|f_n(x) - f(x)\| \) for all \( x \in S \), we see that for all \( x \) we have \( \|f_n(x) - f(x)\| < \epsilon \).

On the other hand, suppose \( \{f_n\} \) converges uniformly to \( f \). Let \( \epsilon > 0 \) be given. Then find \( N \) such that \( \|f_n(x) - f(x)\| < \epsilon \) for all \( x \in S \). Taking the supremum we see that \( \|f_n - f\|_u < \epsilon \). Hence \( \lim_{n \to \infty} \|f_n - f\|_u = 0 \).

Sometimes it is said that \( \{f_n\} \) converges to \( f \) in uniform norm instead of converges uniformly. The proposition says that the two notions are the same thing.

Let \( f_n \colon [0,1] \to \mathbb{R} \) be defined by \( f_n(x) := \frac{nx+ \sin(nx^2)}{n} \). Then we claim \( \{f_n\} \) converges uniformly to \( f(x) := x \). Let us compute:
\[
\lim_{n \to \infty} \|f_n - f\|_u = \sup \{ \left| \frac{nx+ \sin(nx^2)}{n} - x \right| : x \in [0,1] \} = \sup \{ \frac{1}{n} : x \in [0,1] \} = \frac{1}{n}.
\]

Using uniform norm, we define Cauchy sequences in a similar way as we define Cauchy sequences of real numbers.

Let \( f_n \colon S \to \mathbb{R} \) be bounded functions. The sequence is **Cauchy in the uniform norm or uniformly Cauchy** if for every \( \epsilon > 0 \), there exists an \( N \) such that for \( m, k \geq N \) we have \( \|f_m - f_k\|_u < \epsilon \).

[prop:uniformcauchy] Let \( f_n \colon S \to \mathbb{R} \) be bounded functions. Then \( \{f_n\} \) is Cauchy in the uniform norm if and only if there exists an \( f \colon S \to \mathbb{R} \) and \( \{f_n\} \) converges uniformly to \( f \).

Let us first suppose \( \{f_n\} \) is Cauchy in the uniform norm. Let us define \( f \). Fix \( x \), then the sequence \( \{f_n(x)\} \) is Cauchy because \( \|f_m(x) - f_k(x)\|_u \leq \|f_m - f_k\|_u \). Thus \( \{f_n(x)\} \) converges to some real number. Define \( f(x) := \lim_{n \to \infty} f_n(x) \). The sequence \( \{f_n\} \) converges pointwise to \( f \). To show that the convergence is uniform, let \( \epsilon > 0 \) be given. Find an \( N \) such that for \( m, k \geq N \) we have \( \|f_m(x) - f_k(x)\|_u < \epsilon \). In other words for all \( x \) we have \( \|f_m(x) - f_k(x)\|_u < \epsilon \). We take the limit as \( k \) goes to infinity. Then \( \|f_m(x) - f_k(x)\|_u \) goes to zero. Consequently for all \( x \) we get \( \|f_m(x) - f_k(x)\|_u < \epsilon \). And hence \( \{f_n\} \) converges uniformly.

For the other direction, suppose \( \{f_n\} \) converges uniformly to \( f \). Given \( \epsilon > 0 \), find \( N \) such that for all \( n \geq N \) we have \( \|f_n(x) - f(x)\|_u < \epsilon \) for all \( x \in S \). Therefore for all \( m, k \geq N \) we have \( \|f_m(x) - f_k(x)\|_u \) and hence \( \{f_n\} \) converges uniformly.
\[ |f_m(x) - f(x)| + |f(x) - f_k(x)| < \frac{\epsilon}{4} + \frac{\epsilon}{4}. \]

Take supremum over all \(x\) to obtain \[ |f_m(x) - f_k(x)| < \frac{\epsilon}{2} < \epsilon. \] \(\text{qedhere}\)

**Exercises**

Let \((f)\) and \((g)\) be bounded functions on \([a, b])\). Prove \[|f + g|_u \leq |f|_u + |g|_u. \]

a) Find the pointwise limit \(\lim_{n \to \infty} \left(\frac{e^{x/n}}{n}\right)\) for \(x \in \mathbb{R}\).

b) Is the limit uniform on \(\mathbb{R}\)?

c) Is the limit uniform on \([0, 1]\)?

Suppose \(f_n: S \to \mathbb{R}\) are functions that converge uniformly to \(f: S \to \mathbb{R}\). Suppose \(A \subset S\). Show that the sequence of restrictions \(\{f_n|_A\}\) converges uniformly to \(\{f|_A\}\).

Suppose \(\{f_n\}\) and \(\{g_n\}\) defined on some set \(A\) converge to \(f\) and \(g\) respectively pointwise. Show that \(\{f_n + g_n\}\) converges pointwise to \(f + g\).

Suppose \(\{f_n\}\) and \(\{g_n\}\) defined on some set \(A\) converge to \(f\) and \(g\) respectively uniformly on \(A\). Show that \(\{f_n + g_n\}\) converges uniformly to \(f + g\).

Find an example of a sequence of functions \(\{f_n\}\) and \(\{g_n\}\) that converge uniformly to some \(f\) and \(g\) on some set \(A\), but such that \(\{f_n g_n\}\) (the multiple) does not converge uniformly to \(fg\) on \(A\). Hint: Let \(A := \mathbb{R}\), let \(f(x) := g(x) := x\). You can even pick \(f_n = g_n\).

Suppose there exists a sequence of functions \(\{g_n\}\) uniformly converging to \(0\) on \(\mathbb{R}\). Now suppose we have a sequence of functions \(\{f_n\}\) and a function \(f\) on \(\mathbb{R}\) such that \[|f_n(x) - f(x)| \leq g_n(x)\] for all \(x \in \mathbb{R}\). Show that \(\{f_n\}\) converges uniformly to \(f\) on \(\mathbb{R}\).

Let \(\{f_n\}\), \(\{g_n\}\), and \(\{h_n\}\) be sequences of functions on \([a, b]\). Suppose \(\{f_n\}\) and \(\{h_n\}\) converge uniformly to some function \(f: [a, b] \to \mathbb{R}\) and suppose \(\{g_n\}\) converges uniformly to \(h(x)\) for all \(x \in [a, b]\). Show that \(\{f_n g_n\}\) converges uniformly to \(f^2\).

Let \(\{f_n\}\) be a sequence of increasing functions (that is, \(f_n(x) \geq f_n(y)\)) whenever \((x \geq y)\). Suppose \(f_n(0) = 0\) and \(\lim_{n \to \infty} f_n(1) = 0\). Show that \(\{f_n\}\) converges uniformly to \(0\).

Let \(\{f_n\}\) be a sequence of functions defined on \([0, 1]\). Suppose there exists a sequence of distinct numbers \(x_n \in [0, 1]\) such that \(f_n(x_n) = 1\).\) Prove or disprove the following statements:

a) True or false: There exists \(\{f_n\}\) as above that converges to \(0\) pointwise.

b) True or false: There exists \(\{f_n\}\) as above that converges to \(0\) uniformly on \([0, 1]\).
Fix a continuous \( h \colon [a,b] \to \mathbb{R} \). Let \( f(x) := h(x) \) for \( x \in [a,b] \), \( f(x) := h(a) \) for \( x < a \) and \( f(x) := h(b) \) for all \( x > b \). First show that \((f \colon [a,b] \to \mathbb{R})\) is continuous. Now let \((f_n)\) be the function \((g)\) from with \( \epsilon = \frac{1}{n} \), defined on the interval \([a,b]\). Show that \((f_n)\) converges uniformly to \( h \) on \([a,b]\).

Interchange of limits

Note: 1–1.5 lectures

Large parts of modern analysis deal mainly with the question of the interchange of two limiting operations. When we have a chain of two limits, we cannot always just swap the limits. For example,

\[
0 = \lim_{n \to \infty} \left( \lim_{k \to \infty} \frac{n}{k} \right) \not= \lim_{k \to \infty} \left( \lim_{n \to \infty} \frac{n}{k} \right) = 1.
\]

When talking about sequences of functions, interchange of limits comes up quite often. We treat two cases. First we look at continuity of the limit, and second we look at the integral of the limit.

Continuity of the limit

If we have a sequence \( (f_n) \) of continuous functions, is the limit continuous? Suppose \( f \) is the (pointwise) limit of \( (f_n) \). If \( \lim x_k = x \) we are interested in the following interchange of limits. The equality we have to prove (it is not always true) is marked with a question mark. In fact the limits to the left of the question mark might not even exist.

\[
\lim_{k \to \infty} f(x_k) = \lim_{k \to \infty} \left( \lim_{n \to \infty} f_n(x_k) \right) \overset{?}{=} \lim_{n \to \infty} \left( \lim_{k \to \infty} f_n(x_k) \right) = \lim_{n \to \infty} f_n(x) = f(x).
\]

In particular, we wish to find conditions on the sequence \( (f_n) \) so that the above equation holds. It turns out that if we only require pointwise convergence, then the limit of a sequence of functions need not be continuous, and the above equation need not hold.

Let \( (f_n) \colon [0,1] \to \mathbb{R} \) be defined as \( f_n(x) := \begin{cases} 1-nx & \text{if } x < \frac{1}{n}, \\ 0 & \text{if } x \geq \frac{1}{n}. \end{cases} \) Each function \( f_n \) is continuous. Fix an \( x \in (0,1] \). If \( n \geq \frac{1}{x} \), then \( f_n(x) = 0 \). Therefore for \( n \geq \frac{1}{x} \) we have \( f_n(x) = 0 \). On the other hand if \( x=0 \), then \( \lim_{n \to \infty} f_n(0) = \lim_{n \to \infty} 1 = 1 \). Thus the pointwise limit of \( (f_n) \) is the function \( f \colon [0,1] \to \{0,1\} \) defined by \( f(x) := \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{if } x > 0. \end{cases} \) The function \( f \) is not continuous at 0.

If we, however, require the convergence to be uniform, the limits can be interchanged.

Let \( f_n \) be a sequence of continuous functions \( f_n \colon S \to \mathbb{R} \) converging uniformly to \( f \colon S \to \mathbb{R} \). Then \( f \) is continuous.

Let \( x \in S \) be fixed. Let \( f_n \) be a sequence in \( f(S) \) converging to \( f(x) \).
Let $\epsilon > 0$ be given. As $(f_k)$ converges uniformly to $(f)$, we find a $k \in \mathbb{N}$ such that $|f_k(x) - f(x)| < \frac{\epsilon}{3}$ for all $x \in S$. As $(f_k)$ is continuous at $(x)$, we find an $N \in \mathbb{N}$ such that for all $m \geq N$ we have $|f_k(x_m) - f(x_m)| < \frac{\epsilon}{3}$. Thus for $m \geq N$ we have $(f_k(x_m) - f(x_m))$ converges to $(f(x_m) - f(x))$ and hence $(f)$ is continuous at $(x)$. As $(x)$ was arbitrary, $(f)$ is continuous everywhere.

**Integral of the limit**

Again, if we simply require pointwise convergence, then the integral of a limit of a sequence of functions need not be equal to the limit of the integrals.

Let $f_n \colon [0,1] \to \mathbb{R}$ be defined as:

$$f_n(x) := \begin{cases} 0 & \text{if } x = 0, \\ n-n^2x & \text{if } 0 < x < \frac{1}{n}, \\ 0 & \text{if } x \geq \frac{1}{n}. 
\end{cases}$$

Each $f_n$ is Riemann integrable (it is continuous on $(0,1]$ and bounded), and it is easy to see $\int_0^1 f_n = \int_0^{1/n} (n-n^2x)~dx = \frac{1}{2}$. Let us compute the pointwise limit of $(f_n)$. Fix an $x \in (0,1)$. For $n \geq \frac{1}{x}$ we have $x \geq \frac{1}{n}$ and so $f_n(x) = 0$. Therefore $\lim_{n \to \infty} f_n(x) = 0$. We also have $(f_n(0) = 0)$ for all $n$. Therefore the pointwise limit of $(f_n)$ is the zero function. Thus $\lim_{n \to \infty} f_n(x) = 0$. We also have $f_n(0) = 0$ for all $n$. Therefore the pointwise limit of $(f_n)$ is the zero function. Thus $\frac{1}{2} = \lim_{n \to \infty} \int_0^1 f_n(x)~dx \neq \int_0^1 \lim_{n \to \infty} f_n(x)~dx = 0$. But if we again require the convergence to be uniform, the limits can be interchanged.

**Integral interchange theorem** Let $(f_n)$ be a sequence of Riemann integrable functions $f_n \colon [a,b] \to \mathbb{R}$ converging uniformly to $(f)$, then $(f)$ is Riemann integrable and $\int_a^b f = \lim_{n \to \infty} \int_a^b f_n$. Let $\epsilon > 0$ be given. As $(f_n)$ converges to $(f)$, we find an $M \in \mathbb{N}$ such that for all $n \geq M$ we have $|f_n(x) - f(x)| < \frac{\epsilon}{2(b-a)}$ for all $x \in [a,b)$. In particular, by reverse triangle inequality $|f_n(x)| < \frac{\epsilon}{2(b-a)} + |f_n(x)|$ for all $x$, hence $(f)$ is bounded as $(f_n)$ is bounded. Note that $(f_n)$ is integrable and compute $\int_a^b f_n = \int_a^b f_n(x)~dx = \int_a^b f_n(x)~dx + \int_a^b f_n(x)~dx + \int_a^b f_n(x)~dx + \int_a^b f_n(x)~dx + \int_a^b f_n(x)~dx + \int_a^b f_n(x)~dx + \int_a^b f_n(x)~dx$. The first inequality is (it follows as supremum of a sum is less than or equal to the sum of suprema and similarly for infima, see ). The second inequality follows.
from and the fact that for all \(x \in [a,b]\) we have
\[
\frac{-\epsilon}{2(b-a)} < f(x) - f_n(x) < \frac{\epsilon}{2(b-a)}.
\]
As \(\epsilon > 0\) was arbitrary, \(f\) is Riemann integrable.

Finally we compute \(\int_a^b f\). We apply in the calculation. Again, for \(\forall n \geq M\) (where \(M\) is the same as above) we have
\[
\left| \int_a^b f - \int_a^b f_n \right| = \left| \int_a^b (f(x) - f_n(x)) \, dx \right| \\
\leq \frac{\epsilon}{2(b-a)} (b-a) = \frac{\epsilon}{2} < \epsilon.
\]
Therefore \(\{\int_a^b f_n\}\) converges to \(\int_a^b f\).

Suppose we wish to compute
\[
\lim_{n \to \infty} \int_0^1 \frac{nx + \sin(nx^2)}{n} \, dx.
\]
It is impossible to compute the integrals for any particular \(n\) using calculus as \(\sin(nx^2)\) has no closed-form antiderivative. However, we can compute the limit. We have shown before that \(\frac{nx + \sin(nx^2)}{n}\) converges uniformly on \([0,1]\) to \(x\). By , the limit exists and
\[
\lim_{n \to \infty} \int_0^1 \frac{nx + \sin(nx^2)}{n} \, dx = \int_0^1 x \, dx = \frac{1}{2}.
\]
If convergence is only pointwise, the limit need not even be Riemann integrable. On \([0,1]\) define \(f_n(x) := \begin{cases} 1 & \text{if } x = \frac{p}{q} \text{ in lowest terms and } q \leq n, \\ 0 & \text{otherwise.} \end{cases}\) The function \(f_n\) differs from the zero function at finitely many points; there are only finitely many fractions in \([0,1]\) with denominator less than or equal to \(n\). So \(f_n\) is integrable and \(\int_0^1 f_n = \int_0^1 0 = 0\). It is an easy exercise to show that \(f_n\) converges pointwise to the Dirichlet function \(f(x) := \begin{cases} 1 & \text{ if } x \in \mathbb{Q}, \\ 0 & \text{otherwise.} \end{cases}\) which is not Riemann integrable.

In fact, if the convergence is only pointwise, the limit of bounded functions is not necessarily bounded. Define \(f_n(x) := \begin{cases} 0 & \text{ if } x < \frac{1}{n}, \\ \frac{1}{x} & \text{else.} \end{cases}\) For every \(n\) we get that \(f_n(x)\) converges pointwise to \(f(x) := \begin{cases} 0 & \text{ if } x = 0, \\ \frac{1}{x} & \text{else.} \end{cases}\) which is unbounded.

Let us remark that while uniform convergence is enough to swap limits with integrals, it is not, however, enough to swap limits with derivatives, unless you also have uniform convergence of the derivatives themselves. See the exercises below.

## Exercises

While uniform convergence preserves continuity, it does not preserve differentiability. Find an explicit example of a sequence of differentiable functions on \([-1,1]\) that converge uniformly to a function \(f\) such that \(f\) is not differentiable. Hint: Consider \(\left\lfloor 1 + n^2 \right\rfloor \frac{x}{1 + n^2}, \) show that these functions are differentiable, converge uniformly, and then show that the limit is not differentiable.

