Contents

Chapter 0

Remarks & Introduction

0.1 Miscellaneous

About

(1) Who: taught by Prof. Wojciech Ożański; LaTeX notes arranged by Qilin Ye

(2) When: taught in spring 2021; notes arranged in summer

(3) Where: on Zoom...

(4) What: lecture notes for MATH 425a, *Fundamental Concepts in Analysis*

(5) What book: Rudin, *Principle of Mathematical Analysis* (PMA)

Theorem Numbering Scheme

Although I personally prefer to number the theorems according to their corresponding chapters and sections, Prof. Ożański decided to go with the week number. I will follow his numbering scheme, i.e., `Theorem a.b` refers to the b^{th} theorem/definition/etc. on the a^{th} week. I will, however, make the section numbers consistent with those in Rudin's PMA, as this course is heavily based on that book.

0.2 Notations

Throughout this course, we shall adopt the following notations[1].

$\forall$ denotes "for all" †

$\exists$ denotes "there exists" †

$=$ denotes "equals"

$:=$ denotes "is defined by"

$\mathbb{N}$ denotes the set $\{1, 2, \ldots\}$

$\lightning$ denotes "contradiction" †

[!] denotes the exclamation mark; the brackets are to distinguish this from factorial ($x!$ vs. x[!])

$(x > 0) \in \mathbb{R}$ is a shortcut for ($x \in \mathbb{R}$ and $x > 0$)

$\epsilon := \epsilon(n)$ denotes "ϵ whose value depends on n"

$\{x_n\}_{n\in I}$ most likely denotes a set with elements are defined over an index set

$(x_n)_{n\geqslant 1}$ most likely denotes a sequence $(x_1, x_2, \ldots)$

[1] If a symbol is marked with †, then I will most likely spell the word out nevertheless, for I (Qilin Ye) personally prefer to write proofs with more sentences than symbols to enhance readability.

Beginning of Jan. 15, 2021

0.3 Introduction & Motivation

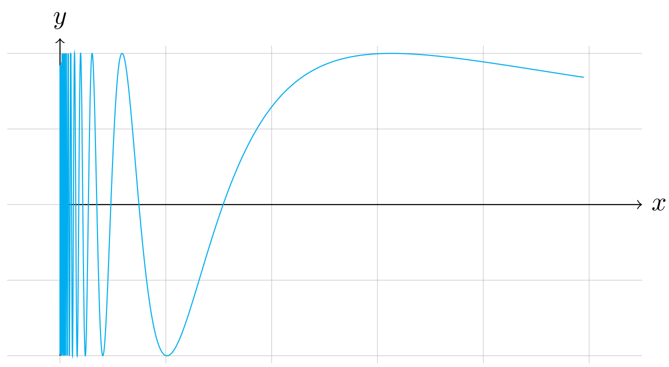
Why analysis? Below is a quote from Professor Kyler Siegel's 425a (the one I took):

> *As you will see, our human intuition can sometimes lead us astray, and many pathologies and exotica arise if we are not extremely careful with our assumptions.*

Before beginning the course, we consider some counterintuitive examples:

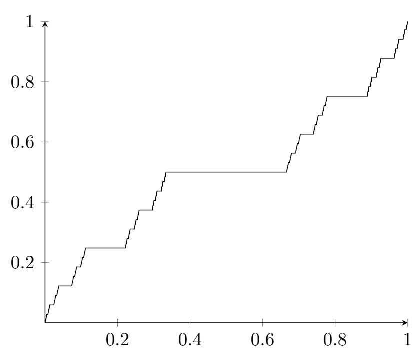
(1) How should we define connectedness of a set?

Example 0.1: the Topologist's Sine Curve / Warsaw Sine Curve. Below is the graph of $f(x) := \sin(1/x)$. We take the union of points on this graph with the points on the line segment between $(0, 1)$ and $(0, -1)$ [points on y-axis] and call it the **closed topologist's sine curve**. It turns out that this set is **connected** (intuitively!) but not **path-connected**, i.e., there exist two points in the set such that we "cannot draw a path connecting them". We will show this rigorously (or not) in Example 8.17.



(2) How about the notion of continuity and differentiability? How should one define continuity over a function?

Example 0.2: the Cantor Function / Devil's Staircase. We construct a function $f : [0, 1] \to [0, 1]$. We first divide $[0, 1]$ into three equal intervals. Let $f(x) = 1/2$ on the middle one (i.e., on $[1/3, 2/3]$). Then for the other two intervals, setting $f(x) = 1/4$ on $[1/9, 2/9]$ and $f(x) = 3/4$ on $[7/9, 8/9]$. We iterate the process and take the *limit* (take it for granted!) function.



It turns out f is continuous *everywhere* and has zero derivative *almost everywhere*, i.e., constant *almost everywhere*, but $f(0) = 0$ and it somehow grows to $f(1) = 1$[!] We will discuss this in-dept in our last lecture!

(3) When can we interchange limits? When not? This is given by the Moore-Smith Theorem.

Example 0.3. Consider a sequence with two indices $(a_{n,k})$. It is not necessarily true that

$$\lim_{k\to\infty}\lim_{n\to\infty} a_{n,k} = \lim_{n\to\infty}\lim_{k\to\infty} a_{n,k}.$$

For example, consider $a_{n,k} = n/(n+k)$. Then

$$\lim_{k\to\infty}\lim_{n\to\infty} \frac{n}{n+k} = \lim_{k\to\infty} 1 = 1,$$

whereas

$$\lim_{n\to\infty}\lim_{k\to\infty} \frac{n}{n+k} = \lim_{n\to\infty} 0 = 0.$$

There are no accidents; we need to know what is going on in each one of these. In mathematical analysis, one need to be careful. Without further ado, let us begin the course.

Chapter 1

The Real and Complex Number Systems

1.1 Introduction

Question. What are real numbers? We know what integers are (or not, lol), and we know how rationals are defined (fractions of integers), but how do we define $\sqrt{2}$, for example?
Remark. It is tempting to say that we can define $\sqrt{2}$ to be the (positive) number x such that $x^2 = 2$, but then we are defining a number by saying it's some number, a circular reasoning.
In this chapter we will develop some tools that solves this issue, even though it might seem unnecessary at first.

1.2 Ordered Sets

Definition 1.1: Subsets & Proper Subsets

If A, B are sets, we say A is a **subset** of B, written $A \subset B$, if $x \in B$ for all $x \in A$. We say A is a **proper subset** of B if $A \subset B$ but $A \neq B$. We denote this by $A \subsetneq B$.

Definition 1.2: Orders & Ordered Sets

A *binary relation* (a relation between two points) $\prec$ on A is an **order** if

(1) (Trichotomy) for all $x, y \in A$, exactly one among the following three is true:

$$x < y \qquad x = y \qquad y < x$$

(2) (Transitivity) if $x, y, z \in A$ and if $x \prec y, y \prec z$, then $x \prec z$.