Let \(f_n(x) := \frac{x^n}{n}\). Show that \(f_n\) converges uniformly to a differentiable function \(f\) on \([0,1]\) (find \(f\)). However, show that \(f(1) \not= \lim_{n \to \infty} f_n(1)\).

Note: The previous two exercises show that we cannot simply swap limits with derivatives, even if the convergence is uniform. See also below.
Let \( f \colon [0,1] \to \mathbb{R} \) be a Riemann integrable (hence bounded) function. Find \( \lim_{n \to \infty} \int_0^1 \frac{f(x)}{n} \, dx \).

Show \( \lim_{n \to \infty} \int_1^2 e^{-nx^2} \, dx = 0 \). Feel free to use what you know about the exponential function from calculus.

Find an example of a sequence of continuous functions on \((0,1))\) that converges pointwise to a continuous function on \((0,1))\), but the convergence is not uniform.

Note: In the previous exercise, \((0,1))\) was picked for simplicity. For a more challenging exercise, replace \((0,1))\) with \([0,1] \).

True/False; prove or find a counterexample to the following statement: If \( \{ f_n \} \) is a sequence of everywhere discontinuous functions on \([0,1])\) that converge uniformly to a function \( f \), then \( f \) is everywhere discontinuous.

For the following two exercises let us define for a Riemann integrable function \( f \colon [0,1] \to \mathbb{R} \) the following number \( \Vert f \Vert_{L^1} := \int_0^1 \lvert f(x) \rvert \, dx \).

It is true that \( \{ f_n \} \) converges uniformly to zero implies \( \lim_{n \to \infty} \int_0^1 f_n \, dx = 0 \). Find a sequence of Riemann integrable functions \( \{ f_n \} \) on \([0,1])\) that converges pointwise to \( 0 \), but \( \lim_{n \to \infty} \int_0^1 f_n \, dx \) does not exist (is \( \infty \)).

For a continuously differentiable function \( f \colon [a,b] \to \mathbb{R} \), define \( \Vert f \Vert_{C^1} := \Vert f \Vert_u + \Vert f' \Vert_u \).

For the following two exercises let us define for a Riemann integrable function \( f \colon [0,1] \to \mathbb{R} \) the following number \( \Vert f \Vert_{C^1} := \int_0^1 \lvert f'(x) \rvert \, dx \).

By Dini’s theorem: Let \( f_n \colon [a,b] \to \mathbb{R} \) be a sequence of continuous functions such that \( 0 \leq f_{n+1}(x) \leq f_n(x) \leq \cdots \leq f_1(x) \). Suppose \( \{ f_n \} \) converges pointwise to \( f \). Show that \( \{ f_n \} \) converges uniformly.

Find a sequence of Riemann integrable functions \( \{ f_n \} \) on \([0,1])\) that converges pointwise to \( f \), and such that a) \( \lim_{n \to \infty} \int_0^1 f_n \, dx \) increases without bound, b) \( \lim_{n \to \infty} \int_0^1 f_n \, dx \) is the sequence \(-1,1,-1,1,-1,1, \ldots\).
It is possible to define a joint limit of a double sequence \( \{ x_{n,m} \} \) of real numbers (that is a function from \( \{ \mathbb{N} \times \mathbb{N} \} \) to \( \mathbb{R} \)). We say \( \lim \) is the joint limit of \( \{ x_{n,m} \} \) and write \( \lim_{(n,m) \to \infty} x_{n,m} = L \) if for every \( \epsilon > 0 \), there exists an \( M \) such that if \( n \geq M \) and \( m \geq M \), then \( |x_{n,m} - L| < \epsilon \).

Suppose the joint limit of \( \{ x_{n,m} \} \) is \( L \), and suppose that for all \( n \), \( \lim_{m \to \infty} x_{n,m} \) exists, and for all \( m \), \( \lim_{n \to \infty} x_{n,m} \) exists. Then show \( \lim_{n \to \infty} \lim_{m \to \infty} x_{n,m} = L \).

A joint limit does not mean the iterated limits even exist. Consider \( x_{n,m} := \min \{n,m \} \).

a) Show that for no \( n \) does \( \lim_{m \to \infty} x_{n,m} \) exist, and for no \( m \) does \( \lim_{n \to \infty} x_{n,m} \) exist. So neither \( \lim_{m \to \infty} \lim_{n \to \infty} x_{n,m} \) nor \( \lim_{n \to \infty} \lim_{m \to \infty} x_{n,m} \) makes any sense at all.

b) Show that the joint limit of \( \{ x_{n,m} \} \) exists and is 0.

**Picard’s theorem**

*Note: 1–2 lectures (can be safely skipped)*

A first semester course in analysis should have a *pièce de résistance* caliber theorem. We pick a theorem whose proof combines everything we have learned. It is more sophisticated than the fundamental theorem of calculus, the first highlight theorem of this course. The theorem we are talking about is Picard’s theorem on existence and uniqueness of a solution to an ordinary differential equation. Both the statement and the proof are beautiful examples of what one can do with all we have learned. It is also a good example of how analysis is applied as differential equations are indispensable in science.

**First order ordinary differential equation**

Modern science is described in the language of *differential equations*. That is, equations involving not only the unknown, but also its derivatives. The simplest nontrivial form of a differential equation is the so-called *first order ordinary differential equation* \( y' = F(x,y) \). Generally we also specify \( y(x_0) = y_0 \). The solution of the equation is a function \( y(x) \) such that \( y(x_0) = y_0 \) and \( y'(x) = F(y(x)) \). When \( F \) involves only the \( x \) variable, the solution is given by the fundamental theorem of calculus. On the other hand, when \( F \) depends on both \( x \) and \( y \) we need far more firepower. It is not always true that a solution exists, and if it does, that it is the unique solution. Picard’s theorem gives us certain sufficient conditions for existence and uniqueness.
The theorem

We need a definition of continuity in two variables. First, a point in the plane \( (\mathbb{R} \times \mathbb{R}) \) is denoted by an ordered pair \((x, y)\). To make matters simple, let us give the following sequential definition of continuity.

Let \( U \subset \mathbb{R}^2 \) be a set and \( F : U \to \mathbb{R} \) be a function. Let \((x, y) \in U\) be a point. The function \( F \) is continuous at \((x, y)\) if for every sequence \( \{ (x_n, y_n) \} \) of points in \( U \) such that \( \lim_{n \to \infty} x_n = x \) and \( \lim_{n \to \infty} y_n = y \), we have \( \lim_{n \to \infty} F(x_n, y_n) = F(x, y) \). We say \( F \) is continuous if it is continuous at all points in \( U \).

Let \( I, J \subset \mathbb{R} \) be closed bounded intervals, let \( \{ I_n \} \) and \( \{ J_n \} \) be their interiors, and let \( \{ (x_n, y_n) \} \subset I_0 \times J_0 \). Suppose \( F : I \times J \to \mathbb{R} \) is continuous and Lipschitz in the second variable, that is, there exists a number \( L \) such that \( \left| F(x, y) - F(x, z) \right| \leq L \left| y - z \right| \) for all \( y, z \in J \), \( x \in I \). Then there exists an \( h > 0 \) and a unique differentiable function \( f : [x_0 - h, x_0 + h] \to J \) such that \( f'(x) = F(x, f(x)) \) and \( f(x_0) = y_0 \).

Suppose we could find a solution \( f \). Using the fundamental theorem of calculus we integrate the equation \( f'(x) = F(x, f(x)) \), \( f(x_0) = y_0 \), and write \([\text{picard:diffeq}]\) as the integral equation \([\text{picard:inteq}]\).

Without loss of generality, suppose \( x_0 = 0 \) (exercise below). Another exercise tells us that \( (F) \) is bounded as it is continuous. Therefore pick some \( \{ M > 0 \} \) so that \( \{ \left| F(x, y) \right| \leq M \} \) for all \( \{ (x, y) \} \in \{ [0,1] \times [0,1] \} \). Pick \( \alpha > 0 \) such that \( \{ [0,1] \} \subset I_0 \times J_0 \). Define \( h := \min \{ \alpha, \frac{\alpha}{M + L \alpha} \} \). Observe \( \{ [-h, h] \} \subset I_0 \times J_0 \).

Set \( f_0(x) := y_0 \). We define \( f_{k+1} \) inductively. Assuming \( \{ f_{k-1} \} \subset \{ [0,1] \} \times \{ [0,1] \} \), we see \( \{ F \} \) is continuous on \( \{ [-h, h] \} \), then \( \{ F(t, f_{k-1}(t)) \} \) is a well defined function of \( t \) for \( t \in [-h, h] \). Further if \( \{ f_{k-1} \} \) is continuous on \( \{ [-h, h] \} \), then \( \{ F(t, f_{k-1}(t)) \} \) is a well defined function of \( t \) for \( t \in [-h, h] \).

We wish to show that the sequence \( \{ f_k \} \) converges uniformly to some function on \( \{ [-h, h] \} \). First, for \( t \in [-h, h] \), we have the following useful bound \( \left| \left| \left| F(t, f_{k}(t)) \right| - \left| \left| F(t, f_{k-1}(t)) \right| \right| \right| \geq \left| \left| \left| F(t, f_{k}(t)) \right| - \left| \left| F(t, f_{k-1}(t)) \right| \right| \right| \frac{\alpha}{M + L \alpha} \geq h \). We now define \( \{ f_{k+1} \} \) and so on, and we have defined a sequence \( \{ f_k \} \) of functions. We need to show that it converges to a function \( f \) that solves the equation \([\text{picard:inteq}]\) and therefore \([\text{picard:diffeq}]\).
uniform norm, that is the supremum of \( \left\| f_n(t) - f_k(t) \right\| \) for \( t \in [-h,h] \). Now note that \( \left\| f_n(t) - f_k(t) \right\| \leq \left\| f_{n-1}(t) - f_{k-1}(t) \right\| \) for \( t \in [-h,h] \). Therefore, \( \left\| f_n(x) - f_k(x) \right\| \leq L \left\| f_{n-1} - f_{k-1} \right\| \left\| x \right\| \). Let \( C := \frac{L\alpha}{M+L\alpha} \) and note that \( C < 1 \). Taking supremum on the left-hand side we get \( \left\| f_n(x) - f_k(x) \right\| \leq C \left\| f_{n-1} - f_{k-1} \right\| \). Without loss of generality, suppose \( n \geq k \). Then by we can show \( \left\| f_n(x) - f_k(x) \right\| \leq C^k \left\| f_{n-k} - f_0 \right\| \). Therefore, \( \left\| f_n(x) - f_k(x) \right\| \leq C^k \left\| f_{n-k} - f_0 \right\| \). As \( C < 1 \), \( \left\{ f_n \right\} \) is uniformly Cauchy and by we obtain that \( \left\| f_n(x) - f_k(x) \right\| \leq C^k \alpha \). Furthermore since all the \( \left\{ f_n([-h,h]) \right\} \) is uniformly limited to \( \mathbb{R} \), then \( f([-h,h]) \) is continuous. Furthermore since all the \( \left\{ f_n([-h,h]) \right\} \) is uniformly limited to \( \mathbb{R} \), then \( f([-h,h]) \) is continuous.

We now need to show that \( f \) solves the Picard integralequation. First, as before we notice \( \left\| f_n(t) - f(t) \right\| \leq L \left\| f_n - f \right\| \). As \( L \left\| f_n - f \right\| \to 0 \) as \( n \to \infty \), then \( f_n(t) \) converges uniformly to \( f(t) \) for \( t \in [-h,h] \). Hence, for \( x \in [-h,h] \) the convergence is uniform for \( t \in [0,x] \). Therefore, \( \left\| f_n(x) - f_0(x) \right\| \leq C^k \left\| f_{n-k} - f_0 \right\| \). As \( C < 1 \), \( \left\{ f_n \right\} \) is uniformly Cauchy and by we obtain that \( \left\| f_n(x) - f_k(x) \right\| \leq C^k \alpha \). Finally, what is left to do is to show uniqueness. Suppose \( g \) is another solution. As before we use the fact that \( \left\| f_n(t) - f(t) \right\| \leq L \left\| f_n - f \right\| \). Therefore, \( \left\| f_n(t) - f(t) \right\| \leq L \left\| f_n - f \right\| \). As \( L \left\| f_n - f \right\| \to 0 \) as \( n \to \infty \), then \( f_n(t) \) converges uniformly to \( f(t) \) for \( t \in [-h,h] \). Hence, for \( x \in [-h,h] \) the convergence is uniform for \( t \in [0,x] \). Therefore, \( \left\| f_n(x) - f_0(x) \right\| \leq C^k \left\| f_{n-k} - f_0 \right\| \). As \( C < 1 \), \( \left\{ f_n \right\} \) is uniformly Cauchy and by we obtain that \( \left\| f_n(x) - f_k(x) \right\| \leq C^k \alpha \). Finally, what is left to do is to show uniqueness. Suppose \( g \) is another solution. As before we use the fact that \( \left\| f_n(t) - f(t) \right\| \leq L \left\| f_n - f \right\| \). Therefore, \( \left\| f_n(t) - f(t) \right\| \leq L \left\| f_n - f \right\| \). As \( L \left\| f_n - f \right\| \to 0 \) as \( n \to \infty \), then \( f_n(t) \) converges uniformly to \( f(t) \) for \( t \in [-h,h] \). Hence, for \( x \in [-h,h] \) the convergence is uniform for \( t \in [0,x] \). Therefore, \( \left\| f_n(x) - f_0(x) \right\| \leq C^k \left\| f_{n-k} - f_0 \right\| \). As \( C < 1 \), \( \left\{ f_n \right\} \) is uniformly Cauchy and by we obtain that \( \left\| f_n(x) - f_k(x) \right\| \leq C^k \alpha \). Finally, we prove that \( f \) is differentiable and its derivative is \( f'(x) \). It is obvious that \( f(0) = y_0 \).

Examples

Let us look at some examples. The proof of the theorem gives us an explicit way to find an \( \left( h \right) \) that works. It does not, however, give us the best \( \left( h \right) \). It is often possible to find a much larger \( \left( h \right) \) for which the conclusion of the theorem holds.

The proof also gives us the Picard iterates as approximations to the solution. So the proof actually tells us how to obtain the solution, not just that the solution exists.
Consider \( f'(x) = f(x), \) \( \text{and we are looking for a function } f \) \( \text{such that } f(x) = f(x). \) We pick any \( (I) \) that contains 0 in the interior. We pick an arbitrary \( (J) \) that contains 1 in its interior. We can use \( (L = 1). \) The theorem guarantees an \( (h > 0) \) such that there exists a unique solution \( f(\mathbb{R}) \) \( \text{to } (f(x) = f(x)). \) This solution is usually denoted by \( e^x := f(x) \) \( \text{We leave it to the reader to verify that by picking } (I) \) \( \text{and } (J) \) \( \text{large enough the proof of the theorem guarantees that we are able to pick } (\alpha) \) \( \text{sufficiently large enough that we get any } (h) \) \( \text{we want as long as } h < \frac{1}{2}. \) We omit the calculation.

Of course, we know this function exists as a function for all \( (x) \), so an arbitrary \( (h) \) ought to work. By same reasoning as above, no matter what \( (x_0) \) \( \text{and } (y_0) \) \( \text{are, the proof guarantees an arbitrary } (h) \) \( \text{as long as } h < \frac{1}{2}. \) Fix such an \( (h) . \) We get a unique function defined on \( [x_0-h,x_0+h] \) \( \text{After defining the function on } (h) \) \( \text{we find a solution on the interval } ([0,2h]) \) \( \text{and notice that the two functions must coincide on } ([0,2h]) \) \( \text{by uniqueness. We thus iteratively construct the exponential for all } (x) \text{in } \mathbb{R}. \) Therefore Picard’s theorem could be used to prove the existence and uniqueness of the exponential.

Let us compute the Picard iterates. We start with the constant function \( f_0(x) := 1 \). Then \[ \begin{aligned} f_1(x) &= 1 + \int_0^x f_0(s)~ds = 1+x, \quad f_2(x) &= 1 + \int_0^x f_1(s)~ds = 1 + \frac{x^2}{2}, \quad f_3(x) &= 1 + \int_0^x f_2(s)~ds = 1 + \frac{x^2}{2} + \frac{x^3}{6}. \end{aligned} \] We recognize the beginning of the Taylor series for the exponential.

Suppose we have the equation \( f'(x) = {f(x)}^2 \) \( \text{and } f(0)=1. \) From elementary differential equations we know \( f(x) = \frac{1}{1-x} \) is the solution. The solution is only defined on \( (-\infty,1) \). That is, we are able to use \( h < 1 \), but never a larger \( h \). The function that takes \( y(x) \) to \( y^2 \) is not Lipschitz as a function on all of \( \mathbb{R} \). As we approach \( x=1 \) \( \text{from the left, the solution becomes larger and larger. The derivative of the solution grows as } \frac{1}{(y-x)^2}, \) \( \text{and therefore the } (L) \text{ required will have to be larger and larger as } y_0 \text{ grows. Thus if we apply the theorem with } (x_0) \text{ close to 1 and } (y_0) = \frac{1}{1-x_0} \text{ we find that the } (h) \text{ that the proof guarantees will be smaller and smaller as } (x_0) \text{ approaches 1.}

By picking \( (\alpha) \) correctly, the proof of the theorem guarantees \( h = 1-\frac{\sqrt{3}}{2} \approx 0.134 \) (we omit the calculation) for \( (x_0=0) \) \( \text{and } (y_0=1), \) even though we saw above that any \( h < 1 \) \text{ should work.}

Consider the equation \( f'(x) = 2 \sqrt{\left| f(x) \right|}, \) \( f(0) = 0. \) The function \( f(x,y) = 2 \sqrt{\left| f(x) \right|} \) \( \text{is continuous, but not Lipschitz in } (y) \text{ (why?). The equation does not satisfy the hypotheses of the theorem. The function } f(x) = \begin{cases} x^2 & \text{if } x \leq 0, \\ -x^2 & \text{if } x > 0, \end{cases} \text{ is a solution, but } f(x) = 0 \text{ is also a solution. A solution exists, but is not unique.}

Consider \( y' = \varphi(x) \) \( \text{where } (\varphi(x) := 0 \text{ if } (x \in \mathbb{Q}) \text{ and } (\varphi(x):=1 \text{ if } (x \not\in \mathbb{Q}). \text{ The equation has no solution regardless of the initial conditions. A solution would have derivative } (\varphi(x) \text{, but } (\varphi(x) \text{ does not have the intermediate value property at any point (why?). No solution exists by . Therefore to obtain existence of a solution, some continuity hypothesis on } (F) \text{ is necessary.} \)
Exercises

Let \( I, J \subset \mathbb{R} \) be intervals. Let \( F : I \times J \to \mathbb{R} \) be a continuous function of two variables and suppose \( f : I \to J \) be a continuous function. Show that \( F(x, f(x)) \) is a continuous function on \( I \).

Let \( I, J \subset \mathbb{R} \) be closed bounded intervals. Show that if \( F : I \times J \to \mathbb{R} \) is continuous, then \( F \) is bounded.

We proved Picard’s theorem under the assumption that \( x_0 = 0 \). Prove the full statement of Picard’s theorem for an arbitrary \( x_0 \).

Let \( f(x) = x f(x) \) be our equation. Start with the initial condition \( f(0) = 2 \) and find the Picard iterates \( f_0, f_1, f_2, f_3, f_4 \).

Suppose \( F : I \times J \to \mathbb{R} \) is a function that is continuous in the first variable, that is, for any fixed \( y \) the function that takes \( x \) to \( F(x, y) \) is continuous. Further, suppose \( F \) is Lipschitz in the second variable, that is, there exists a number \( L \) such that \[ \left| F(x, y) - F(x, z) \right| \leq L \left| y - z \right| \quad \text{for all } y, z \in J, \ x \in I \]. Show that \( F \) is continuous as a function of two variables. Therefore, the hypotheses in the theorem could be made even weaker.

A common type of equation one encounters are linear first order differential equations, that is equations of the form \( y' + p(x) y = q(x) \). Prove Picard’s theorem for linear equations. Suppose \( I \) is an interval, \( x_0 \in I \), and \( (p : I \to \mathbb{R}) \) and \( (q : I \to \mathbb{R}) \) are continuous. Show that there exists a unique differentiable \( f : I \to \mathbb{R} \), such that \( y = f(x) \) satisfies the equation and the initial condition. Hint: Assume existence of the exponential function and use the integrating factor formula for existence of \( f(x) \) (prove that it works):

\[
f(x) := e^{\int_{x_0}^{x} p(s) \, ds} \left( \int_{x_0}^{x} e^{\int_{x_0}^{s} p(t) \, dt} q(t) \, dt + y_0 \right) .
\]

Metric Spaces

Metric spaces

Note: 1.5 lectures

As mentioned in the introduction, the main idea in analysis is to take limits. In we learned to take limits of sequences of real numbers. And in we learned to take limits of functions as a real number approached some other real number.

We want to take limits in more complicated contexts. For example, we want to have sequences of points in 3-dimensional space. We wish to define continuous functions of several variables. We even want to define functions on spaces that are a little harder to describe, such as the surface of the earth. We still want to talk about limits there.
Finally, we have seen the limit of a sequence of functions in . We wish to unify all these notions so that we do not have to reprove theorems over and over again in each context. The concept of a metric space is an elementary yet powerful tool in analysis. And while it is not sufficient to describe every type of limit we find in modern analysis, it gets us very far indeed.

Let \((X)\) be a set, and let \(d : X \times X \to \mathbb{R}\) be a function such that

i. \([\text{metric:pos}]\) \(d(x,y) \geq 0\) for all \((x, y)\) in \(X\),

ii. \([\text{metric:zero}]\) \(d(x,y) = 0\) if and only if \((x = y)\),

iii. \([\text{metric:com}]\) \(d(x,y) = d(y,x)\),

iv. \([\text{metric:triang}]\) \(d(x,z) \leq d(x,y) + d(y,z)\) (triangle inequality).

Then the pair \((X,d)\) is called a metric space. The function \(d\) is called the metric or sometimes the distance function. Sometimes we just say \((X)\) is a metric space if the metric is clear from context.

The geometric idea is that \(d\) is the distance between two points. Items \([\text{metric:pos}]–[\text{metric:com}]\) have obvious geometric interpretation: distance is always nonnegative, the only point that is distance 0 away from \((x)\) is \((x)\) itself, and finally that the distance from \((x)\) to \((y)\) is the same as the distance from \((y)\) to \((x)\). The triangle inequality \([\text{metric:triang}]\) has the interpretation given in .

For the purposes of drawing, it is convenient to draw figures and diagrams in the plane and have the metric be the standard distance. However, that is only one particular metric space. Just because a certain fact seems to be clear from drawing a picture does not mean it is true. You might be getting sidetracked by intuition from euclidean geometry, whereas the concept of a metric space is a lot more general.

Let us give some examples of metric spaces.

The set of real numbers \((\mathbb{R})\) is a metric space with the metric \(|d(x,y) := \left| x-y \right|\). Items \([\text{metric:pos}]–[\text{metric:com}]\) of the definition are easy to verify. The triangle inequality \([\text{metric:triang}]\) follows immediately from the standard triangle inequality for real numbers: \(|d(x,z) = \left| x-z \right| = \left| x-y+y-z \right| \leq \left| x-y \right| + \left| y-z \right| = d(x,y) + d(y,z)\). This metric is the standard metric on \((\mathbb{R})\). If we talk about \((\mathbb{R})\) as a metric space without mentioning a specific metric, we mean this particular metric.

We can also put a different metric on the set of real numbers. For example, take the set of real numbers \((\mathbb{R})\) together with the metric \(|d(x,y) := \frac{\left| x-y \right|}{\left| x-y \right|+1}\). Items \([\text{metric:pos}]–[\text{metric:com}]\) are again easy to verify. The triangle inequality \([\text{metric:triang}]\) is a little bit more difficult. Note that \(|d(x,y) = \text{varphi}(\left| x-y \right|)\) where \(|\text{varphi}(t) = \frac{t}{t+1}\) and \(|\text{varphi}(t)\) is an increasing function (positive derivative). Hence \(|\begin{split} d(x,z) & = \text{varphi}(\left| x-z \right|) = \text{varphi}(\left| x-y+y-z \right|) \\
& = \text{varphi}(\left| x-y \right|+\left| y-z \right|) \leq \text{varphi}(\left| x-y \right|+\left| y-z \right|+\left| y-z \right|) = \frac{\left| x-y \right|+\left| y-z \right|+\left| x-y \right|}{\left| x-y \right|+\left| y-z \right|+\left| x-y \right|} \end{split}|\).
An important metric space is the \( (n) \)-dimensional euclidean space \( (\mathbb{R}^n) \). We use the following notation for points: \( x = (x_1, x_2, \ldots, x_n) \). We also simply write \((0, 0, \ldots, 0) \in (\mathbb{R}^n) \). Before making \((\mathbb{R}^n)\) a metric space, let us prove an important inequality, the so-called Cauchy-Schwarz inequality.

Take \( x = (x_1, x_2, \ldots, x_n) \) and \( y = (y_1, y_2, \ldots, y_n) \). Then \( \left( \sum_{j=1}^n x_j y_j \right)^2 \leq \left( \sum_{j=1}^n x_j^2 \right) \left( \sum_{j=1}^n y_j^2 \right) \).

Any square of a real number is nonnegative. Hence any sum of squares is nonnegative: \( \sum_{j=1}^n \left( x_j y_k - x_k y_j \right)^2 = \sum_{j=1}^n \left( x_j^2 y_k^2 + x_k^2 y_j^2 - 2 x_j x_k y_j y_k \right) \). We relabel and divide by 2 to obtain \( \sum_{j=1}^n \left( x_j - y_j \right)^2 + \sum_{j=1}^n \left( y_j - z_j \right)^2 = \left( \sum_{j=1}^n x_j - y_j \right)^2 + \left( \sum_{j=1}^n y_j - z_j \right)^2 \).

Let us construct the standard metric for \( (\mathbb{R}^n) \). Define \( d(x,y) := \sqrt{ \sum_{j=1}^n (x_j - y_j)^2 } \).

An example to keep in mind is the so-called discrete metric. Let \( X \) be any set and define \( d(x,y) := \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases} \) That is, all points are equally distant from each other. When \( X \) is an infinite set such as the real numbers.

While this particular example seldom comes up in practice, it gives a useful “smell test.” If you make a statement about metric spaces, try it with the discrete metric. To show that \((X,d)\) is indeed a metric space is left as an exercise.

**Example:** Let \( C([a,b], \mathbb{R}) \) be the set of continuous real-valued functions on the interval \([a,b]\). Define the metric on \( C([a,b], \mathbb{R}) \) as \( d(f,g) := \sup_{x \in [a,b]} \left| f(x) - g(x) \right| \). Let us check the triangle inequality.
properties. First, \(|d(f,g)|\) is finite as \(|f(x)-g(x)|\) is a continuous function on a closed bounded interval \([a,b]\), and so is bounded. It is clear that \(|d(f,g)\geq 0\), it is the supremum of nonnegative numbers. If \(f = g\) then \(||f(x)-g(x)|\) \(|\text{right}|\) is bounded. It is clear that \(|d(f,g)\geq 0\), it is the supremum of nonnegative numbers. If \(f = g\) then \(||f(x)-g(x)|\) \(|\text{right}|\) is bounded. Conversely if \(d(f,g) = 0\), then for any \(x\) we have \(||f(x)-g(x)|\) \(|\text{right}|\) is bounded. That \(|d(f,g)| = |d(g,f)|\) is \(|\text{right}|\) trivial. To show the triangle inequality we use the standard triangle inequality. \[\begin{split} d(f,g) & = \sup_{x \in [a,b]} |f(x)-g(x)| = \sup_{x \in [a,b]} |f(x)-h(x)+h(x)-g(x)| \leq \sup_{x \in [a,b]} (|f(x)-h(x)|+|h(x)-g(x)|) \leq \sup_{x \in [a,b]} |f(x)-h(x)|+\sup_{x \in [a,b]} |h(x)-g(x)| = d(f,h) + d(h,g). \end{split}\] When treating \((C([a,b], \mathbb{R}))\) as a metric space without mentioning a metric, we mean this particular metric. Notice that \(|d(f,g)| = \|f-g\|_u\), the uniform norm of .

This example may seem esoteric at first, but it turns out that working with spaces such as \((C([a,b], \mathbb{R}))\) is really the meat of a large part of modern analysis. Treating sets of functions as metric spaces allows us to abstract away a lot of the grubby detail and prove powerful results such as Picard’s theorem with less work.

Oftentimes it is useful to consider a subset of a larger metric space as a metric space itself. We obtain the following proposition, which has a trivial proof.

Let \((X,d)\) be a metric space and \((Y \subset X)\), then the restriction \((d|_{\{Y \times Y\}})\) is a metric on \((Y)\).

If \((X,d)\) is a metric space, \((Y \subset X)\), and \((d' := d|_{\{Y \times Y\}})\), then \((Y,d')\) is said to be a \textit{subspace} of \((X,d)\).

It is common to simply write \((d')\) for the metric on \((Y)\), as it is the restriction of the metric on \((X)\). Sometimes we say \((d')\) is the \textit{subspace metric} and \((Y)\) has the \textit{subspace topology}.

A subset of the real numbers is bounded whenever all its elements are at most some fixed distance from 0. We also define bounded sets in a metric space. When dealing with an arbitrary metric space there may not be some natural fixed point 0. For the purposes of boundedness it does not matter.

Let \((X,d)\) be a metric space. A subset \((S \subset X)\) is said to be \textit{bounded} if there exists a \((p \in X)\) and a \((B \in \{\mathbb{R}\})\) such that \(|d(p,x)\leq B \quad \forall x \in S\}. \) We say \((X,d)\) is bounded if \((X)\) itself is a bounded subset.

For example, the set of real numbers with the standard metric is not a bounded metric space. It is not hard to see that a subset of the real numbers is bounded in the sense of if and only if it is bounded as a subset of the metric space of real numbers with the standard metric.

On the other hand, if we take the real numbers with the discrete metric, then we obtain a bounded metric space. In fact, any set with the discrete metric is bounded.
Exercises

Show that for any set \( \langle X \rangle \), the discrete metric \( \langle d(x,y) = 1 \rangle \) if \( x \not= y \) and \( d(x,x) = 0 \)) does give a metric space \( \langle (X,d) \rangle \).

Let \( \langle X := \{ 0 \} \rangle \) be a set. Can you make it into a metric space?

Let \( \langle X := \{ a, b \} \rangle \) be a set. Can you make it into two distinct metric spaces? (define two distinct metrics on it)

Let the set \( \langle X := \{ A, B, C \} \rangle \) represent 3 buildings on campus. Suppose we wish our distance to be the time it takes to walk from one building to the other. It takes 5 minutes either way between buildings \( \langle A \rangle \) and \( \langle B \rangle \). However, building \( \langle C \rangle \) is on a hill and it takes 10 minutes from \( \langle A \rangle \) and 15 minutes from \( \langle B \rangle \) to get to \( \langle C \rangle \). On the other hand it takes 5 minutes to go from \( \langle C \rangle \) to \( \langle A \rangle \) and 7 minutes to go from \( \langle C \rangle \) to \( \langle B \rangle \), as we are going downhill. Do these distances define a metric? If so, prove it, if not, say why not.

Suppose \( \langle (X,d) \rangle \) is a metric space and \( \langle \varphi \colon [0,\infty) \to \{ \text{mathbb \{R\} } \rangle \) is a function such that \( \langle \varphi(t) \geq 0 \rangle \) for all \( t \) and \( \langle \varphi(t) = 0 \rangle \) if and only if \( (t=0) \). Also suppose \( \langle \varphi \rangle \) is \textit{subadditive}, that is, \( \langle \varphi(s+t) \leq \varphi(s)+\varphi(t) \rangle \). Show that with \( \langle d'(x,y) := \varphi(bigr(d(x,y)bigr)) \rangle \), we obtain a new metric space \( \langle (X,d') \rangle \).

[exercise:mscross] Let \( \langle (X,d) \rangle \) and \( \langle (Y,d) \rangle \) be metric spaces.

a) Show that \( \langle (X \times Y,d) \rangle \) with \( \langle d((x_1,y_1),(x_2,y_2)) := d_X(x_1,x_2) + d_Y(y_1,y_2) \rangle \) is a metric space.

b) Show that \( \langle (X \times Y,d) \rangle \) with \( \langle d((x_1,y_1),(x_2,y_2)) := \max \{ d_X(x_1,x_2), d_Y(y_1,y_2) \} \rangle \) is a metric space.

Let \( \langle X \rangle \) be the set of continuous functions on \( \langle [0,1] \rangle \). Let \( \langle \varphi \colon [0,1] \to (0,\infty) \rangle \) be continuous. Define \( \langle d(f,g) := \int_0^1 |f(x)-g(x)| \varphi(x)\ dx \rangle \) Show that \( \langle (X,d) \rangle \) is a metric space.