We say a set A is an **ordered set** if it is equipped with an order relation. We write $x \leq y$ if ($x \prec y$ or $x = y$). *It is more convenient to simply write $<$ instead of $\prec$, so I will use $<$ to denote order in the future.*
Future reference: Definition 1.6 (supremum)

Example 1.3. Let $A := \mathbb{Q}$ with relation $<$ (less than), and we say

$$\frac{a}{b} < \frac{c}{d} \quad \text{if} \quad \frac{a}{b} - \frac{c}{d} = \frac{ad - bc}{bd} \quad \text{is negative.}$$

By doing so, we reduced the relation between rational numbers to that between integers, and we know $(ad - bc)/(bd)$ is negative if and only if (exactly) one between the numerator and denominator is negative.

Example 1.4. Let $\mathbb{Q}^2 := \{(q,r) : q, r \in \mathbb{Q}\}$. It is not clear how to define an order on $\mathbb{Q}^2$.
For example, if we let $(q,v) < (w,p)$ if $q < w$, then $(0,1),(0,2)$ are not comparable. Trichotomy implies $(0,1) = (0,2)$, which is absurd.
For another example, consider $(q,v) < (w,p)$ if $p < w$ and $v < p$, but then $(-1,1)$ and $(0,0)$ cause an issue.

Definition 1.5: Bounded Sets

Let A be an ordered set and let $E \subset A$. We say E is **bounded** (form **above**, in A) if there exists $\alpha \in A$ such that $x \leqslant \alpha$ for all $x \in E$. If so we call α an **upper bound**.
Analogously E is **bouneded below** if there exists a **lower bound** $\beta \in A$ if $\beta \leqslant x$ for all $x \in E$.

Beginning of Jan. 20, 2021

Definition 1.6: Suprema & Infima

Let A be an ordered set and $E \subset A$ bounded above. We say $\alpha \in A$ is the **supremum** of E, written $\alpha = \sup E$, if it is the least upper bound of E, i.e.,

(1) $x \leqslant \alpha$ for all $x \in E$, (*i.e., α is an upper bound of E*) and

(2) if $\gamma < \alpha$ for some $\gamma \in A$ then γ is not an upper bound of E, *i.e., α is the least upper bound of E. In particular, supremum is unique, should it exist: if α, α' are suprema and $\alpha \neq \alpha'$ then the definition of order, either $\alpha < \alpha'$ or $\alpha' < \alpha$, and in either case one between them is not an upper bound of E.*

If every bounded subset E has a supremum, then we say that A has the **least upper bound property** (LUBP).
Analogously we can define the **infimum** (written $\inf E$) and the **greatest lower bound property** (GLBP).
Future reference: Characterization of supremum, characterization of supremum

Remark. The supremum and the infimum, should they exist, may or may not be elements of E. If they are, they agree with the maximum and minimum. (Example: $[1,2)$ contains its infimum but not supremum, and so $[1,2)$ has a minimal element 1 but no maximal element: its supremum $2 \notin [1,2)$.)

Example 1.7. Let $A := \mathbb{Q}$ and define $E := \{q \in \mathbb{Q} : q^3 < 2\}$. Clearly $E \subset \mathbb{Q}$ and any $q \in \mathbb{Q}$ with $q^3 \geqslant 2$ is an upper bound for E. However, does E have a supremum? The answer is no. *We take it for granted that $\sqrt[3]{2}$ is irrational, the proof of which can be easily derived by contradiction, setting $(m/n)^3 = 2$ for integers m, n, and using divisibility of 2.*
Suppose for contradiction that some $p \in \mathbb{Q}$ we have $p = \sup E$. Then either $p^3 < 2, p^3 > 2$, or $p^3 = 2$. The italic sentence above shows $p^3 = 3$ gives a contradiction.
If $p^3 < 2$, we claim that there exists $q \in \mathbb{Q}$ such that $q > p$ and $q^3 < 2$. *We can find an explicit formula for q but this is most likely going to be very complicated. However, once we get to the notion of limits, we can easily circumvent the algebraic construction and obtain our desired result.* This contradicts p's being the supremum.
Likewise, if $p^3 > 2$, there exists $q \in \mathbb{Q}$ with $q < p$ and $q^3 \geqslant 2$ (or > 2). Then q is a smaller upper bound, again contradicting p's being the supremum. Every single case gives a contradcition, completing the proof.
This example shows that $\mathbb{Q}$ does not have the LUBP, and this is not a nice thing...
Future reference: Example 1.15

In fact, the LUBP is equivalent to the GLBP, as we present the following theorem:

Theorem 1.8: LUBP ⇒ GLBP

If A an ordered set with LUBP and $E \subset A$ is bounded from below, then

$$\inf E = \sup\{\alpha \in A : \alpha \text{ is a lower bound of } E\}.$$

In particular A has the GLBP.

Note that the set above (the set of lower bounds of E) is nonempty and bounded: since E is bounded from below, the set contains some lower bound of E and is nonempty, and the set is bounded because any $x \in E$ bounds this set from above.

Proof. Let β denote the RHS. We want to show that (1) β is a lower bound of E and (2) if $\beta < \gamma$ for $\gamma \in A$ then γ is not a lower bound of E.
For (1), if β is not a lower bound of E, then there must exist some $x \in E$ such that $x < \beta$. By definition of supremum, x is not an upper bound of the RHS set. This means there exists some α, some lower bound of E, that is greater than x, contradicting $x \in E$.
For (2), suppose for contradiction that there exists $\gamma > \beta$ that is a lower bound of E. In particular, γ does not belong to the RHS set, so γ is not a lower bound of E, contradiction again! □

Corollary 1.9: GLBP ⇒ LUBP

Similarly, GLBP ⇒ LUBP, and so

$$A \text{ has LUBP} \iff A \text{ has GLBP}. \qquad \text{(Eq.1.1)}$$

Now that we have defined ordered sets, we would like to have some operations on the set, e.g., addition and multiplication. These are necessary tools for our construction $\mathbb{R}$.

1.3 Fields

This is probably the least interesting section in the entire course as it is highly axiomatic. There is no need to fluently memorize all axioms, and they are not the main focus of our course, either.