[exercise:mshausdorffpseudo] Let \( \langle (X,d) \rangle \) be a metric space. For nonempty bounded subsets \( \langle A \rangle \) and \( \langle B \rangle \) let \( \langle d(A,B) := \inf \{ d(x,b) : b \in B \} \rangle \) qquad \text{and} \langle d(A,B) := \sup \{ d(a,B) : a \in A \} \rangle \) Now define the \textit{Hausdorff metric} as \( \langle d_H(A,B) := \max \{ d(A,B), d(B,A) \} \rangle \) \text{Note:} \( \langle d_H \rangle \) can be defined for arbitrary nonempty subsets if we allow the extended reals.

a) Let \( \langle Y \rangle \subset \{ \text{mathcal \{P\} } \} \) be the set of bounded nonempty subsets. Prove that \( \langle (Y,d) \rangle \) is a so-called \textit{pseudometric space}: \( \langle d(H) \rangle \) satisfies the metric properties \( \langle \text{metric:pos} \rangle \), \( \langle \text{metric:com} \rangle \), \( \langle \text{metric:triang} \rangle \), and further \( \langle d(H,A) = 0 \rangle \) for all \( \langle A \in Y \rangle \).

b) Show by example that \( \langle d \rangle \) itself is not symmetric, that is \( \langle d(A,B) \not= d(B,A) \rangle \).

c) Find a metric space \( \langle X \rangle \) and two different nonempty bounded subsets \( \langle A \rangle \) and \( \langle B \rangle \) such that \( \langle d(H,A,B) = 0 \rangle \).

Open and closed sets

Note: 2 lectures
Topology

It is useful to define a so-called topology. That is we define closed and open sets in a metric space. Before doing so, let us define two special sets.

Let \((X,d)\) be a metric space, \(x \in X\) and \(\delta > 0\). Then define the open ball or simply ball of radius \(\delta\) around \(x\) as

\[B(x,\delta) := \{ y \in X : d(x,y) < \delta \} .\]

Similarly we define the closed ball as

\[C(x,\delta) := \{ y \in X : d(x,y) \leq \delta \} .\]

When we are dealing with different metric spaces, it is sometimes convenient to emphasize which metric space the ball is in. We do this by writing \(B_X(x,\delta) := B(x,\delta)\) or \((C_X(x,\delta) := C(x,\delta))\).

Take the metric space \((\mathbb{R},d)\) with the standard metric. For \(x \in \mathbb{R}\), and \(\delta > 0\) we get

\[B(x,\delta) = (x-\delta,x+\delta) \quad \text{and} \quad C(x,\delta) = [x-\delta,x+\delta] .\]

Be careful when working on a subspace. Suppose we take the metric space \(\{[0,1]\}\) as a subspace of \((\mathbb{R},d)\). Then in \(\{[0,1]\}\) we get

\[B(0,\frac{1}{2}) = B_{[0,1]}(0,\frac{1}{2}) = [0,\frac{1}{2}) .\]

The important thing to keep in mind is which metric space we are working in.

Let \((X,d)\) be a metric space. A set \(V \subset X\) is open if for every \(x \in V\), there exists a \(\delta > 0\) such that \(B(x,\delta) \subset V\). A set \(E \subset X\) is closed if the complement \(E^c = X \setminus E\) is open. When the ambient space \(X\) is not clear from context we say \(V\) is open in \(X\) and \(E\) is closed in \(X\).

If \(x \in V\) and \(V\) is open, then we say \(V\) is an open neighborhood of \(x\) (or sometimes just neighborhood).

Intuitively, an open set is a set that does not include its “boundary,” wherever we are at in the set, we are allowed to “wiggle” a little bit and stay in the set. Note that not every set is either open or closed, in fact generally most subsets are neither.

The set \(\{(0,1)\} \subset \{\mathbb{R}\}\) is neither open nor closed. First, every ball in \(\{\mathbb{R}\}\) around \(0\), \(\{(-\delta,\delta)\}\) contains negative numbers and hence is not contained in \(\{(0,1)\}\) and so \(\{(0,1)\}\) is not open. Second, every ball in \(\{\mathbb{R}\}\) around \(1\), \((1-\delta,1+\delta)\) contains numbers strictly less than 1 and greater than 0 (e.g. \(1-\nicefrac{\delta}{2}\) as long as \(\delta < 2\)). Thus \(\{\mathbb{R}\} \setminus \{0,1\}\) is not open, and so \(\{(0,1)\}\) is not closed.

[prop:topology:open] Let \((X,d)\) be a metric space.

i. [topology:openi] \(\emptyset\) and \(X\) are open in \(X\).
ii. [topology:openii] If \( V_1, V_2, \ldots, V_k \) are open then \( \bigcap_{j=1}^k V_j \) is also open. That is, finite intersection of open sets is open.

iii. [topology:openiii] If \( \{ V_{\lambda} \}_{\lambda \in I} \) is an arbitrary collection of open sets, then \( \bigcup_{\lambda \in I} V_\lambda \) is also open. That is, union of open sets is open.

Note that the index set in [topology:openiii] is arbitrarily large. By \( \bigcup_{\lambda \in I} V_\lambda \) we simply mean the set of all \( x \) such that \( x \in V_\lambda \) for at least one \( \lambda \in I \).

The sets \( (X) \) and \( (\emptyset) \) are obviously open in \( (X) \).

Let us prove [topology:openii]. If \( x \in \bigcap_{j=1}^k V_j \), then \( x \in V_j \) for all \( j \). As \( V_j \) are all open, for every \( j \) there exists a \( \delta_j > 0 \) such that \( B(x, \delta_j) \subset V_j \). Take \( \delta := \min \{ \delta_1, \delta_2, \ldots, \delta_k \} \) and notice \( \delta > 0 \). We have \( B(x, \delta) \subset B(x, \delta_j) \subset V_j \) for every \( j \) and so \( B(x, \delta) \subset \bigcap_{j=1}^k V_j \). Consequently the intersection is open.

Let us prove [topology:openiii]. If \( x \in \bigcup_{\lambda \in I} V_\lambda \), then \( x \in V_\lambda \) for some \( \lambda \in I \). As \( V_\lambda \) is open, there exists a \( \delta > 0 \) such that \( B(x, \delta) \subset V_\lambda \). But then \( B(x, \delta) \subset \bigcup_{\lambda \in I} V_\lambda \) and so the union is open.

The main thing to notice is the difference between items [topology:openii] and [topology:openiii]. Item [topology:openii] is not true for an arbitrary intersection, for example \( \bigcap_{n=1}^\infty (-\nicefrac{1}{n}, \nicefrac{1}{n}) = \{ 0 \} \), which is not open.

The proof of the following analogous proposition for closed sets is left as an exercise.

[prop:topology:closed] Let \( (X,d) \) be a metric space.

i. [topology:closedi] \( (\emptyset) \) and \( (X) \) are closed in \( (X) \).

ii. [topology:closedii] If \( \{ E_{\lambda} \}_{\lambda \in I} \) is an arbitrary collection of closed sets, then \( \bigcap_{\lambda \in I} E_\lambda \) is also closed. That is, intersection of closed sets is closed.

iii. [topology:closediii] If \( (E_1, E_2, \ldots, E_k) \) are closed then \( \bigcup_{j=1}^k E_j \) is also closed. That is, finite union of closed sets is closed.

We have not yet shown that the open ball is open and the closed ball is closed. Let us show this fact now to justify the terminology.

[prop:topology:ballsopenclosed] Let \( (X,d) \) be a metric space, \( (x \in X) \), and \( (\delta > 0) \). Then \( B(x, \delta) \) is open and \( C(x, \delta) \) is closed.
Let \( y \in B(x, \delta) \). Let \( \alpha := \delta - d(x, y) \). Of course \( \alpha > 0 \). Now let \( z \in B(y, \alpha) \). Then
\[
d(x, z) \leq d(x, y) + d(y, z) < d(x, y) + \alpha = d(x, y) + \delta - d(x, y) = \delta.
\]
Therefore \( z \in B(x, \delta) \) for every \( z \in B(y, \alpha) \). So \( B(y, \alpha) \subseteq B(x, \delta) \) and \( B(x, \delta) \) is open.

The proof that \( C(x, \delta) \) is closed is left as an exercise.

Again be careful about what is the ambient metric space. As \( ([0, \frac{1}{2}] \subseteq \mathbb{R}) \) is an open ball in \( ([0, 1]) \), this means that \( ([0, \frac{1}{2}] \subseteq \mathbb{R}) \) is an open set in \( ([0, 1]) \). On the other hand \( ([0, \frac{1}{2}] \subseteq \mathbb{R}) \) is neither open nor closed in \( ([\mathbb{R}, \delta]) \).

A useful way to think about an open set is as a union of open balls. If \( \bigcup \{ U \} \) is open, then for each \( x \in U \), there is a \( \delta > 0 \) (depending on \( x \)) such that \( B(x, \delta) \subseteq U \). Then \( \bigcup \{ x \in U \} B(x, \delta) \subseteq U \).

The proof of the following proposition is left as an exercise. Note that there are many other open and closed sets in \( \{ \mathbb{R}, \delta \} \).

### Connected sets

A nonempty metric space \( (X, d) \) is **connected** if the only subsets of \( X \) that are both open and closed are \( \emptyset \) and \( X \) itself. If \( (X, d) \) is not connected we say it is **disconnected**.

When we apply the term **connected** to a nonempty subset \( A \subseteq X \), we simply mean that \( A \) with the subspace topology is connected.

In other words, a nonempty \( X \) is connected if whenever we write \( X = X_1 \cup X_2 \) where \( X_1 \cap X_2 = \emptyset \) and \( X_1 \) and \( X_2 \) are open, then either \( X_1 = \emptyset \) or \( X_2 = \emptyset \). So to show \( X \) is disconnected, we need to find nonempty disjoint open sets \( X_1 \) and \( X_2 \) whose union is \( X \). For subsets, we state this idea as a proposition.

Let \( (X, d) \) be a metric space. A nonempty set \( S \subseteq X \) is **not connected** if and only if there exist open sets \( U_1 \) and \( U_2 \) in \( \{ X, \delta \} \), such that \( U_1 \cap U_2 \cap S = \emptyset \), \( U_1 \cap S \not= \emptyset \), \( U_2 \cap S \not= \emptyset \), and \( S = U_1 \cup U_2 \cap S \).

If \( U_1 \) is open in \( X \), then \( U_1 \cap S \) is open in \( S \) in the subspace topology (with subspace metric). To see this, note that if \( (B_S(x, \delta) \subseteq U_1 \cap S) \), then \( (B_S(x, \delta) \subseteq S \cap B_S(x, \delta)) \), we have \( B_S(x, \delta) \subseteq U_1 \cap S \).

The proof of the other direction follows by using to find \( U_1 \) and \( U_2 \) from two open disjoint subsets of \( S \).

Let \( S \subseteq \mathbb{R} \) be such that \( x < z < y \) with \( (x, y) \in S \) and \( (z \notin S) \). Claim: \( S \) is not connected.

**Proof:** Notice \( (\{ x \in S \} \cap S) \cup (\{ z \in S \} \cap S) = S \).
A nonempty set $S \subset \mathbb{R}$ is connected if and only if it is an interval or a single point.

Suppose $\forall (S)$ is connected. If $\forall (S)$ is a single point then we are done. So suppose $\forall (x < y)$ and $\forall (x,y \in S)$. If $\forall (z)$ is such that $\forall (x < z < y)$, then $\varnothing \subset \mathbb{R}$ is nonempty and $\varnothing \subset \mathbb{R}$ is nonempty. The two sets are disjoint. As $\forall (S)$ is connected, we must have they their union is not $\forall (S)$, so $\forall (z \in S)$.

Suppose $\forall (S)$ is bounded, connected, but not a single point. Let $\forall (\alpha := \inf, S)$ and $\forall (\beta := \sup, S)$ and note that $\forall (\alpha < \beta)$. Suppose $\forall (\alpha < z < \beta)$. As $\forall (\alpha)$ is the infimum, then there is an $\forall (x \in S)$ such that $\forall (\alpha \leq z < x)$. Similarly there is a $\forall (y \in S)$ such that $\forall (y \geq z > \beta)$. We have shown above that $\forall (z \in S)$, so $\forall (\alpha, \beta) \subset S$. If $\forall (w \in \alpha)$, then $\forall (w \not\in S)$ as $\forall (\alpha)$ was the infimum, similarly if $\forall (w > \beta)$ then $\forall (w \not\in S)$. Therefore the only possibilities for $\forall (S)$ are $\forall (\alpha, \beta), \forall (\alpha, \beta), \forall (\alpha, \beta), \forall (\alpha, \beta)$. The proof that an unbounded connected $\forall (S)$ is an interval is left as an exercise.

On the other hand suppose $\forall (S)$ is an interval. Suppose $\forall (U_1)$ and $\forall (U_2)$ are open subsets of $\forall (\mathbb{R})$, $\forall (U_1 \cap S)$ and $\forall (U_2 \cap S)$ are nonempty, and $\forall (S = \bigl( U_1 \cap S \bigr) \cup \bigl( U_2 \cap S \bigr)$). We will show that $\forall (U_1 \cap S)$ and $\forall (U_2 \cap S)$ contain a common point, so they are not disjoint, and hence $\forall (S)$ must be connected.

Suppose there is $\forall (x \in U_1 \cap S)$ and $\forall (y \in U_2 \cap S)$. Without loss of generality, assume $\forall (x < y)$. As $\forall (S)$ is an interval $\forall ([x,y] \subset S)$. Let $\forall (z := \inf \bigl( U_2 \cap [x,y] \bigr))$. If $\forall (z = x)$, then $\forall (z \in U_1)$. If $\forall (z > x)$, then for any $\forall (\delta > 0)$ the ball $\forall (B(z,\delta) = (z-\delta, z+\delta))$ contains points of $\forall (S)$ that are not in $\forall (U_2)$, and so $\forall (z \not\in U_2)$ as $\forall (U_2)$ is open. Therefore, $\forall (z \in U_1)$. As $\forall (U_1)$ is open, $\forall (B(z,\delta) \subset \forall (U_1))$ for a small enough $\forall (\delta > 0)$. As $\forall (z)$ is the infimum of $\forall (U_2 \cap [x,y])$, there must exist some $\forall (w \in U_2 \cap [x,y])$ such that $\forall (w \in [z,z+\delta) \subset \forall (B(z,\delta) \subset \forall (U_1))$. Therefore $\forall (w \in U_1 \cap U_2 \cap [x,y])$. So $\forall (U_1 \cap S)$ and $\forall (U_2 \cap S)$ are not disjoint and hence $\forall (S)$ is connected.

In many cases a ball $\forall (B(x,\delta))$ is connected. But this is not necessarily true in every metric space. For a simplest example, take a two point space $\forall (\{ a, b \})$ with the discrete metric. Then $\forall (B(a,2) = \forall (\{ a, b \}))$, which is not connected as $\forall (B(a,1) = \forall (\{ a \})$ and $\forall (B(b,1) = \forall (\{ b \}))$ are open and disjoint.

**Closure and boundary**

Sometimes we wish to take a set and throw in everything that we can approach from the set. This concept is called the closure.

Let $\forall ([X,d])$ be a metric space and $\forall (A \subset X)$. Then the closure of $\forall (A)$ is the set $\forall (\overline{A} := \bigcap \{ E \subset X : \forall (E \text{ is closed and } A \subset E) \})$. That is, $\forall (\overline{A})$ is the intersection of all closed sets that contain $\forall (A)$.

Let $\forall ([X,d])$ be a metric space and $\forall (A \subset X)$. The closure $\forall (\overline{A})$ is closed. Furthermore if $\forall (A)$ is closed then $\forall (\overline{A} = A)$.

First, the closure is the intersection of closed sets, so it is closed. Second, if $\forall (A)$ is closed, then take $\forall (E = A)$, hence the intersection of all closed sets $\forall (E)$ containing $\forall (A)$ must be equal to $\forall (A)$. 
The closure of \((0,1)\) in \(\mathbb{R}\) is \([0,1]\). Proof: Simply notice that if \(E\) is closed and contains \((0,1)\), then \(\overline{E}\) must contain \(0\) and \(1\) (why?). Thus \([0,1]\ subset E\). But \([0,1]\) is also closed. Therefore the closure \(\overline{(0,1)} = [0,1]\).

Be careful to notice what ambient metric space you are working with. If \(X = (0,\infty)\), then the closure of \((0,1)\) in \((0,\infty)\) is \((0,1]\). Proof: Similarly as above \((0,1]\) is closed in \((0,\infty)\) (why?). Any closed set \(E\) that contains \((0,1]\) must contain 1 (why?). Therefore \((0,1]\ subset E\), and hence \(\overline{(0,1]} = (0,1]\) when working in \((0,\infty)\).

Let us justify the statement that the closure is everything that we can “approach” from the set.

[prop:msclosureappr] Let \((X,d)\) be a metric space and \(A subset X\). Then \(x \in \overline{A}\) if and only if for every \(\delta > 0\), \((B(x,\delta) \ cap A) \ not= \emptyset\).

Let us prove the two contrapositives. Let us show that \(x \notin \overline{A}\) if and only if there exists a \(\delta > 0\) such that \((B(x,\delta) \ cap A) = \emptyset\).

First suppose \(x \notin \overline{A}\). We know \(\overline{A}\) is closed. Thus there is a \(\delta > 0\) such that \((B(x,\delta) \ subset \overline{A}^c)\). As \(A \ subset \overline{A}\) we see that \((B(x,\delta) \ subset A^c)\) and hence \((B(x,\delta) \ cap A) = \emptyset\).

On the other hand suppose there is a \(\delta > 0\) such that \((B(x,\delta) \ cap A) = \emptyset\). Then \((B(x,\delta) \ cap A) = \emptyset\) is a closed set and we have that \((A \ subset \overline{B(x,\delta)} \ cap A)\), but \(x \notin \overline{B(x,\delta)} \ cap A\). Thus as \(\overline{A}\) is the intersection of closed sets containing \(\overline{A}\), we have \(x \notin \overline{A}\).

We can also talk about what is in the interior of a set and what is on the boundary.

Let \((X,d)\) be a metric space and \(A subset X\), then the interior of \(A\) is the set \(A^{\circ} := \{ x in A : there exists a \delta > 0 \ such that B(x,\delta) subset A \} \subset X\). The boundary of \(A\) is the set \(\partial A := \overline{A} \ setminus A^{\circ}\).

Suppose \(A=(0,1]\) and \(X = \mathbb{R}\). Then it is not hard to see that \(\overline{A}=[0,1]\), \(A^{\circ} = (0,1)\), and \(\partial A = \{ 0, 1 \}\).

Suppose \(X = \{ a, b \}\) with the discrete metric. Let \(A = \{ a \}\), then \(\overline{A} = A^{\circ} = A\) and \(\partial A = \emptyset\).

Let \((X,d)\) be a metric space and \(A subset X\). Then \(A^{\circ}\) is open and \(\partial A\) is closed.

Given \(x in A^{\circ}\) we have \(\delta > 0\) such that \((B(x,\delta) \ subset A)\). If \(z in B(x,\delta)\), then as open balls are open, there is an \(\epsilon > 0\) such that \((B(z,\epsilon) \ subset B(x,\delta) \ subset A)\), so \((z)\) is in \(A^{\circ}\). Therefore \((B(x,\delta) \ subset A^{\circ})\) and so \((A^{\circ})\) is open.

As \(A^{\circ}\) is open, then \(\partial A = \overline{A} \ setminus A^{\circ} = \overline{A} \ cap \overline{(A^{\circ})^c}\) is closed.
The boundary is the set of points that are close to both the set and its complement.

Let \( (X,d) \) be a metric space and \( (A \subset X) \). Then \( x \in \partial A \) if and only if for every \( \delta > 0 \), \( B(x,\delta) \cap A \) and \( B(x,\delta) \cap A^c \) are both nonempty.

Suppose \( x \in \partial A = \overline{A} \setminus A^\circ \) and let \( \delta > 0 \) be arbitrary. By , \( B(x,\delta) \) contains a point from \( \overline{A} \). If \( B(x,\delta) \) contained no points of \( A^\circ \), then \( x \) would be in \( A^\circ \). Hence \( B(x,\delta) \) contains a point of \( A^c \) as well.

Let us prove the other direction by contrapositive. If \( x \notin \partial A = \overline{A} \setminus A^\circ \), then there is some \( \delta > 0 \) such that \( B(x,\delta) \subset \overline{A}^c \) as \( \overline{A} \) is closed. So \( B(x,\delta) \) contains no points of \( A^\circ \), because \( \overline{A}^c \subset A^c \).

Now suppose \( x \in A^\circ \), then there exists a \( \delta > 0 \) such that \( B(x,\delta) \subset A \), but that means \( B(x,\delta) \) contains no points of \( A^\circ \).

We obtain the following immediate corollary about closures of \( (A) \) and \( (A^c) \). We simply apply .

Let \( (X,d) \) be a metric space and \( (A \subset X) \). Then \( \partial A = \overline{A} \cap \overline{A^c} \).

**Exercises**

Prove . Hint: consider the complements of the sets and apply .

Finish the proof of by proving that \( \overline{C(x,\delta)} \) is closed.

Prove .

Suppose \( (X,d) \) is a nonempty metric space with the discrete topology. Show that \( \overline{X} \) is connected if and only if it contains exactly one element.

Show that if \( S \subset \{\mathbb{R}\} \) is a connected unbounded set, then it is an (unbounded) interval.

Show that every open set can be written as a union of closed sets.

a) Show that \( \overline{E} \) is closed if and only if \( \partial E \subset E \). b) Show that \( \overline{U} \) is open if and only if \( \partial U \cap U = \emptyset \).

a) Show that \( \overline{A} \) is open if and only if \( A^\circ = A \). b) Suppose that \( \overline{U} \) is an open set and \( U \subset A \). Show that \( U \subset A^\circ \).

Let \( (X) \) be a set and \( (d), (d') \) be two metrics on \( (X) \). Suppose there exists an \( \alpha > 0 \) and \( \beta > 0 \) such that \( \alpha d(x,y) \leq d'(x,y) \leq \beta d(x,y) \) for all \( x,y \in X \). Show that \( \overline{U} \) is open in \( (X,d) \) if and only if \( \overline{U} \) is open in \( (X,d') \). That is, the topologies of \( (X,d) \) and \( (X,d') \) are the same.
Suppose \( \{ S_i \} \), \( i \in \mathbb{N} \) is a collection of connected subsets of a metric space \( (X,d) \). Suppose there exists an \( x \in X \) such that \( x \in S_i \) for all \( i \in \mathbb{N} \). Show that \( \bigcup_{i=1}^{\infty} S_i \) is connected.

Let \( \{ A \} \) be a connected set. a) Is \( \overline{A} \) connected? Prove or find a counterexample. b) Is \( A^\circ \) connected? Prove or find a counterexample. Hint: Think of sets in \( \{ \text{mathbb(R)} \}^{\geq 2} \).

The definition of open sets in the following exercise is usually called the subpace topology. You are asked to show that we obtain the same topology by considering the subspace metric.

[exercise:mssubspace] Suppose \( (X,d) \) is a metric space and \( \{ Y \in \text{subset X} \} \). Show that with the subspace metric on \( \{ Y \} \), a set \( \{ U \in \text{subset Y} \} \) is open (in \( \{ Y \} \)) whenever there exists an open set \( \{ V \in \text{subset X} \} \) such that \( \{ U = V \text{ cap Y} \} \).

Let \( (X,d) \) be a metric space. a) For any \( x \in X \) and \( \{ \text{delta} > 0 \} \), show \( \overline{\text{B}(x,\text{delta})} \in \text{subset C(x,\text{delta})} \). b) Is it always true that \( \overline{\text{B}(x,\text{delta})} = C(x,\text{delta}) \)? Prove or find a counterexample.

Let \( (X,d) \) be a metric space and \( \{ A \in \text{subset X} \} \). Show that \( \{ A^\circ = \bigcup \{ V : V \in \text{subset A text{ is open} \} \} \} \).

**Sequences and convergence**

**Note:** 1 lecture

**Sequences**

The notion of a sequence in a metric space is very similar to a sequence of real numbers. The definitions are essentially the same as those for real numbers in the sense of where \( \{ \text{mathbb(R)} \} \) with the standard metric \( d(x,y)=\lvert x-y \rvert \) is replaced by an arbitrary metric space \( ((X,d)) \).

A sequence in a metric space \( (X,d) \) is a function \( x : \mathbb{N} \to X \). As before we write \( x_n \) for the \( n \)th element in the sequence and use the notation \( n \{ x_n \} \), or more precisely \( n \{ x_n \}_{n=1}^{\infty} \). A sequence \( \{ x_n \} \) is bounded if there exists a point \( p \in \text{subset X} \) and \( B \in \mathbb{R} \) such that \( \{ d(p,x_n) \leq B \text{ for all } n \in \text{subset \{ x_n \}} \} \). In other words, the sequence \( \{ x_n \} \) is bounded whenever the set \( \{ n : n \in \text{subset \{ x_n \}} \} \) is bounded.

If \( \{ n_j \}_{j=1}^{\infty} \) is a sequence of natural numbers such that \( n_{j+1} > n_j \) for all \( j \), then the sequence \( \{ x_{n_j} \}_{j=1}^{\infty} \) is said to be a subsequence of \( \{ x_n \} \).

Similarly we also define convergence. Again, we will be cheating a little bit and we will use the definite article in front of the word limit before we prove that the limit is unique.

A sequence \( \{ x_n \} \) in a metric space \( (X,d) \) is said to converge to a point \( p \in \text{subset X} \), if for every \( \epsilon > 0 \), there exists an \( M \in \text{subset \{ x_n \}} \) such that \( d(x_n,p) < \epsilon \) for all \( n \geq M \). The point \( p \) is said to be the limit of \( \{ x_n \} \). We write \( \lim_{n \to \infty} x_n := p \).
A sequence that converges is said to be *convergent*. Otherwise, the sequence is said to be *divergent*.

Let us prove that the limit is unique. Note that the proof is almost identical to the proof of the same fact for sequences of real numbers. Many results we know for sequences of real numbers can be proved in the more general settings of metric spaces. We must replace \(|x-y|\) with \(d(x,y)\) in the proofs and apply the triangle inequality correctly.

[prop:mslimisunique] A convergent sequence in a metric space has a unique limit.

Suppose the sequence \(\{x_n\}\) has the limit \(x\) and the limit \(y\). Take an arbitrary \(\epsilon > 0\). From the definition find an \(M_1\) such that for all \(n \geq M_1\), \(d(x_n, x) < \frac{\epsilon}{2}\). Similarly find an \(M_2\) such that for all \(n \geq M_2\) we have \(d(x_n, y) < \frac{\epsilon}{2}\). Now take an \(n\) such that \(n \geq M_1\) and also \(n \geq M_2\).

\[
\begin{split}
d(y,x) & \leq d(y,x_n) + d(x_n,x) \\
& < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{split}
\]

As \(d(y,x) < \epsilon\) for all \(\epsilon > 0\), then \(d(y,x) = 0\) and \(y = x\). Hence the limit (if it exists) is unique.

The proofs of the following propositions are left as exercises.

[prop:msconvbound] A convergent sequence in a metric space is bounded.

[prop:msconvifa] A sequence \(\{x_n\}\) in a metric space \((X,d)\) converges to \((p \in X)\) if and only if there exists a sequence \(\{a_n\}\) of real numbers such that \(d(x_n,p) \leq a_n\) for all \(n \in \mathbb{N}\) and \(\lim_{n \to \infty} a_n = 0\).

[prop:mssubseq] Let \(\{x_n\}\) be a sequence in a metric space \((X,d)\).

i. If \(\{x_n\}\) converges to \((p \in X)\), then every subsequence \(\{x_{n_k}\}\) converges to \((p)\).

ii. If for some \((K \in \mathbb{N})\) the \(K\)-tail \(\{x_n\}_{n=K+1}^{\infty}\) converges to \((p \in X)\), then \(\{x_n\}\) converges to \((p)\).

**Convergence in euclidean space**

It is useful to note what convergence means in the euclidean space \(\{\text{mathbb}\{\mathbb{R}\}^{\times n}\}\).

[prop:msconveuc] Let \(\{x_j\}_{j=1}^{\infty}\) be a sequence in \(\{\text{mathbb}\{\mathbb{R}\}^{\times n}\}\), where we write \(x_j = (x_{j,1}, x_{j,2}, \ldots, x_{j,n}) \in \{\text{mathbb}\{\mathbb{R}\}^{\times n}\}\). Then \(\{x_j\}_{j=1}^{\infty}\) converges if and only if \(\{x_{j,1}\}_{j=1}^{\infty}, \{x_{j,2}\}_{j=1}^{\infty}, \ldots, \{x_{j,n}\}_{j=1}^{\infty}\) converges for every \(j\), in which case \(\lim_{j \to \infty} x_j = (\lim_{j \to \infty} x_{j,1}, \ldots, \lim_{j \to \infty} x_{j,n})\).

Let \(\{x_j\}_{j=1}^{\infty}\) be a convergent sequence in \(\{\text{mathbb}\{\mathbb{R}\}^{\times n}\}\), where we write \(x_j = (x_{j,1}, x_{j,2}, \ldots, x_{j,n}) \in \{\text{mathbb}\{\mathbb{R}\}^{\times n}\}\). Let \(y = (y_1, y_2, \ldots, y_n)\) in \(\{\text{mathbb}\{\mathbb{R}\}^{\times n}\}\) be the limit. Given \(\epsilon > 0\), there exists an \(M\) such that for all \(j \geq M\) we have \(d(y,x_j) < \epsilon\). Fix some \(k\). For \(j \geq M\) we have \(d(y - x_j, k) \leq d(y, k)\) \(\forall k\)
\[ \sqrt{\sum_{\ell=1}^{n} \bigl(y_{\ell}-x_{j,\ell}\bigr)^2} = d(y,x_j) < \epsilon. \] Hence the sequence \( \{ x_{j,k} \}_{j=1}^{\infty} \) converges to \( y_k \).

For the other direction suppose \( \{ x_{j,k} \}_{j=1}^{\infty} \) converges to \( y_k \) for every \( k=1,2,\ldots,n \). Hence, given \( \epsilon > 0 \), pick an \( (M,k) \), such that if \( j \geq M \), then \( \bigl| y_k-x_{j,k}\bigr| < \epsilon \). Then \[ d(y,x_j) = \sqrt{\sum_{k=1}^{n} \bigl(y_k-x_{j,k}\bigr)^2} < \sqrt{\sum_{k=1}^{n} \bigl| y_k-x_{j,k}\bigr|^2} < \frac{\epsilon}{\sqrt{n}} \] (since \( \sum_{k=1}^{n} \bigl| y_k-x_{j,k}\bigr|^2 \) is the sum of \( n \) squares, each less than \( \frac{\epsilon^2}{n} \)).

The sequence \( \{ x_j \} \) converges to \( y \in \mathbb{R}^n \) and we are done.

Convergence and topology

The topology, that is, the set of open sets of a space encodes which sequences converge.

[prop:msconvtoto] Let \((X,d)\) be a metric space and \( \{ x_{n} \} \) a sequence in \( X \). Then \( \{ x_{n} \} \) converges to \( x \in X \) if and only if for every open neighborhood \( U \) of \( x \), there exists an \( M \in \mathbb{N} \) such that for all \( n \geq M \) we have \( x_{n} \in U \).

First suppose \( \{ x_{n} \} \) converges. Let \( U \) be an open neighborhood of \( x \), then there exists an \( \epsilon > 0 \) such that \( \{ B(x,\epsilon) \} \) \subset U. As the sequence converges, find an \( M \) such that for all \( n \geq M \) we have \( d(x,x_{n}) < \epsilon \) or in other words \( x_{n} \in U \).

Let us prove the other direction. Given \( \epsilon > 0 \) let \( U := B(x,\epsilon) \) be the neighborhood of \( x \). Then there is an \( M \) such that for \( n \geq M \) we have \( x_{n} \in U = B(x,\epsilon) \) or in other words, \( d(x,x_{n}) < \epsilon \).

A set is closed when it contains the limits of its convergent sequences.

[prop:msclosedlim] Let \((X,d)\) be a metric space, \( \{ x_{n} \} \) a sequence in \( X \) that converges to \( x \). Then \( x \in \overline{X} \).

Let us prove the contrapositive. Suppose \( \{ x_{n} \} \) is a sequence in \( X \) that converges to \( x \). As \( \{ E \} \) is open, says there is an \( (M,k) \) such that for all \( x_{n} \in E \), \( x_{n} \in E \). So \( \{ x_{n} \} \) is not a sequence in \( E \).

When we take a closure of a set \( A \), we really throw in precisely those points that are limits of sequences in \( \{ x \} \).

[prop:msclosureaprseq] Let \((X,d)\) be a metric space and \( A \) a subset of \( X \). Then \( x \in \overline{A} \) if and only if there exists a sequence \( \{ x_{n} \} \) of elements of \( A \) such that \( \lim_{n \to \infty} x_{n} = x \).
Let \( x \in \overline{A} \). We know by that given \( \frac{1}{n} \), there exists a point \( x_n \in B(x, \frac{1}{n}) \cap A \). As \( d(x, x_n) < \frac{1}{n} \), we have \( \lim_{n \to \infty} x_n = x \).