Definition 1.10: Fields

A set F equipped with two operations, **addition** and **multiplication**, is called a **field** if it satisfies the following *field axioms* (A), (M), and (D):

(A) Axioms for addition:

- (A1) Closure: $x + y \in F$ for all $x, y \in F$.
- (A2) Commutativity: $x + y = y + x$ for all $x, y \in F$.
- (A3) Associativity: $(x + y) + z = x + (y + z)$ for all $x, y, z \in F$.
- (A4) Additive identity: there exists $0 \in F$ such that $0 + x = x$ for all $x \in F$.
- (A5) Additive inverse: for each $x \in F$ there exists an (unique) element $-x \in F$ with $x + (-x) = 0$. *It is in fact not completely trivial that this inverse is unique, and we write $x + (-x)$ as $x - x$ for convenience.*

(M) Axioms for multiplication

- (M1) Closure: $xy \in F$ for all $x, y \in F$.
- (M2) Commutativity: $xy = yx$ for all $x, y \in F$.
- (M3) Associativity: $(xy)z = x(yz)$ for all $x, y, z \in F$.
- (M4) Multiplicative identity: there exists $[1 \neq 0] \in F$ such that $1x = x$ for all $x \in F$.
- (M5) If $x \in F$ and $x \neq 0$ then there exists an (unique) element $1/x \in F$ with $x(1/x) = 1$. *It also takes a small proof to prove uniqueness; once shown, the notion $(1/x)$ is well-defined. We write $x^{-1} := 1/x$.*

(D) The distributive law:

$$x(y + z) = xy + xz$$

for all $x, y, z \in F$. *This connects addition with multiplication.*

Future reference: Example 2.2

Example 1.11. $\mathbb{Q}$ is a field, whereas $\mathbb{Z}$ is not a field as it violates (M5): $1/2 \notin \mathbb{Z}$ for example.

Lemma 1.12: Properties of Fields

Let F be a field. Then for all $x, y, z \in F$, we have

(1) Related to cancellation law for addition:

- (i) $x + y = x + z \implies y = z$.
- (ii) $x + y = x \implies y = 0$.
- (iii) $x + y = 0 \implies y = -x$.
- (iv) $-(-x) = x$.

(2) Related to cancellation law for multiplication: assuming $x \neq 0$,

(i) $xy = xz \implies y = z$.

(ii) $xyx \implies y = 1$.

(iii) $xy = 1 \implies y = x^{-1}$.

(iv) $(x^{-1})^{-1} = x$.

(3) Related to field itself:

(i) $0x = 0$.

(ii) $x \neq 0, y \neq 0 \implies xy \neq 0$.

(iii) $(-x)y = -(xy) = x(-y)$.

(iv) $(-x)(-y) = xy$.

Having defined ordered sets and fields, the next natural construction is a combination of them:

Definition 1.13: Ordered Fields

An **ordered field** is a field F which is also an ordered set, such that

(1) $x + y < x + z$ if $x, y, z \in F$ with $y < z$, and

(2) $xy > 0$ if $x, y \in F$, $x > 0$, and $y > 0$.

These express the compatibility of the order relation with the field axioms.

If $x > 0$ we call x **positive**; if $x < 0$ we call x **negative**.

Beginning of Jan. 22, 2021

Lemma 1.14: Properties of an Ordered Field

Let F be an ordered field and let $x, y \in F$. Then

(1) $x > 0$ if and only if $-x < 0$.

(2) If $x > 0$ and $y < z$ then $xy < xz$.

(3) If $x < 0$ and $y < z$ then $xy > xz$.

(4) If $x \neq 0$ then $x^2 > 0$. *In particular, this shows* $1 > 0$.

(5) If $0 < x < y$ then $0 < 1/y < 1/x$.

Future reference: $\mathbb{C}$ is not ordered

Example 1.15. $\mathbb{Q}$ is an ordered field, even though it does not have LUBP.

The issue with $\mathbb{Q}$ and LUBP is that, while each rational number corresponds to some integer fraction, there are numbers *on the real line that do not correspond to any integer fraction. We need to define the real numbers as an extension that "fills the gaps' via what is called the* ***Dedekind cuts****.*

1.4 The Real Field

Theorem 1.16: Dedekind Cuts and the Construction of $\mathbb{R}$

There exists an ordered field (which we call $\mathbb{R}$) that has the LUBP and contains $\mathbb{Q}$ as a subfield, i.e., $\mathbb{Q} \subset \mathbb{R}$ and $(\mathbb{Q}, +, \cdot)$ is a field where $+, \cdot$ are inherited from $(\mathbb{R}, +, \cdot)$. The members of $\mathbb{R}$ are called **real numbers**.

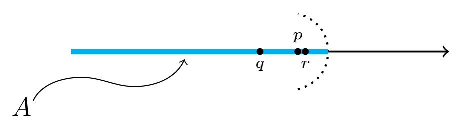
Proof Sketch. We say a set $A \subset \mathbb{Q}$ is a **Dedekind cut** if it satisfies the following property:

(1) $A \neq \varnothing, A \neq \mathbb{Q}$,

(2) If $p \in A, q \in \mathbb{Q}$ are such that $q < p$, then $q \in A$.

(3) If $p \in A$ then there exists $r \in A$ with $r > p$, i.e., A has no largest element.

The diagram below serves as a heuristic visualization:



We claim that each point on the real line can be represented as a cut. For example, the example of $\sqrt[3]{2}$ is represented by the set $\{x \in \mathbb{Q} : x^3 < 2\}$. Now we provide a *sketch* of the proof, as the actual one is rather long (see Rudin's Ch1 Appendix for a complete proof).

(Step 1) Define Dedekind cuts as above.

(Step 2) Define $\mathbb{R}$ to be the set of all cuts with the order relation $A < B$ if $A \subset B$.

(Step 3) Define addition of two cuts by

$$A + B := \{a + b : a \in A, b \in B\}.$$

In particular the zero element 0^* is the set of all rational numbers, i.e., $0^* := \{q \in \mathbb{Q} : q < 0\}$. For $A > 0^*$, define $-A$ to be the cut such that $A + (-A) = 0^*$.

(Step 4) Define a cut A to be positive if $A > 0^*$. Multiplication of two positive cuts A, B is given by

$$\{q : q \leqslant ab \text{ for some } (a > 0) \in A, (b > 0) \in B\}.$$

(Step 5) Complete the multiplication by setting $A0^* = 0^*A = 0^*$ and

$$AB = \begin{cases} (-A)(-B) & A < 0^*, B < 0^* \\ -[(-A)B] & A < 0^*, B > 0^* \\ -[A(-B)] & A > 0^*, B < 0^*. \end{cases}$$

Then $\mathbb{R}$ with $<, +, \cdot$ defined as such forms an ordered field with LUBP and contains $\mathbb{Q}$ as a subfield by associating a rational number by a cut at a rational number.

Future reference: Archimedean Property of $\mathbb{R}$ □

Remark. From now on, we will treat $\mathbb{R}$ "as usual", i.e., we no longer need to worry about addition, multiplication, etc. We may also freely use the fact that $\mathbb{R}$ has the LUBP, i.e., the supremum and infimum exist for *any* set $A \subset \mathbb{R}$. This claim does need to be supplemented with some extra constructions. By convention,

(1) The supremum of a set unbounded from above is ∞.

(2) The infimum of a set unbounded from below is $-\infty$.

(3) $\inf \varnothing = \infty$ as any number is trivially a lower bound of $\varnothing$.

(4) $\sup \varnothing = -\infty$ as any number is trivially an upper bound of $\varnothing$.

Example 1.17: Archimedean Property of $\mathbb{R}$. We present a seemingly trivial claim:

For all $x, y \in \mathbb{R}$ such that $x > 0$, there exists $n \in \mathbb{N}$ such that $nx > y$.