For the other direction, see .

**Exercises**

[exercise:reverseclosedseq] Let \((X,d)\) be a metric space and let \(A \subseteq X\). Let \(E\) be the set of all \( x \in X \) such that there exists a sequence \( \{ x_n \} \) in \((A)\) that converges to \(x\). Show \( E = \overline{A} \).

a) Show that \( d(x,y) := \min \{ 1, |x-y| \} \) defines a metric on \((\mathbb{R},d)\). b) Show that a sequence converges in \((\mathbb{R},d)\) if and only if it converges in the standard metric. c) Find a bounded sequence in \((\mathbb{R},d)\) that contains no convergent subsequence.

Prove .

Prove .

Suppose \( \{ x_n \} \) converges to \( x \). Suppose \( \{ f(n) \} \) is a one-to-one function. Show that \( \{ x_{f(n)} \} \) converges to \( x \).

If \((X,d)\) is a metric space where \(d\) is the discrete metric. Suppose \( \{ x_n \} \) is a convergent sequence in \((X)\). Show that there exists a \( K \in \{ x_n \} \) such that for all \( n \geq K \) we have \( x_n = x_K \).

A set \( S \subseteq X \) is said to be dense in \( X \) if for every \( x \in X \), there exists a sequence \( \{ x_n \} \) in \( S \) that converges to \( x \). Prove that \( \{ x_{f(n)} \} \) converges to \( x \).

Suppose \( \{ U_n \} \) is a decreasing \( (U_{n+1} \subseteq U_n) \) sequence of open sets in a metric space \((X,d)\) such that \( \bigcap_{n=1}^\infty U_n = \{ p \} \). Suppose \( \{ x_n \} \) is a sequence of points in \((X)\) such that \( x_n \in U_n \). Does \( \{ x_n \} \) necessarily converge to \( p \)? Prove or construct a counterexample.

Let \( E \subseteq X \) be closed and let \( \{ x_n \} \) be a sequence in \((X)\) converging to \( p \in X \). Suppose \( x_n \in E \) for infinitely many \( n \in \mathbb{N} \). Show \( p \in E \).

Take \( \{ x_n \} \) to be the extended reals. Define \( d(x,y) := |\frac{x}{1+|x|} - \frac{y}{1+|y|}| \) for all \( x, y \). Let \( \{ x_n \} \) be a sequence of real numbers such that for every \( M \in \mathbb{R} \), there exists an \( N \) such that \( \{ x_n \} \) for all \( n \geq N \). Show that \( \lim_{n \to \infty} x_n = \infty \) in \((\mathbb{R},d)\). c) Show that a sequence of real numbers converges to a real number in \((\mathbb{R},d)\) if and only if it converges in \((\mathbb{R},d)\) with the standard metric.

UC Davis ChemWiki is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.
Suppose \( \{ V_n \}_{n=1}^{\infty} \) is a collection of open sets in \((X,d)\) such that \( V_{n+1} \supset V_n \). Let \( \{ x_n \} \) be a sequence such that \( x_n \in V_{n+1} \setminus V_n \) and suppose \( \{ x_n \} \) converges to \( p \in X \). Show that \( p \in \partial V \) where \( V = \bigcup_{n=1}^{\infty} V_n \).

Prove.

**Completeness and compactness**

*Note: 2 lectures*

**Cauchy sequences and completeness**

Just like with sequences of real numbers we define Cauchy sequences.

Let \((X,d)\) be a metric space. A sequence \( \{ x_n \} \) in \( X \) is a Cauchy sequence if for every \( \epsilon > 0 \) there exists an \( M \) such that for all \( n, k \geq M \) we have \( d(x_n, x_k) < \epsilon \).

The definition is again simply a translation of the concept from the real numbers to metric spaces. So a sequence of real numbers is Cauchy in the sense of if and only if it is Cauchy in the sense above, provided we equip the real numbers with the standard metric \( d(x,y) = |x-y| \).

A convergent sequence in a metric space is Cauchy.

Suppose \( \{ x_n \} \) converges to \( x \). Given \( \epsilon > 0 \) there exists an \( M \) such that for all \( n, k \geq M \) we have \( d(x_n, x_k) < \epsilon \). Hence for all \( n, k \geq M \) we have \( d(x_n, x_k) \leq d(x_n, x) + d(x, x_k) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \).

Let \((X,d)\) be a metric space. We say \( X \) is complete or Cauchy-complete if every Cauchy sequence \( \{ x_n \} \) in \( X \) converges to an \( x \in X \).

The space \( \{ \mathbb{R} \}_{n=1}^{\infty} \) with the standard metric is a complete metric space.

For \( \{ \mathbb{R} \}_{n=1}^{\infty} \) completeness was proved in . The proof of the above proposition is a reduction to the one dimensional case.

Let \( \{ x_j \}_{j=1}^{\infty} \) be a Cauchy sequence in \( \{ \mathbb{R} \}_{n=1}^{\infty} \), where we write \( x_j = \{ x_{j,1}, x_{j,2}, \ldots, x_{j,n} \} \) in \( \{ \mathbb{R} \}_{n=1}^{\infty} \). As the sequence is Cauchy, given \( \epsilon > 0 \), there exists an \( M \) such that for all \( i, j \geq M \) we have \( d(x_i, x_j) < \epsilon \).

Fix some \( k = 1, 2, \ldots, n \), for \( i, j \geq M \) we have \( \sqrt{\sum_{\ell=1}^{n} (x_{i,\ell} - x_{j,\ell})^2} = d(x_i, x_j) < \epsilon \). Hence the sequence \( \{ x_{j,k} \}_{j=1}^{\infty} \) is Cauchy. As \( \{ \mathbb{R} \}_{n=1}^{\infty} \) is complete the sequence converges; there exists an \( y_k \in \{ \mathbb{R} \}_{n=1}^{\infty} \) such that \( y_k = \lim_{j \to \infty} x_{j,k} \).
Write \( y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n \). By we have that \( \langle x_j \rangle \) converges to \( y \in \mathbb{R}^n \) and hence \( \mathbb{R}^n \) is complete.

Note that a subset of \( \mathbb{R}^n \) with the subspace metric need not be complete. For example, \( (0,1) \) with the subspace metric is not complete as \( \langle \frac{1}{n} \rangle \) is a Cauchy sequence in \( (0,1) \) with no limit in \( (0,1) \). But see also .

**Compactness**

Let \( (X,d) \) be a metric space and \( K \subset X \). The set \( K \) is said to be **compact** if for any collection of open sets \( \{ U_{\lambda} \}_{\lambda \in I} \) such that \( K \subset \bigcup_{\lambda \in I} U_{\lambda} \), there exists a finite subset \( \{ \lambda_1, \lambda_2, \ldots, \lambda_k \} \subset I \) such that \( K \subset \bigcup_{j=1}^k U_{\lambda_j} \).

A collection of open sets \( \{ U_{\lambda} \}_{\lambda \in I} \) as above is said to be an **open cover** of \( K \). So a way to say that \( K \) is compact is to say that every open cover of \( K \) has a finite subcover.

Let \( (X,d) \) be a metric space. A compact set \( K \subset X \) is closed and bounded. First, we prove that a compact set is bounded. Fix \( p \in X \). We have the open cover \( K \subset \bigcup_{n=1}^{\infty} B(p,n) = X \). If \( K \) is compact, then there exists some set of indices \( \{ n_1 < n_2 < \ldots < n_k \} \) such that \( K \subset \bigcup_{j=1}^k B(p,n_j) = B(p,n_k) \). As \( K \) is contained in a ball, \( K \) is bounded.

Next, we show a set that is not closed is not compact. Suppose \( \overline{K} \neq K \), that is, there is a point \( x \in \overline{K} \) \( \{ \text{not in K} \} \). If \( \{ y \{ \text{not in K} \} \} \), then for \( \{ n \} \) with \( \{ \text{nicefrac} \{ 1 \} \{ n \} \} < d(x,y) \) we have \( \{ y \{ \text{not in K} \} \} \). Furthermore \( \{ x \{ \text{not in K} \} \} \), so \( K \subset \bigcup_{n=1}^{\infty} C(x, \text{nicefrac} \{ 1 \} \{ n \}) \). As a closed ball is closed, \( \{ C(x, \text{nicefrac} \{ 1 \} \{ n \}) \} \) is open, and so we have an open cover. If we take any finite collection of indices \( \{ n_1 < n_2 < \ldots < n_k \} \), then \( \bigcup_{j=1}^k C(x, \text{nicefrac} \{ 1 \} \{ n_j \}) = C(x, \text{nicefrac} \{ 1 \} \{ n_k \}) \). As \( x \) is in the closure, \( C(x, \text{nicefrac} \{ 1 \} \{ n_k \}) \) is not compact. Therefore, there is no finite subcover and \( K \) is not compact.

We prove below that in finite dimensional euclidean space every closed bounded set is compact. So closed bounded sets of \( \mathbb{R}^n \) are examples of compact sets. It is not true that in every metric space, closed and bounded is equivalent to compact. A simple example would be an incomplete metric space such as \( (0,1) \) with the subspace metric. But there are many complete and very useful metric spaces where closed and bounded is not enough to give compactness, see : \( (C([a,b], \mathbb{R})) \) is a complete metric space, but the closed unit ball \( C(0,1) \) is not compact. However, see .

A useful property of compact sets in a metric space is that every sequence has a convergent subsequence. Such sets are sometimes called **sequentially compact**. Let us prove that in the context of metric spaces, a set is compact if and only if it is sequentially compact. First we prove a lemma.

---

**Lemma:** Let \( (X,d) \) be a metric space and \( K \subset X \). Suppose every sequence in \( K \) has a subsequence convergent in \( K \). Given an open cover \( \{ \bigcup_{\lambda \in I} U_{\lambda} \} \) of \( K \), there exists a \( \{ \delta > 0 \} \) such that for every \( x \in K \), there exists a \( \{ \lambda \in I \} \) with \( B(x, \delta) \subset U_{\lambda} \).

---
It is important to recognize what the lemma says. It says that given any cover there is a single \((\delta > 0)\). The \((\delta)\) can depend on the cover, but of course it does not depend on \((x)\).

Let us prove the lemma by contrapositive. If the conclusion is not true, then there is an open cover \(\{ U_\lambda \}_{\lambda \in I} \) of \((K)\) with the following property. For every \(n \in \mathbb{N}\) there exists an \((x_n \in K)\) such that \((B(x_n, \nicefrac{1}{n}))\) is not a subset of any \((U_\lambda)\). Given any \((x \in K)\), there is a \((\lambda \in I)\) such that \((x \in U_\lambda)\). Hence there is an \((\epsilon > 0)\) such that \((B(x, \epsilon) \subseteq U_\lambda)\). Take \((M)\) such that \((\nicefrac{1}{M} < \epsilon)\). If \((y \in B(x, \nicefrac{1}{M}))\), then by triangle inequality \((B(y, \nicefrac{1}{M}) \subseteq B(x, \epsilon))\). Hence the sequence cannot have a subsequence converging to \((x)\). As \((x \in K)\) was arbitrary we are done.

Let \((K \subseteq X)\) be a set and \(\{ x_n \}\) a sequence in \((K)\). Suppose that for each \((x \in K)\), there is a ball \((B(x, \alpha_x))\) for some \((\alpha_x > 0)\) such that \((x_n \in B(x, \alpha_x))\) for only finitely many \((n \in \mathbb{N})\). Then \((K \subseteq \bigcup_{x \in K} B(x, \alpha_x))\). Any finite collection of these balls is going to contain only finitely many \((x_n)\). Thus for any finite collection of such balls there is an \((x \in K)\) that is not in the union. Therefore, \((K)\) is not compact.

For the other direction, suppose every sequence in \((K)\) has a subsequence converging in \((K)\). Take an open cover \(\{ U_\lambda \}_{\lambda \in I} \) of \((K)\). Using the Lebesgue covering lemma above, we find a \((\delta > 0)\) such that for every \((x)\), there is a \((\lambda \in I)\) with \((B(x, \delta) \subseteq U_\lambda)\).

Pick \((x_1 \in K)\) and find \((\lambda_1 \in I)\) such that \((B(x_1, \delta) \subseteq U_{\lambda_1})\). If \((K \subseteq \lambda_1 \subseteq \lambda_2 \subseteq \cdots \subseteq \lambda)\), we stop as we have found a finite subcover. Otherwise, there must be a point \((x_2 \in K \setminus \lambda)\). Note that \((d(x_2, x_1) \geq \delta)\). There must exist some \((\lambda_2 \in I)\) such that \((B(x_2, \delta) \subseteq U_{\lambda_2})\). We work inductively. Suppose \((\lambda_{n-1})\) is defined. Either \((U_{\lambda_{n-1}} \cup B(x_n, \delta) \subseteq U_{\lambda_n})\) in which case we stop, or there must be a point \((x_n \in K \setminus \lambda)\). Note that \((d(x_n, x_{n-j}) \geq \delta)\). Next, there must be some \((\lambda_n \in I)\) such that \((B(x_n, \delta) \subseteq U_{\lambda_n})\).

Either at some point we obtain a finite subcover of \((K)\) or we obtain an infinite sequence \(\{ x_n \}\) as above. For contradiction suppose that there is no finite subcover and we have the sequence \(\{ x_n \}\). For all \((n)\) and \((\lambda)\), \((n \not= \lambda)\).
k\), we have \(d(x_n, x_k) \geq \delta\), so no subsequence of \(\{x_n\}\) can be Cauchy. Hence no subsequence of \(\{x_n\}\) can be convergent, which is a contradiction.

The Bolzano-Weierstrass theorem for sequences of real numbers () says that any bounded sequence in \(\mathbb{R}\) has a convergent subsequence. Therefore any sequence in a closed interval \([a, b] \subset \mathbb{R}\) has a convergent subsequence. The limit must also be in \([a, b]\) as limits preserve non-strict inequalities. Hence a closed bounded interval \([a, b] \subset \mathbb{R}\) is compact.

Let \((X, d)\) be a metric space and let \((K \setminus X)\) be compact. If \((E \setminus K)\) is a closed set, then \((E)\) is compact.

Let \(\{x_n\}\) be a sequence in \((E)\). It is also a sequence in \((K)\). Therefore it has a convergent subsequence \(\{x_{n_j}\}\) that converges to some \(\lim x \in K\). As \((E)\) is closed, the limit of a sequence in \((E)\) is also in \((E)\) and so \(\lim x \in E\). Thus \((E)\) must be compact.

[thm:msbw] A closed bounded subset \((K \setminus \{\text{mathbb}\{R\}\}^\text{n})\) is compact.

So subsets of \((\text{mathbb}\{R\})^\text{n}\) are compact if and only if they are closed and bounded, a condition that is much easier to check. Let us reiterate that the Heine-Borel theorem only holds for \((\text{mathbb}\{R\})^\text{n}\) and not for metric spaces in general. In general, compact implies closed and bounded, but not vice versa.

For \((\text{mathbb}\{R\}) = \{\text{mathbb}\{R\}^\text{n}\}\) if \((K \setminus \{\text{mathbb}\{R\}\}^\text{n})\) is closed and bounded, then any sequence \(\{x_k\}\) in \((K)\) is bounded, so it has a convergent subsequence by Bolzano-Weierstrass theorem (). As \((K)\) is closed, the limit of the subsequence must be an element of \((K)\). So \((K)\) is compact.

Let us carry out the proof for \((n=2)\) and leave arbitrary \((n)\) as an exercise. As \((K \setminus \{\text{mathbb}\{R\}\}^2)\) is bounded, there exists a set \((B=[a,b] \times [c,d] \subset \{\text{mathbb}\{R\}\}^2)\) such that \((K \setminus B)\). We will show that \((B)\) is compact. Then \((K)\), being a closed subset of a compact \((B)\), is also compact.

Let \(\{x_k, y_k\}_{k=1}^\infty\) be a sequence in \((B)\). That is, \((a \leq x_k \leq b)\) and \((c \leq y_k \leq d)\) for all \((k)\). A bounded sequence of real numbers has a convergent subsequence so there is a subsequence \(\{x_{k_j}\}_{j=1}^\infty\) that is convergent. The subsequence \(\{y_{k_j}\}_{j=1}^\infty\) is also a bounded sequence so there exists a subsequence \(\{y_{k_{i_j}}\}_{i=1}^\infty\) that is convergent. A subsequence of a convergent sequence is still convergent, so \(\{x_{k_{i_j}}\}_{i=1}^\infty\) is convergent. Let \(\lim x := \lim_{i \to \infty} x_{k_{i_j}}\) and \(\lim y := \lim_{i \to \infty} y_{k_{i_j}}\). By \(\lim_{i \to \infty} (x_{k_{i_j}}) \equiv (y_{k_{i_j}})\) for all \((k)\), we know that \((x, y) \in B)\).

The discrete metric provides interesting counterexamples again. Let \((X, d)\) be a metric space with the discrete metric, that is \((d(x, y) = 1)\) if \((x \not= y)\). Suppose \((X)\) is an infinite set. Then:

i. \((X, d)\) is a complete metric space.

ii. Any subset \((K \setminus X)\) is closed and bounded.

iii. A subset \((K \setminus X)\) is compact if and only if it is a finite set.
iv. The conclusion of the Lebesgue covering lemma is always satisfied with e.g. \( \delta = \frac{1}{2} \), even for noncompact \( K \subset X \).

The proofs of these statements are either trivial or are relegated to the exercises below.

**Exercises**

Let \((X,d)\) be a metric space and \( \{A\} \) a finite subset of \( X \). Show that \( \{A\} \) is compact.

Let \( A = \{ \frac{1}{n} : n \in \mathbb{N} \} \subset \mathbb{R} \). a) Show that \( \{A\} \) is not compact directly using the definition. b) Show that \( \{A \cup \{0\}\} \) is compact directly using the definition.

Let \((X,d)\) be a metric space with the discrete metric. a) Prove that \( \{X\} \) is complete. b) Prove that \( \{X\} \) is compact if and only if \( \{X\} \) is a finite set.

a) Show that the union of finitely many compact sets is a compact set. b) Find an example where the union of infinitely many compact sets is not compact.

Prove for arbitrary dimension. Hint: The trick is to use the correct notation.

Show that a compact set \( \{K\} \) is a complete metric space (using the subspace metric).

[exercise:CabRcomplete] Let \( \{C([a,b],\mathbb{R})\} \) be the metric space as in . Show that \( \{C([a,b],\mathbb{R})\} \) is a complete metric space.

[exercise:msclbounnotcompt] Let \( \{C([0,1],\mathbb{R})\} \) be the metric space of . Let \( \{0\} \) denote the zero function. Then show that the closed ball \( \{C([0,1])\} \) is not compact (even though it is closed and bounded). Hints: Construct a sequence of distinct continuous functions \( \{f_n\} \) such that \( d(f_n,0) = 1 \) and \( d(f_n,f_k) = 1 \) for all \( n \neq k \). Show that the set \( \{f_n : n \in \mathbb{N}\} \subset C(0,1) \) is closed but not compact. See for inspiration.

Show that there exists a metric on \( \{\mathbb{R}\} \) that makes \( \{\mathbb{R}\} \) into a compact set.

Suppose \((X,d)\) is complete and suppose we have a countably infinite collection of nonempty compact sets \( \{E_1 \cup \ldots \cup E_n : n \in \mathbb{N}\} \) then prove \( \{E_1 \cup \ldots \cup E_n : n \in \mathbb{N}\} \) is compact.

[exercise:CabRcomplete] Let \( \{C([0,1],\mathbb{R})\} \) be the metric space of . Let \( K \) be the set of \( f \in C([0,1],\mathbb{R}) \) such that \( f \) is a quadratic polynomial, i.e. \( f(x) = ax^2 + bx + c \), and such that \( \left\| f(x) \right\|_{\infty} \leq 1 \) for all \( x \in [0,1] \), that is \( f \in C([0,1]) \). Show that \( K \) is compact.

[exercise:msclbounnotcompt] Let \((X,d)\) be a complete metric space. Show that \( \{K \subset X\} \) is compact if and only if \( \{K\} \) is closed and such that for every \( \epsilon > 0 \) there exists a finite set of points \( \{x_1,x_2,\ldots,x_n\} \) with \( \{K \cap \operatorname{B}(x_j,\epsilon) : j = 1,\ldots,n\} \) closed and such that for every \( \epsilon > 0 \) there exists a finite set of points \( \{x_1,x_2,\ldots,x_n\} \) with \( \{K \cap \operatorname{B}(x_j,\epsilon) : j = 1,\ldots,n\} \) closed and such that for every \( \epsilon > 0 \) there exists a finite set of points \( \{x_1,x_2,\ldots,x_n\} \) with \( \{K \cap \operatorname{B}(x_j,\epsilon) : j = 1,\ldots,n\} \). Note: Such a set \( \{K\} \) is said to be totally bounded, so in a complete metric space a set is compact if and only if it is closed and totally bounded.
Take $\{\mathbb{N}\} \subset \{\mathbb{R}\}$ using the standard metric. Find an open cover of $\{\mathbb{N}\}$ such that the conclusion of the Lebesgue covering lemma does not hold.

Prove the general Bolzano-Weierstrass theorem: Any bounded sequence $\{x_k\}$ in $\{\mathbb{R}\}^n$ has a convergent subsequence.

Let $\langle X \rangle$ be a metric space and $\langle C \subset \{\text{mathcal{P}}\} \rangle$ the set of nonempty compact subsets of $\langle X \rangle$. Using the Hausdorff metric from , show that $\langle (C,d_H) \rangle$ is a metric space. That is, show that if $\langle L \rangle$ and $\langle K \rangle$ are nonempty compact subsets then $\langle d_H(L,K) = 0 \rangle$ if and only if $\langle L=K \rangle$.

[exercise:closedcomplete] Let $\langle (X,d) \rangle$ be a complete metric space and $\langle E \subset X \rangle$ a closed set. Show that $\langle E \rangle$ with the subspace metric is a complete metric space.

### Continuous functions

**Note: 1 lecture**

#### Continuity

Let $\langle (X,d_X) \rangle$ and $\langle (Y,d_Y) \rangle$ be metric spaces and $\langle c \in X \rangle$. Then $\langle f: X \to Y \rangle$ is *continuous at* $\langle c \in X \rangle$ if for every $\langle \epsilon > 0 \rangle$ there is a $\langle \delta > 0 \rangle$ such that whenever $\langle x \in X \rangle$ and $\langle d_X(x,c) < \delta \rangle$, then $\langle d_Y(f(x),f(c)) < \epsilon \rangle$.

When $\langle f: X \to Y \rangle$ is continuous at all $\langle c \in X \rangle$, then we simply say that $\langle f \rangle$ is a *continuous function*.

The definition agrees with the definition from when $\langle f \rangle$ is a real-valued function on the real line, if we take the standard metric on $\langle \{\mathbb{R}\} \rangle$.

[prop:contiscont] Let $\langle (X,d_X) \rangle$ and $\langle (Y,d_Y) \rangle$ be metric spaces. Then $\langle f: X \to Y \rangle$ is continuous at $\langle c \in X \rangle$ if and only if for every sequence $\langle \{ x_n \} \rangle$ in $\langle X \rangle$ converging to $\langle c \rangle$, the sequence $\langle \{ f(x_n) \} \rangle$ converges to $\langle f(c) \rangle$.

Suppose $\langle f \rangle$ is continuous at $\langle c \rangle$. Let $\langle \{ x_n \} \rangle$ be a sequence in $\langle X \rangle$ converging to $\langle c \rangle$. Given $\langle \epsilon > 0 \rangle$, there is a $\langle \delta > 0 \rangle$ such that whenever $\langle d_X(x,c) < \delta \rangle$, then $\langle d_Y(f(x),f(c)) < \epsilon \rangle$. Hence $\langle \{ f(x_n) \} \rangle$ converges to $\langle f(c) \rangle$.

On the other hand suppose $\langle f \rangle$ is not continuous at $\langle c \rangle$. Then there exists an $\langle \epsilon > 0 \rangle$, such that for every $\langle n \in \{\mathbb{N}\} \rangle$ there exists an $\langle x_n \in X \rangle$, with $\langle d_X(x_n,c) < \frac{1}{n} \rangle$ such that $\langle d_Y(f(x_n),f(c)) \geq \epsilon \rangle$. Then $\langle \{ x_n \} \rangle$ converges to $\langle c \rangle$, but $\langle \{ f(x_n) \} \rangle$ does not converge to $\langle f(c) \rangle$.

Suppose $\langle f: \{\mathbb{R}\} \to \{\mathbb{R}\} \rangle$ is a polynomial. That is, $\langle f(x,y) = \sum_{j=0}^d \sum_{k=0}^j a_{jk} x^j y^k \rangle$ for some $\langle d \in \{\mathbb{N}\} \rangle$ (the degree) and $\langle (a_{jk}) \in \{\mathbb{R}\} \rangle$. Then we claim $\langle f \rangle$ is continuous.
\((\{(x_n, y_n)\}_{n=1}^{\infty}\) be a sequence in \(\{(\mathbb{R}^2)^\times\}\) that converges to \((x, y) \in \{(\mathbb{R}^2)^\times\}\). We have proved that this means that \(\lim_{n \to \infty} x_n = x \) and \(\lim_{n \to \infty} y_n = y\). So by we have \(\lim_{n \to \infty} f(x_n, y_n) = \lim_{n \to \infty} \sum_{j=0}^{d} \sum_{k=0}^{d-j} a_{jk} x_n^j y_n^k = \sum_{j=0}^{d} \sum_{k=0}^{d-j} a_{jk} x^j y^k = f(x, y)\). So \(f\) is continuous at \((x, y)\), and as \((x, y)\) was arbitrary \(f\) is continuous everywhere. Similarly, a polynomial in \(n\) variables is continuous.

**Compactness and continuity**

Continuous maps do not map closed sets to closed sets. For example, \(f: (0,1) \to \mathbb{R}\) defined by \(f(x) := x\) takes the set \((0,1)\), which is closed in \((0,1)\), to the set \((0,1)\), which is not closed in \(\{(\mathbb{R}^2)^\times\}\). On the other hand continuous maps do preserve compact sets.

\[\text{[lemma:continuouscompact]}\]

Let \((X, d_X)\) and \((Y, d_Y)\) be metric spaces and \(f: X \to Y\) a continuous function. If \(K \subset X\) is a compact set, then \(f(K)\) is a compact set.

A sequence in \(f(K)\) can be written as \(\{f(x_n)\}_{n=1}^{\infty}\), where \(\{x_n\}_{n=1}^{\infty}\) is a sequence in \(K\). The set \(K\) is compact and therefore there is a subsequence \(\{x_{n_i}\}_{i=1}^{\infty}\) that converges to some \((x, y)\) in \(K\). By continuity, \(\lim_{i \to \infty} f(x_{n_i}) = f(x) \in f(K)\). So every sequence in \(f(K)\) has a subsequence convergent to a point in \(f(K)\), and \(f(K)\) is compact by .

As before, \(f: X \to \{(\mathbb{R}^2)^\times\}\) achieves an absolute minimum at \(c \in X\) if \(f(x) \geq f(c)\) for all \(x \in X\). On the other hand, \(f\) achieves an absolute maximum at \(c \in X\) if \(f(x) \leq f(c)\) for all \(x \in X\).

Let \((X, d)\) be a compact metric space and \(f: X \to \{(\mathbb{R}^2)^\times\}\) a continuous function. Then \(f\) is bounded and in fact \(f\) achieves an absolute minimum and an absolute maximum on \(X\).

As \(X\) is compact and \(f\) is continuous, we have that \(f(X) \subset \{(\mathbb{R}^2)^\times\}\) is compact. Hence \(f(X)\) is closed and bounded. In particular, \(\sup f(X) \in f(X)\) and \(\inf f(X) \in f(X)\), because both the sup and inf can be achieved by sequences in \(f(X)\) and \(f(X)\) is closed. Therefore there is some \((x, y)\) such that \(f(x) = \sup f(X)\) and some \((y, x)\) such that \(f(y) = \inf f(X)\).

**Continuity and topology**

Let us see how to define continuity in terms of the topology, that is, the open sets. We have already seen that topology determines which sequences converge, and so it is no wonder that the topology also determines continuity of functions.

\[\text{[lemma:mstopocontloc]}\]

Let \((X, d_X)\) and \((Y, d_Y)\) be metric spaces. A function \(f: X \to Y\) is continuous at \(c \in X\) if and only if for every open neighborhood \(U\) of \((f(c))\) in \((Y)\), the set \(f^{-1}(U)\) contains an open neighborhood \(\text{of}(c)\) in \(X\).

First suppose that \(f\) is continuous at \(c\). Let \(U\) be an open neighborhood of \((f(c))\) in \((Y)\), then \(B_{Y - \text{big}(f(c), \epsilon) \text{big}(\epsilon, \biggr) \subset U}\) for some \(\epsilon > 0\). By continuity of \(f\), there exists a \(\delta > 0\) such that
whenever \((x)\) is such that \(d_X(x,c) < \epsilon\), then \(d_Y(f(x),f(c)) < \epsilon\). In other words, \([B_X(c,\delta) \subset f^{-1}(B_Y(f(c),\epsilon))] \subset f^{-1}(U)\) and \(B_X(c,\epsilon)\) is an open neighborhood of \(c\).

For the other direction, let \((\epsilon > 0)\) be given. If \((f^{-1}(B_Y(f(c),\epsilon)))\) contains an open neighborhood \(W\) of \(c\), it contains a ball. That is, there is some \(\delta > 0\) such that \([B_X(c,\delta) \subset f^{-1}(B_Y(f(c),\epsilon))\) \] That means precisely that if \(d_X(x,c) < \delta\) then \(d_Y(f(x),f(c)) < \epsilon\), and so \(f\) is continuous at \(c\).

Let \((X,d_X)\) and \((Y,d_Y)\) be metric spaces. A function \(f : X \to Y\) is continuous if and only if for every open \(U \subset Y\), \(f^{-1}(U)\) is open in \(X\).

The proof follows from and is left as an exercise.

Let \(f : X \to Y\) be a continuous function. tells us that if \((E \subset Y)\) is closed, then \(f^{-1}(E = X \setminus \text{setminus f^{-1}(E^c)})\) is also closed. Therefore if we have a continuous function \((f : X \to \mathbb{R})\), then the zero set of \(\langle f \rangle\), that is, \(f^{-1}(0) = \{ x \in X : f(x) = 0 \}\), is closed. We have just proved the most basic result in algebraic geometry, the study of zero sets of polynomials.

Similarly the set where \(f\) is nonnegative, that is, \(f^{-1}([0,\infty)) = \{ x \in X : f(x) \geq 0 \}\) is closed. On the other hand the set where \(f\) is positive, \(f^{-1}((0,\infty)) = \{ x \in X : f(x) > 0 \}\) is open.

Uniform continuity

As for continuous functions on the real line, in the definition of continuity it is sometimes convenient to be able to pick one \(\delta\) for all points.

Let \((X,d_X)\) and \((Y,d_Y)\) be metric spaces. Then \(f : X \to Y\) is uniformly continuous if for every \(\epsilon > 0\) there is a \(\delta > 0\) such that whenever \((x,c \in X)\) and \(d_X(x,c) < \delta\), then \(d_Y(f(x),f(c)) < \epsilon\).

A uniformly continuous function is continuous, but not necessarily vice-versa as we have seen.

Let \((X,d_X)\) and \((Y,d_Y)\) be metric spaces. Suppose \(f : X \to Y\) is continuous and \((X)\) compact. Then \(f\) is uniformly continuous.

Let \(\epsilon > 0\) be given. For each \(c \in X\), pick \(\delta > 0\) such that \(d_Y(f(x),f(c)) < \epsilon\) whenever \((d_X(x,c) < \delta)\). The balls \((B(c,\delta))\) cover \((X)\), and the space \((X)\) is compact. Apply the to obtain a \(\delta > 0\) such that for every \((x \in X)\), there is a \(c \in X\) for which \((B(x,\delta) \subset B(c,\delta))\).

If \((x_1,x_2 \in X)\) where \((d_X(x_1,x_2) < \delta)\), find a \(c \in X\) such that \((B(x_1,\delta) \subset B(c,\delta))\). Then \((x_2 \in B(c,\delta))\). By the triangle inequality and the definition of \((\delta)\) we have \(d_Y(f(x_1),f(x_2)) \leq \epsilon\).
Useful examples of uniformly continuous functions are again the so-called Lipschitz continuous functions. That is if \((X,d_X)\) and \((Y,d_Y)\) are metric spaces, then \(f : X \to Y\) is called Lipschitz or \(K\)-Lipschitz if there exists a \(K \in \mathbb{R}\) such that \(d_Y(f(x),f(c)) \leq K d_X(x,c)\) for all \(x,c \in X\). It is not difficult to prove that Lipschitz implies uniformly continuous, just take \(\delta = \frac{\epsilon}{K}\). And we already saw in the case of functions on the real line, a function can be uniformly continuous but not Lipschitz.

It is worth mentioning that, if a function is Lipschitz, it tends to be easiest to simply show it is Lipschitz even if we are only interested in knowing continuity.