Proof. Define $A := \{nx : n \in \mathbb{N}\}$. Suppose the claim is false so all elements of A is less than y, i.e., y is an upper bound of A. We define $\alpha := \sup A$, which exists because $\mathbb{R}$ has the LUBP. Note that $\alpha - x < \alpha$ since $x > 0$. It follows that $\alpha - x$ cannot be an upper bound of A as α is the supremum (LUB). Thus there exists $m \in \mathbb{N}$ such that $\alpha - x < mx \implies \alpha < (m+1)x \in A$, contradicting $\alpha = \sup A$. Hence the claim must hold. □

Beginning of Jan. 25, 2021

Example 2.1. Another seeming trivial statement that claims we are able to take roots in $\mathbb{R}$:

For all $(x > 0) \in \mathbb{R}$ and $n \in \mathbb{N}$, there exists a unique $(y > 0) \in \mathbb{R}$ such that $y^n = x$.

Future reference: Example 2.2

Proof. We first show uniqueness. Suppose for $0 < y_1 < y_2$ we have $y_1^n = x = y_2^n$. Using Lemma 1.14 we get $y_1^n < y_2^n$, an immediate contradiction. Now it remains to show existence.
Define $E := \{t > 0 : t^n < x\}$. We know E is nonempty:

$$\frac{x}{1+x} < x \text{ and } \left(\frac{x}{1+x}\right)^n < \left(\frac{x}{1+x}\right) < x \implies \frac{x}{1+x} \in E.$$

It is also bounded above, as $(1+x)^n > 1+x > x$. Therefore it has a finite supremum and we define $y := \sup E$. We will show that $y^n = x$ which will complete the proof. Notice that, for $0 < a < b$,

$$b^n - a^n = (b-a)\sum_{i=0}^{n-1} a^i b^{(n-1)-i} < (b-a)nb^{n-1}. \tag{Δ}$$

If $y^n < x$ then $x - y^n > 0$, consider a positive number $\dfrac{x-y^n}{n(y+1)^{n-1}}$ and any $k \in (0,1)$ smaller than the fraction. Now we apply (Δ) to $0 < y < y+k$:

$$(y+k)^n - y^n < kn(y+k)^{n-1} < kn(y+1)^{n-1} < x - y^n,$$

meaning that $(y+k)^n < x$ and so $y+k \in E$, contradiction.
If $y^n > x$ then $y^n - x > 0$. Similarly, we set $k := \dfrac{y^n - x}{ny^{n-1}} < \dfrac{y^n}{ny^{n-1}} \leqslant \dfrac{y^n}{y^{n-1}} = y$ and claim $y - k$ is still an upper bound of

E (so a contradiction arises). If $y-k$ is not an upper bound, then there exists $t > y-k$ such that $t^n < x$, and so

$$y^n - x < y^n - t^n < y^n - (y-k)^n < kny^{n-1} = y^n - x,$$

where the last $<$ uses (Δ) and the $=$ uses the definition of k. Since $y^n - x < y^n - x$ is absurd, we are done. □

Remark. In fact, we can replace the exponent n by any nonzero real number.

Corollary 2.2

If $a, b \in \mathbb{R}$ are positive numbers and if $n \in \mathbb{N}$, then $(ab)^{1/n} = a^{1/n} b^{1/n}$.

Proof. Let $\alpha := a^{1/n}$ and $\beta := b^{1/n}$ so that $\alpha^n = a$ and $\beta^n = b$. Using Def 1.10 (M2) (commutativity of multiplication in a field),

$$ab = \underbrace{\alpha \cdot \ldots \cdot \alpha}_{n} \cdot \underbrace{\beta \cdot \ldots \cdot \beta}_{n} = \underbrace{(\alpha\beta) \cdot \ldots \cdot (\alpha\beta)}_{n} = (\alpha\beta)^n.$$

Therefore we have found one solution to $x^n = ab$, and by Example 2.1 it is the unique one. □

Example 2.3. Let $I = (a, b] \subset \mathbb{R}$ (or $[a, b), (a, b), [a, b]$) for some $a < b$. Then $a = \inf I$ and $b = \sup I$.

Proof. We first show that $\inf I = a$. First notice that a is a lower bound for I; in particular, its infimum exists. Suppose $a \neq \inf I$; then there exists $a' > a$ also a lower bound of I. However, (a, a') is nonempty and $(a, a') \cap I$ is also nonempty, so there exists $e \in (a, a') \cap I$. In particular, such $e \in I$ and $e < a'$, contradicting the assumption that a' is a lower bound. We can show analogously that $b = \sup I$ and the proof is omitted. □

Example 2.4. (A harder example) Find the supremum and infimum of

$$A := \{\sqrt{n} - \lfloor \sqrt{n} \rfloor : n \in \mathbb{N}\}.$$

Note that since it is well-defined to take square roots as justified previously, the set is well defined, and so are its supremum and infimum.

Solution. It is obvious that the infimum is 0 and that the supremum $\leqslant 1$ as $\sqrt{n} - \lfloor \sqrt{n} \rfloor < 1$.
We will now show that $\sup A = 1$ as anyone would guess. If not, then there exists $\epsilon > 0$ such that

$$\sqrt{n} - \lfloor \sqrt{n} \rfloor \leqslant 1 - \epsilon \qquad \text{for all } n \in \mathbb{N}.$$

If we take $n := k^2 + 2k$ for some $k \in \mathbb{N}$, then $k < \sqrt{k^2+2k} < k+1 = \sqrt{k^2+2k+1}$, so $\left\lfloor \sqrt{k^2+2k} \right\rfloor = k$, and

$$\begin{aligned} \sqrt{n} - \lfloor \sqrt{n} \rfloor &= \sqrt{k^2+2k} - \left\lfloor \sqrt{k^2+2k} \right\rfloor = \sqrt{k^2+2k} - k \\ &= \frac{(\sqrt{k^2+2k} - k)(\sqrt{k^2+2k} + k)}{\sqrt{k^2+2k} + k} \\ &= \frac{2k}{\sqrt{k^2+2k} + k} = \frac{2}{\sqrt{1 + 2/k} + 1}. \end{aligned}$$

For sufficiently large k [to be precise, if $k > (1-\epsilon)^2/(2\epsilon)$], the fraction $> 1 - \epsilon$. Therefore $\sup A = 1$. □

1.5 The Euclidean Space

Definition 2.5: $\mathbb{R}^n$, Inner Product, & Norm

For each integer n, we define $\mathbb{R}^n$ to be the collection of n-tuples with real entries:

$$\mathbb{R}^n := \{x = (x_1, ..., x_n) : x_i \in \mathbb{R}\}.$$

We see $\mathbb{R}^n$ is a **vector field**, in which addition on $\mathbb{R}^n$ is defined by[1]

$$x + y := (x_1 + y_1, ..., x_n + y_n),$$

and *scalar* multiplication is defined by

$$\lambda x := (\lambda x_1, ..., \lambda x_n).$$

However, there is no mulplication defined on $\mathbb{R}^n$, so $\mathbb{R}^n$ is not a field.
Nevertheless, we have the **inner product**, a mapping $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$\langle x, y \rangle = x \cdot y := \sum_{i=1}^{n} x_i y_i =: x_1 y_1 + ... + x_n y_n,$$