**Exercises**

Consider \(\mathbb{N} \subset \mathbb{R}\) with the standard metric. Let \((X,d)\) be a metric space and \(f : X \to \mathbb{N}\) a continuous function. a) Prove that if \((X)\) is connected, then \(f\) is constant (the range of \(f\) is a single value). b) Find an example where \((X)\) is disconnected and \(f\) is not constant.

Let \(f : \mathbb{R} \to \mathbb{R}\) be defined by \(f(0,0) := 0\), and \(f(x,y) := \frac{xy}{x^2+y^2}\) if \((x,y) \neq (0,0)\). a) Show that for any fixed \((x)\), the function that takes \((y)\) to \(f(x,y)\) is continuous. Similarly for any fixed \((y)\), the function that takes \((x)\) to \(f(x,y)\) is continuous. b) Show that \(f\) is not continuous.

Suppose that \(f : X \to Y\) is continuous for metric spaces \((X,d_X)\) and \((Y,d_Y)\). Let \((A \subset X)\). a) Show that \(f(\overline{A}) \subset \overline{f(A)}\). b) Show that the subset can be proper.

Prove . Hint: Use .

[exercise:msconnconn] Suppose \(f : X \to Y\) is continuous for metric spaces \((X,d_X)\) and \((Y,d_Y)\). Show that if \((X)\) is connected, then \(f(X)\) is connected.

Prove the following version of the . Let \((X,d)\) be a connected metric space and \(f : X \to \mathbb{R}\) a continuous function. Suppose that there exist \((x_0,x_1 \in X)\) and \((y \in \mathbb{R})\) such that \(f(x_0) < y < f(x_1)\). Then prove that there exists a \((z \in X)\) such that \(f(z) = y\). Hint: See .

A continuous function \(f : X \to Y\) for metric spaces \((X,d_X)\) and \((Y,d_Y)\) is said to be proper if for every compact set \((K \subset Y)\), the set \(f^{-1}(K)\) is compact. Suppose a continuous \(f : (0,1) \to (0,1)\) is proper and \(\{x_n\}\) is a sequence in \((0,1)\) that converges to \(0\). Show that \(\{f(x_n)\}\) has no subsequence that converges in \((0,1)\).

Let \((X,d_X)\) and \((Y,d_Y)\) be a one-to-one and onto continuous function. Suppose \((X)\) is compact. Prove that the inverse \(f^{-1} : Y \to X\) is continuous.
Take the metric space of continuous functions \( C([0,1], \mathbb{R}) \). Let \( k : [0,1] \times [0,1] \to \mathbb{R} \) be a continuous function. Given \( k \in C([0,1], \mathbb{R}) \) define \( \varphi_f(x) := \int_0^1 k(x,y) f(y) \, dy \). a) Show that \( T(f) := \varphi_f \) defines a function \( T : C([0,1], \mathbb{R}) \to C([0,1], \mathbb{R}) \).

b) Show that \( T \) is continuous.

Let \( (X,d) \) be a metric space.

a) If \( |p \in X| \), show that \( f \) defined by \( f(x) := d(x,p) \) is continuous.
b) Define a metric on \( (X \times X) \) as in part b, and show that \( g : (X \times X) \to \mathbb{R} \) defined by \( g(x,y) := d(x,y) \) is continuous.
c) Show that if \( (K_1) \) and \( (K_2) \) are compact subsets of \( (X) \), then there exists a \( (p \in K_1) \) and \( (q \in K_2) \) such that \( d(p,q) \) is minimal, that is, \( d(p,q) = \inf \{ (x,y) \colon x \in K_1, y \in K_2 \} \).

**Fixed point theorem and Picard’s theorem again**

*Note: 1 lecture (optional, does not require)*

In this section we prove the fixed point theorem for contraction mappings. As an application we prove Picard’s theorem, which we proved without metric spaces in . The proof we present here is similar, but the proof goes a lot smoother with metric spaces and the fixed point theorem.

**Fixed point theorem**

Let \( (X,d) \) and \( (Y,d') \) be metric spaces. \( f : X \to Y \) is said to be a *contraction* (or a contractive map) if it is a \( k \)-Lipschitz map for some \( k < 1 \), i.e. if there exists a \( k < 1 \) such that \[ d'(f(x),f(y)) \leq k d(x,y) \quad \text{for all } x,y \in X. \]

If \( f : X \to X \) is a map, \( x \in X \) is called a *fixed point* if \( f(x) = x \).

[Contraction mapping principle or Fixed point theorem] [thm:contr] Let \( (X,d) \) be a nonempty complete metric space and \( f : X \to X \) a contraction. Then \( f \) has a unique fixed point. The words *complete* and *contraction* are necessary. See .

Pick any \( x_0 \in X \). Define a sequence \( \{ x_n \} \) by \( x_0 := f(x_0) \), \( d(x_n, x_{n+1}) = d(f(x_n), f(x_{n+1})) \leq k d(x_n, x_{n-1}) \). Suppose \( m > n \), then \[ d(x_m, x_n) \leq \sum_{i=n}^{m-1} d(x_{i+1}, x_i) \leq k \sum_{i=n}^{m-1} d(x_i, x_{i-1}) \leq \cdots \leq k^n d(x_1, x_0) \]

In particular the sequence is Cauchy (why?). Since \( (X) \) is complete we let \( x := \lim x_n \), and we claim that \( (x) \) is our unique fixed point.

Fixed point? The function \( f \) is continuous as it is a contraction, so Lipschitz continuous. Hence \[ f(x) = f(\lim x_n) = \lim f(x_n) = \lim x_{n+1} = x \ . \]
Unique? Let \( \langle x \rangle \) and \( \langle y \rangle \) both be fixed points. \( \|d(x,y) = d(f(x),f(y)) \| \leq k \, d(x,y) \). As \( \langle k < 1 \rangle \) this means that \( \|d(x,y) = 0\) and hence \( \langle x = y \rangle \). The theorem is proved.

The proof is constructive. Not only do we know a unique fixed point exists. We also know how to find it. We start with any point \( x_0 \in X \) and simply iterate \( \langle f(x_0) \rangle \), \( \langle f(f(x_0)) \rangle \), \( \langle f(f(f(x_0))) \rangle \), etc... In fact, you can even find how far away from the fixed point you are, see the exercises. The idea of the proof is therefore used in real world applications.

### Picard’s theorem

Before we get to Picard, let us mention what metric space we will be applying the fixed point theorem to. We will use the metric space \( \langle C([a,b], \mathbb{R}) \rangle \) of . That is, \( \langle C([a,b], \mathbb{R}) \rangle \) is the space of continuous functions \( \langle f\colon [a,b] \to \mathbb{R} \rangle \) with the metric \( \|d(f,g) = \sup \{ x \in [a,b] \} \| f(x) - g(x) \| \) Convergence in this metric is convergence in uniform norm, or in other words, uniform convergence. Therefore, see \( \langle C([a,b], \mathbb{R}) \rangle \) is a complete metric space.

Let us use the fixed point theorem to prove the classical Picard theorem on the existence and uniqueness of ordinary differential equations. Consider the equation \( \| dy \| = F(x,y) \). Given some \( x_0, y_0 \) we are looking for a function \( \langle y = f(x) \rangle \) such that \( \langle f(x_0) = y_0 \rangle \) and such that \( \| F(x) = F(f(x),f(x)) \| \). To avoid having to come up with many names we often simply write \( \| y' = F(x,y) \| \) and \( \langle y(x) \rangle \) for the solution.

The simplest example is for example the equation \( \langle y' = y \rangle, \langle y(0) = 1 \rangle \). The solution is the exponential \( \langle y(x) = e^{x} \rangle \). A somewhat more complicated example is \( \langle y' = -2xy \rangle, \langle y(0) = 1 \rangle \), whose solution is the Gaussian \( \langle y(x) = e^{-x^2} \rangle \).

There are some subtle issues, for example how long does the solution exist. Look at the equation \( \langle y' = y^2 \rangle, \langle y(0) = 1 \rangle \). Then \( \langle y(x) = \sqrt{1-x} \rangle \) is a solution. While \( \langle F(x) \) is a reasonably “nice” function and in particular exists for all \( \langle x \rangle \) and \( \langle y \rangle \), the solution “blows up” at \( \langle x=1 \rangle \). For more examples related to Picard’s theorem see .

Let \( \langle I, J \subset \mathbb{R} \rangle \) be compact intervals, let \( \langle I_0 \rangle \) and \( \langle J_0 \rangle \) be their interiors, and let \( \langle (x_0,y_0) \in I_0 \times J_0 \rangle \). Suppose \( \langle F\colon \times \times J \rangle \) is continuous and Lipschitz in the second variable, that is, there exists an \( \langle L \in \mathbb{R} \rangle \) such that \( \langle \left\| F(x,y) - F(x,z) \right\| = L \times \right\| y - z \right\| \) for all \( \times y, z \in J \), \( \times x \in I \). Then there exists an \( \langle h > 0 \rangle \) and a unique differentiable function \( \langle f\colon [x_0-h, x_0+h] \to J \rangle \) such that \( \langle f(x) = F(f(x),f(x)) \rangle \) and \( \langle y_0 \rangle \rangle \). That is, \( \langle J \rangle \) is the space of continuous functions on \( \langle I \rangle \) with values in \( \langle J \rangle \), in other words, exactly those functions where \( \langle F(f(x),f(x)) \rangle \) makes sense. The metric used is the standard metric given above.

Without loss of generality assume \( \langle x_0 = 0 \rangle \). As \( \langle I \rangle \) is compact and \( \langle F(x,y) \rangle \) is continuous, it is bounded. So find an \( \langle M > 0 \rangle \), such that \( \langle \left\| F(x,y) \right\| \leq M \) for all \( \langle x,y \rangle \in I \). Pick \( \langle \alpha > 0 \rangle \) such that \( \langle [\alpha,\alpha] \rangle \) \( \langle y_0 - \alpha, y_0 + \alpha \rangle \) \( \langle \right\| \times \rangle \) \( \langle \alpha, \alpha \rangle \) \( \langle M+L\alpha \rangle \) \( \langle \right\| \rangle \). Note \( \langle [-h,h] \rangle \) \( \langle \times \rangle \). Define the set \( \langle Y := \{ f \in C([-h,h], \langle \times \rangle ) \} \rangle \) \( \langle [f([-h,h]) \times \langle \times \rangle \rangle \). That is, \( \langle Y \rangle \) is the space of continuous functions on \( \langle [-h,h] \rangle \) with values in \( \langle J \rangle \), in other words, exactly those functions where \( \langle F(f(x),f(x)) \rangle \) makes sense. The metric used is the standard metric given above.

Show that \( \langle Y \times C([-h,h], \langle \times \rangle \rangle \rangle \) is closed. Hint: \( \langle J \rangle \) is closed.
The space \( C([-h,h], \mathbb{R}) \) is complete, and a closed subset of a complete metric space is a complete metric space with the subspace metric, see . So \( Y \) with the subspace metric is complete.

Define a mapping \( (T \colon Y \to C([-h,h], \mathbb{R})) \) by \( (T(f)(x) := y_0 + \int_0^x F(t,f(t)) \, dt) \).

Show that if \( f \colon [-h,h] \to J \) is continuous then \( F(t,f(t)) \) is continuous on \([-h,h]\) as a function of \( t \). Use this to show that \( (T) \) is well defined and that \( (T(f)) \in C([-h,h], \mathbb{R}) \).

We claim \( (T \colon Y \to Y) \) is a contraction. First, for \( (x \in [-h,h]) \) and \( (f,g \in Y) \) we have \( |F(x,f(x)) - F(x,g(x))| \leq L \, |f(x) - g(x)| \leq L \, d(f,g) \). Therefore, \( |T(f)(x) - T(g)(x)| \leq L \, d(f,g) \leq \frac{L\alpha}{M+L\alpha} \, d(f,g) \).

We apply the fixed point theorem (to find a unique \( (f \in Y) \) such that \( (T(f) = f) \), that is, \( (T(f)) \) is the unique differentiable function whose derivative is \( (F(x,f(x))) \) and \( (T(f)(0) = y_0) \). Therefore \( (f) \) is the unique solution of \( (F(x) = F(x,f(x))) \) and \( (f(0) = y_0) \).

Exercises

For more exercises related to Picard’s theorem see .

Let \( (F \colon \mathbb{R} \to \mathbb{R}) \) be defined by \( (F(x) := kx + b) \) where \( (0 < k < 1) \), \( (b \in \mathbb{R}) \).

a) Show that \( (F) \) is a contraction.

b) Find the fixed point and show directly that it is unique.

Let \( (f \colon [0,\frac{1}{4}] \to [0,\frac{1}{4}]) \) be defined by \( (f(x) := x^2) \) is a contraction.

a) Show that \( (f) \) is a contraction, and find the best (smallest) \( (k) \) from the definition that works.

b) Find the fixed point and show directly that it is unique.

a) Find an example of a contraction \( (f \colon X \to X) \) of non-complete metric space \( (X) \) with no fixed point. b) Find a 1-Lipschitz map \( (f \colon X \to X) \) of a complete metric space \( (X) \) with no fixed point.
Consider \(y' = y^2\), \(y(0) = 1\). Use the iteration scheme from the proof of the contraction mapping principle. Start with \(f_0(x) = 1\). Find a few iterates (at least up to \(f_2\)). Prove that the pointwise limit of \(f_n(x)\) is \(\frac{1}{1-x}\), that is for every \(x\) with \(|x| < h\) for some \(h > 0\), prove that \(\lim_{n \to \infty} f_n(x) = \frac{1}{1-x}\).

Suppose \(f : X \to X\) is a contraction for \(k < 1\). Suppose you use the iteration procedure with \(x_{n+1} := f(x_n)\) as in the proof of the fixed point theorem. Suppose \(\gamma(x)\) is the fixed point of \(f\).

a) Show that \(d(x, x_n) \leq k^n d(x_1, x_0) \frac{1}{1-k}\) for all \(n \in \mathbb{N}\).

b) Suppose \(d(x_1, y_2) \leq 16\) for all \(x_1, y_2 \in X\), and \(k = \frac{1}{16}\). Find an \(N\) such that starting at any point \((x_0 \in X)\), \(d(x, x_n) \leq 2^{-16}\) for all \(n \geq N\).

Let \(f(x) = x - \frac{x^2 - 2}{2x}\). (You may recognize Newton’s method for \(\sqrt{2}\))

a) Prove \(\big([1, \infty) \subseteq [1, \infty)\).

b) Prove that \(f(x) = \frac{1}{x}\) is a contraction.

c) Apply the fixed point theorem to find an \(x \geq 1\) such that \(f(x) = x\), and show that \(x = \sqrt{2}\).

Suppose \(f : X \to X\) is a contraction, and \((X, d)\) is a metric space with the discrete metric, that is \(d(x, y) = 1\) whenever \(x \neq y\). Show that \(f\) is constant, that is, there exists a \(c \in X\) such that \(f(x) = c\) for all \(x \in X\).

---

1. The term “modern” refers to late 19th century up to the present.

2. For the fans of the TV show *Futurama*, there is a movie theater in one episode called an \(\aleph_0\)-plex.

3. An algebraist would say that \(\{\mathbb{Z}\}\) is an ordered ring, or perhaps more precisely a commutative ordered ring.

4. Uniqueness is up to isomorphism, but we wish to avoid excessive use of algebra. For us, it is simply enough to assume that a set of real numbers exists. See Rudin for the construction and more details.

5. Named after the Swiss mathematician Jacob Bernoulli (1655 – 1705).

6. The boundedness hypothesis is for simplicity, it can be dropped if we allow for the extended real numbers.

7. use the notation \(\langle x_n \rangle\) to denote a sequence instead of \(\{x_n\}\), which is what uses. Both are common.


9. Named after the Czech mathematician Bernhard Placidus Johann Nepomuk Bolzano (1781 – 1848), and the German mathematician Karl Theodor Wilhelm Weierstrass (1815 – 1897).

10. Sometimes it is said that \(\langle x_n \rangle\) converges to infinity.

12. The divergence of the harmonic series was known before the theory of series was made rigorous. In fact the proof we give is the earliest proof and was given by Nicole Oresme (1323–1382).

13. Demonstration of this fact is what made the Swiss mathematician Leonhard Paul Euler (1707 – 1783) famous.


15. Named after the German mathematician Johannes Karl Thomae (1840 – 1921).


17. Named for the German mathematician Gottfried Wilhelm Leibniz (1646–1716).


20. Named after the German mathematician Georg Friedrich Bernhard Riemann (1826–1866).


22. Such an \( h \) is said to be of bounded variation.

23. Compare this hypothesis to .


26. The notation nor terminology is not completely standardized. The norm is also called the sup norm or infinity norm, and in addition to \( \| f \|_u \) and \( \| f \|_S \) it is sometimes written as \( \| f \|_\infty \) or \( \| f \|_{\infty,S} \).