(*both notations are fine, but $\langle \cdot, \cdot \rangle$ seems more convenient and unambiguous*) and we also have the (Euclidean standard / Euclidean 2-) **norm**[2]of x (a function $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$) defined by

$$\|x\|_2 = |x| := \sqrt{\langle x, x \rangle} = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$

Beginning of Jan. 27, 2021

Remark. *The lecture seemed to have taken the following result for granted, but I will nevertheless include it for the sake of completeness.* The $\mathbb{R}^n$ inner product $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is *bilinear*, meaning that

$$\langle x, y_1 + \lambda y_2 \rangle = \langle x, y_1 \rangle + \lambda \langle x, y_2 \rangle \quad \text{and} \quad \langle x_1 + \lambda x_2, y \rangle = \langle x_1, y \rangle + \lambda \langle x_2, y \rangle.$$

This can be easily verified using the direct definition.

Lemma 2.6: Properties of Norms

For simplicity we consider $|\cdot|$ on $\mathbb{R}^n$. It satisfies the following:

(1) (Non-degeneracy) $\|x\| \geqslant 0$ for all x and $\|x\| = 0$ if and only if $x = 0$.

(2) (Absolute homogeneity) $\|\lambda x\| = |\lambda| \|x\|$ ($|\lambda|$ denotes the absolute value).

(3) (**Cauchy-Schwarz Inequality**) $|\langle x, y \rangle| \leqslant \|x\| \|y\|$, and $\langle x, y \rangle = \|x\| \|y\|$ if and only if y is a scalar multiple of x, i.e., $y = \lambda x$ for some $\lambda \in \mathbb{R}$.

Future reference: Complex Cauchy-Schwarz Inequality

(4) (**Triangle inequality**) $\|x + y\| \leqslant \|x\| + \|y\|$.

In fact, (1), (2), and (4) characterize a norm.

[1] When context is clear, we write $x \in \mathbb{R}^n$ instead of $(x_1, ..., x_n) \in \mathbb{R}^n$ for convenience.
[2] Currently in $\mathbb{R}^n$, unless otherwise specified, we let $|\cdot|$ be the 2-norm. $|\cdot|$, $\|\cdot\|$, and $\|\cdot\|_2$ are interchangeable when context is clear.

Proof of Cauchy-Schwarz Inequality. Consider points of form $x - \lambda y$ ($x, y \in \mathbb{R}^n$ so $x - \lambda y \in \mathbb{R}^n$), and consider their norms, which must be nonnegative by *non-degeneracy*. On the other hand, using the bilinearity of inner products, we obtain

$$\begin{aligned} 0 \leqslant \|x - \lambda y\|^2 &= \langle x - \lambda y, x - \lambda y \rangle \\ &= \langle x, x \rangle - 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle \\ &= \|x\|^2 - (2\langle x, y \rangle)\lambda + \|y\|^2 \lambda^2.[3] \end{aligned} \tag{Δ}$$

Picking a specific $\lambda := \langle x, y \rangle / \|y\|^2$, we see that

$$\|x\|^2 - 2\langle x, y \rangle \frac{\langle x, y \rangle}{\|y\|^2} + \|y^2\| \frac{\langle x, y \rangle^2}{\|y\|^4} = \|x\|^2 - \frac{\langle x, y \rangle^2}{\|y\|^2} \geqslant 0,$$

so indeed $\|x\|^2 \|y\|^2 \geqslant \langle x, y \rangle^2$, and taking roots on both sides completes the proof of $\leqslant$. "=" is attained if and only if $x - \lambda y = 0$, and this is precisely the claim. □

Proof of triangle inequality. Taking squares on both sides,

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 \\ &\leqslant \|x\|^2 + 2|\langle x, y \rangle| + \|y\|^2 \\ \text{[Cauchy-Schwarz]} &\leqslant \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2. \end{aligned}$$

□

[3] Another proof (the one I learned from Pugh's book and Siegel's lectures): (Δ) is a quadratic equation in terms of λ. Since it is forced to be nonnegative by non-degeneracy of norms, the discriminant must be nonpositive, so $4\langle x, y \rangle^2 \geqslant 4\|x\|^2\|y\|^2$, and the claim follows.

1.6 The Complex Field

Recall that we said $\mathbb{R}^n$ is not a field as there is no multiplication defined. However, we can introduce a notion of "multiplication" on $\mathbb{R}^2$, and this process makes it a field, which we call the **complex field**.

For $(a,b),(c,d) \in \mathbb{R}^2$, we define the notions of addition and multiplication for this construction to be

$$(a,b)+(c,d) := (a+c, b+d) \qquad \text{and} \qquad (a,b)\cdot(c,d) := (ac-bd, ad+bc), \tag{Eq.2.1}$$

and we define $\mathbb{C}$ to be $\mathbb{R}^2$ equipped with such operations. We call it the **complex field / complex plane**.
Future reference: $i^2 = 1$

Theorem 2.7: $\mathbb{C}$ is a Field

$\mathbb{C}$ is a field; in particular, $(0,0)$ is the additive identity, $(1,0)$ the multiplicative identity, and

$$(a,b)^{-1} := \left(\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2}\right) \qquad \text{for } (a,b) \neq (0,0).$$

It is easy to verify that $(a,b)\cdot\left(\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2}\right) = \left(\frac{a^2}{a^2+b^2}+\frac{b^2}{a^2+b^2}, -\frac{ab}{a^2+b^2}+\frac{ba}{a^2+b^2}\right) = (1,0).$

Definition 2.8: Imaginary Unit

We define $i := (0,1) \in \mathbb{C}$ to be the **imaginary unit**.

Theorem 2.9: $i^2 = 1$

$i^2 = 1$. *Quick proof:* $i^2 = (0,1)\cdot(0,1) = (-1,0)$ *by Equation 2.1.*

Theorem 2.10: Representation of Complex Numbers

Any $(a,b) \in \mathbb{C}$ (where $a,b \in \mathbb{R}$) can be represented as $a+bi$.
Quick proof: $a+bi = (a,0)+(b,0)\cdot(0,1) = (a,0)+(0,b) = (a,b)$.

From now on we will write any complex number $z = (a,b) \in \mathbb{C}$ as $a+bi$ and keep in mind that $i^2 = -1$.

Definition 2.11

For $z = a+bi \in \mathbb{C}$, we define

(1) (**Complex conjugate**) $\overline{z} := a - bi$.

(2) (**Modulus**) $|z| := \sqrt{a^2+b^2}$ (which is equal to $\|(a,b)\|$).

(3) a to be the **real part** of z, written $\mathfrak{Re}(z)$, and b to be the **imaginary part** of z, written $\mathfrak{Im}(z)$.

Theorem 2.12: Properties of $\mathbb{C}$

Let $z, w \in \mathbb{C}$.

(1) $|z|^2 = z\overline{z}$, $\overline{z+w} = \overline{z}+\overline{w}$, $\overline{zw} = \overline{z}\cdot\overline{w}$, $z+\overline{z} = 2\mathfrak{Re}(z)$, and $z-\overline{z} = 2i\mathfrak{Im}(z)$.

(2) $|zw| = |z||w|$.

(3) $\mathfrak{Re}(z) \leqslant |z|$ and $|\mathfrak{Im}(z)| \leqslant |z|$.

(4) $|z + w| \leqslant |z| + |w|$ *which is immediate by the triangle inequality for* $\mathbb{R}^2$.

Proof. For convenience we write $z = a + bi$ and $w = c + di$.

(1) $z\overline{z} = (a + bi)(a - bi) = a^2 - (bi)^2 = a^2 + b^2 = |z^2|$. The addition one is also clear, and so are the last two statements, so we will only prove the third:

$$\overline{zw} = \overline{(a + bi)(c + di)} = \overline{(ac - bd) + i(ad + bc)} = (ac - bd) - i(bc + ad)$$

and

$$\overline{z} \cdot \overline{w} = (a - bi)(c - di) = (ac - bd) + i(-ad - bc) = \overline{zw}.$$

(2) From above we already know $zw = (ac - bd) + i(bc + ad)$, so

$$\begin{aligned} |zw|^2 &= (ac - bd)^2 + (bc + ad)^2 \\ &= (a^2c^-2abcd + b^2d^2) + (b^2c^2 + 2abcd + a^2d^2) \\ &= a^2c^2 + b^2d^2 + b^2c^2 + a^2d^2 = (a^2 + b^2)(c^2 + d^2) = |z|^2|w|^2. \end{aligned}$$

(3) Clearly $|\mathfrak{Re}(z)|^2 = a^2 \leqslant a^2 + b^2 = |z^2|$ and the other one is analogous.

□

Theorem 2.13: $\mathbb{C}$ is not Ordered

While $\mathbb{C}$ is a field, it is not an ordered field: there is no order on $\mathbb{C}$.

Proof. Suppose there were. By Lemma 1.14.4, $i \neq 0$ implies $i^2 > 0$, but $-1 < 0$ by Lemma 1.14.1 and .4, so

$$0 < i^2 = -1 < 0, \qquad \text{contradiction.}$$ □

Theorem 2.14: Complex Cauchy-Schwarz Inequality

If $a_i, b_i \in \mathbb{C}$, then

$$\left|\sum_{i=1}^{n} a_i b_i\right|^2 \leqslant \sum_{i=1}^{n} |a_i|^2 \sum_{i=1}^{n} |b_i|^2.$$

Proof. Each $|a_i|, |b_i|$ is real, and $|a_i b_i| = |a_i||b_i|$ as shown above. Then the real Cauchy-Schwarz applies. □

Example 2.15: Applications of Complex Numbers.

(1) (Electrical engineering). In electronic circuits, one represents *alternating current* (AC) as a vector rotating in the complex plane, whose formula is given by

$$U(t) = A\exp(i(\omega t + \theta_0)),$$

where A represents the *amplitude* (maximum), ω the frequency, t time, θ_0 the initial phase, and the entire exponential the rotation. In the US $\omega = 60$ Hz and in EU/China $\omega = 50$ Hz.

(2) (**Euler's formula**) $e^{ix} = \cos x + i \sin x$. In particular taking $x := \pi$ gives $e^{i\pi} = -1$, "*the most beautiful equation in math*" as it relates four essential constants in math.

(3) (Fluid dynamics) Modeling 2D flows near obstacles via *conformal mapping*.

(4) (Fourier analysis, signal analysis, control theory) Given $f(x)$, we can define its *Fourier transform*

$$\hat{f}(\xi) = \int f(x) e^{ix\xi} \, \mathrm{d}x.$$

The most classical example: we can decompose signals into a combination of sinusoidal waves.[4]

(5) (Holomorphic functions) If a function is complex differentiable it is smooth, and we call such functions *holomorphic*. The *Cauchy Inrtegral Formula* gives

$$f^{(n)}(x_0) = \frac{n!}{2\pi i} \oint_C \frac{f(z)}{(z - x_0)^{n+1}} \, \mathrm{d}z.$$

[4] I have written some notes on Fourier transforms on finite groups as part of my MATH 410 exam project. They can be found here.

Chapter 2

Basic Topology

Beginning of Jan. 29, 2021

2.1 Functions

Definition 2.16: Functions

(1) We say f is a **function** from A to B, denoted $f : A \to B$ it associates every $x \in A$ with a point in B, i.e., $f(x) \in B$ for all $x \in A$. We call A the **domain** of f and B the **codomain**.

(2) For $E \subset A$, we define $f(E) := \{b \in B : f(x) = b \text{ for some } x \in A\}$ to be the **image** of E under f.

(3) We say f is **surjective** (or f is a **surjection**) if, for every $b \in B$, there exists $x \in A$ with $f(x) = b$. *In other words, each $b \in B$ corresponds to at least one $x \in A$.*

(4) We say f is **injective** (or f is a **injection**) if, for all $x, y \in A$, $f(x) = f(y) \Rightarrow x = y$. *In other words, each $b \in B$ corresponds to at most one $x \in A$.*

(5) We say f is **bijective** (or f is a **bijection**) if f is both injective and surjective. *In other words, each $b \in B$ corresponds to exactly one $x \in A$.*

(6) If there exists a bijection $f : A \to B$, then we say A and B are **equicardinal** (or they have the same **cardinal number**). *They have the same "number of elements".* We write $A \sim B$.

2.2 Finite, Countable, & Uncountable Sets

Definition 2.17: Cardinality

Let A be a set.

(1) We say A is **finite** if $A \sim \{1, ..., n\}$ for some $n \in \mathbb{N}$. If so, we say the **cardinality** of A, written $|A|$ [1], is n.

(2) If A is not finite, we say it is **infinite**.

(i) If A is infinite and $A \sim \mathbb{N}$, we say A is **countable**[2] and we say $|A| = \aleph_0$ (aleph-null). *If so, we can find an **enumeration** $\{a_i\}_{i \geqslant 1}$ that lists all elements of A.*

(ii) If A is infinite and not countable then it is **uncountable**. *Uncountable sets can have different cardinalities, some of which we will examine later.*

Definition 2.18: Unions & Intersections over Indexed Families of Sets

Let I be an **index set** (countable or uncountable), and let $\{A_\alpha\}$ where $\alpha \in I$ to be an *indexed family of sets*. We define the **union** and **intersection** over this indexed family of sets as the following:

$$\bigcup_{\alpha \in I} A_\alpha := \{x : x \in A_\alpha \text{ for some } \alpha \in I\} \quad \text{and} \quad \bigcap_{\alpha \in I} A_\alpha := \{x \in A_\alpha \text{ for all } \alpha \in I\}.$$

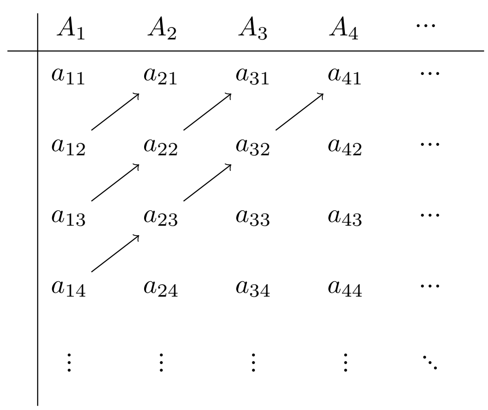
We provide a heuristic example of a union over an uncountably indexed family of sets: $[0,1] = \bigcup_{n \in [0,1]} \{n\}$. *(We'll show such index set is uncountable soon.)*

Lemma 2.19

A(n at most) countable union of (at most) countable set is (at most) countable. (*At most countable means countable or finite.*)

Future reference: Example 2.20.4

Proof. We omit the parenthesized words (if the "all countable" version holds, then clearly the claim still holds if we replace one or more "countable" by "finite" as we will see in the proof) and use the *diagonal numbering*. Let A_n be a countable set for each n and define $A := \bigcup_{n \in \mathbb{N}} A_n$ to be our countable union of countable sets. If we enumerate the elements of each A_n in the n^{th} column as shown below, we can find a way to traverse through all elements in A.



Let us define a new sequence whose first few terms are $a_{11}, a_{12}, a_{21}, a_{13}, \dots$ (traversing through the most upper-left diagonal remaining before moving to the next one). In fact, we can justify rigorously that we will reach $a_{n,m}$ (the m^{th} element of A_n) in this manner. Since the sums of indices of different elements on the same diagonal are the

[1] Notation of cardinality varies from text to text, e.g., $\text{card}(A), |A|, \#A$, and more. I have chosen $|A|$, the notion my lectures used when I took 425a, partly because it is the most convenient to write (and even till today, I still haven't found an aesthetically pleasing hashtag in LaTeX...). This is not to be confused with norms, modulus, or absolute values (and hopefully they shouldn't; sets are usually denoted in capital letters).

[2] The definition of countable sets also vary; some texts exclude finite sets from countable sets whereas some consider finite sets to be countable also. Pugh's book (and hence my 425a) used the latter, whereas Rudin chose the former. We will follow Rudin's option throughout this 425a.

same, $a_{m,n}$ is precisely the n^{th} element on the $(m+n-1)^{\text{th}}$ diagonal, and this corresponds to the

$$n + \sum_{k=1}^{m+n-2} k = \frac{2n + (m+n-2)(m+n-1)}{2}$$

-$^{\text{th}}$ element in our sequence. Define $f : A \to \mathbb{N}$ by $f(a_{n.m})$ to be the position of $a_{n,m}$ on our newly constructed sequence. It follows that f is surjective. It must also be injective, as each point in A has a unique pair of indices, and $a_{n,m}$ clearly only appears once in our sequence. Hence f is injective and thus bijective, and we conclude that $A \sim \mathbb{N}$ is countable.

If some $a_{n,m}$ is missing because either the index set is finite or because some A_n is finite, we simply skip that element when traversing through the list. The proof is still valid. □

Example 2.20.

(1) $\mathbb{Z}$ is countable.

Valid proof: set $f : \mathbb{N} \to \mathbb{Z}$ by defining

$$\begin{cases} f(2k+1) = k & k \geqslant 0 \\ f(2k) := -k & k \geqslant 1. \end{cases}$$

This is a bijection from $\mathbb{N} \to \mathbb{Z}$ so $\mathbb{Z} \sim \mathbb{N}$.

Valid proof: enumerate $\mathbb{Z}$ as $\{0, -1, 1, -2, 2, \ldots\}$ and verify that each $z \in \mathbb{Z}$ is included.

INVALID proof: write $\mathbb{Z}$ as $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

(2) Any subset of a(n at most countable) countable set $E \subset A$ is (at most) countable.

Proof sketch: if E is finite then the claim is trivial; if E is infinite, express $(E \subset A) := \{e_i\}_{i \geqslant 1}$ and let $(S \subset A) \subset \{s_j\}_{j \geqslant 1}$ be such that $s_i := f(e_i)$, where $f(e_i)$ denotes the smallest integer k such that exactly i elements among $\{a_1, \ldots, a_k\}$ lie in E. Since E is finite, this construction is valid. Then S is countable and $S \sim E$. □

> **Remark.** In particular, this shows that "countable" is the "smallest" infinity we can have (where "uncountable" refers to "higher levels" of infinity).

(3) $\mathbb{Q}$ is countable:

$$\mathbb{Q} = \bigcup_{n=1}^{\infty} \left\{ \frac{x}{n} : x \in \mathbb{Z}, \gcd(x, n) = 1 \right\} \subset \bigcup_{n=1}^{\infty} \left\{ \frac{x}{n} : x \in \mathbb{Z} \right\}.$$

The RHS is a countable union of countable sets and is therefore countable, whereas $\mathbb{Q}$ is a subset, so it is at most countable. Of course we know $\mathbb{Q}$ is not finite: $\mathbb{N} \subset \mathbb{Q}$.

(4) Let A be any collection (finite, countable, or uncountable) of pairwise disjoint intervals $I \subset \mathbb{R}$. Then A is at most countable.

Proof sketch: note that $A = \{I : I \in A, |I| > 0\}$, (since $|I| > 0$ adds no additional restriction), so

$$A = \bigcup_{n=1}^{\infty} \{I \in A : |I| \geqslant 1/n\}.$$

Note that

$$\{I \in A : |I| \geqslant 1/n\} = \bigcup_{k \in \mathbb{N}} \{\text{disjoint } I \in A : |I| \geqslant 1/n \text{ and } I \cap [k, k+1) \neq \varnothing\},$$

where the last requirement is because each I must lie in some $[k, k+1)$. Also notice that at most $n+1$ disjoint open intervals can intersect with the same $[k, k+1)$. Therefore, each of these RHS sets is countable, and using Lemma 2.19 twice, we conclude that A must also be countable. □

Beginning of Feb. 1, 2021

Example 3.1: Cantor's Diagonalization. Let $B := \{(\{a_n\}_{n\geqslant 1} : a_n \in \{0,1\}\}$, the set of all sequences whose digits are either 0 or 1, is uncountable.

Proof. Suppose B is countable; let $\{B_i\}_{i\geqslant 1}$ be an enumeration of B. Let us list the B_i's as the columns of a table with infinite rows and infinite columns. For example, it may look like

	b_1	b_2	b_3	$\cdots$
	1	0	1	$\cdots$
	0	1	1	$\cdots$
	1	0	0	$\cdots$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$

We consider the *diagonal sequence* $\hat{a} := (a_n)_{n\geqslant 1}$ defined by $\hat{a}_i = 1 - b_{i,i}$. For example, the first three terms of $\hat{a}$ are $0, 0, 1$, respectively, since b_{11}, b_{22}, b_{33} are $1, 1, 0$, respectively. Clearly this is an element of B, but does it appear in the enumeration $\{B_i\}$? No! Its first term disagrees with that of b_1, and likewise its n^{th} term disagrees with that of b_n. Therefore we obtain a contradiction, and B must be uncountable!
*This proof is due to Cantor, hence the name **Cantor's Diagonalization**.* □

Now we let the *real fun* begin — introducing the (ϵ-δ) language!

2.3 Metric Spaces

Definition 3.2: Metric

Let A be a set. We say a function $d : A \times A \to \mathbb{R}$ is called a **metric**[3] on A (or **distance** function) if

(1) (Non-degeneracy) $d(a, b) \geqslant 0$ with $d(a, b) = 0$ if and only if $a = b$,

(2) (Symmetry) $d(a, b) = d(b, a)$, and

(3) (Triangle inequality) $d(a, b) \leqslant d(a, c) + d(c, b)$.

If so, we say (A, d) is a **metric space**. *It is customary to simply say "A is a metric space" for convenience.*
Future reference: $(C(K), \|\cdot\|_{\sup})$ is complete

Example 3.3: Some Metric Spaces.

(1) If $K \subset \mathbb{R}^n$ then $(K, \|\cdot\|)$ is a metric space.

(2) $(S^2, \mathrm{d}g)$, the sphere with the *geodesic distance*, is a metric space.

Unless otherwise specified, from now on we will look at an abstract metric space (X, d).

Definition 3.4: "ϵ-N" Sequential Convergence

Given a sequence $(x_n)_{n\geqslant 1} \subset X$, we say x_n **converges** to x and say x is *the* **limit** of (x_n) if

For all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n, x) \leqslant \epsilon$ for all $n \geqslant N$.

For notation, we write $x_n \to x$ or $(x_n) \to x$ and $x = \lim_{n\to\infty} x_n$.
Future reference: (Open) neighborhoods

Lemma 3.5: Limits are Unique

In a metric space, the limits of a sequence, should they exist, are unique. Also see a function version.

Proof. Suppose that $(x_n) \to x$ and $(x_n) \to y$. Let $\epsilon > 0$. By assumption there exists $N_x := N(x) \in \mathbb{N}$ [depending on (x_n) and x] such that

$$n \geqslant N_x \implies d(x_n, x) < \frac{\epsilon}{2}. \tag{1}$$

Similarly, there exists $N_y := N(y) \in \mathbb{N}$ such that

$$n \geqslant N_y \implies d(x_n, y) < \frac{\epsilon}{2}. \tag{2}$$

Therefore, if $n \geqslant \max(N_x, N_y)$, (1) and (2) both hold, so

$$d(x, y) < d(x, x_n) + d(x_n, y) = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Since ϵ is arbitrary, $d(x, y) < \epsilon$ implies $d(x, y) = 0$, i.e., $x = y$ and limits are unique. *Heuristically, if two limits are apart, how can the tail of a sequence be arbitrarily close to both limits at the same time?* □

[3] Since ∞ is not an element of $\mathbb{R}$, the distance between two points is always finite.

Example 3.6. The sequence $(x_n)_{n\geqslant 1}$ defined by $x_n := 1/n$ converges to 0 in $(\mathbb{R}, \|\cdot\|)$.
The proof is a one-liner: given $\epsilon > 0$, pick $N > 1/\epsilon$; if $n \geqslant N$ then $d(x_n, 0) = x_n = 1/n \leqslant 1/N < \epsilon$. □

However, the choice of metric space matters! Consider the same example but this time with $((0,1), \|\cdot\|)$ as the metric space. The number 0 does *not* lie in this open interval, and we say (x_n) **diverges** in this space!
*To prove this rigorously, we know that (x_n) converges to 0 in the metric space $(\mathbb{R}, \|\cdot\|)$. Since our metric space $((0,1), \|\cdot\|)$ **inherits** the metric from $(\mathbb{R}, \|\cdot\|)$, if the sequence converges, by uniqueness of limits, it must converge to 0, which is outside $(0,1)$. Hence (x_n) does not converge in our metric space.*

Example 3.7: Convergence in $\mathbb{R}^k$ ⇔ Component-Wise Convergence. In $(\mathbb{R}^k, \|\cdot\|)$, a sequence $x^{(n)} \to x$[4] if and only if $x_i^{(n)} \to x_i$ for all $1 \leqslant i \leqslant k$, i.e., a sequence converges to a limit if and only if the i^{th} component of each term of the sequence converges to that of the limit.
Future reference: Compactness of k-cells, Example 7.10

Proof. We first prove $\Longrightarrow$. Let $\epsilon > 0$ be given and pick some $i \in [1, k]$. Notice that

$$|x_i^{(n)} - x_i| \leqslant |x^{(n)} - x| < \epsilon$$

for sufficiently large[5] n. The inequality is because $y_i^2 \leqslant y_1^2 + \ldots + y_i^2 + \ldots + y_n^2 = \|y\|^2$. It follows that the **tail** of $(x_i^{(n)})$ can be made arbitrarily close to x_i, so $x_i^{(n)} \to x_i$. This proves $\Longrightarrow$.
For $\Longleftarrow$, also let $\epsilon > 0$ be given. By assumption, for all $1 \leqslant j \leqslant k$, $x_j^{(n)}$ converges to x_j. Therefore, for each j, there exists $N_j := N(j) \in \mathbb{N}$ such that

$$|x_j^{(n)} - x_j| < \sqrt{\epsilon/k} \text{ for all } n \geqslant N_j. \tag{1}$$

(*We shall see very soon where $\sqrt{\epsilon/k}$ comes from. In real analysis, it is often helpful to think backwards; once we are able to bound our target of interest by a multiple/function of ϵ, we can go back and adjust our initial bound accordingly. In this example, we will see if we start with ϵ we would end up with $k \cdot \epsilon^2$, so if we start with $\sqrt{\epsilon/k}$ we would end up with precisely ϵ, which makes things look nicer... and mathematicians like this.*)
Therefore, if we define $N := \max_{1\leqslant j\leqslant k} N_j$, (1) holds for every single component, and thus

$$|x^{(n)} - x|^2 = \sum_{i=1}^{k} |x_j^{(n)} - x_j|^2 < k \cdot \frac{\epsilon}{k} = \epsilon.$$

This shows that the tail of $x^{(n)}$ can be made arbitrarily close to x, which proves the convergence. □

Beginning of Feb. 3, 2021

Definition 3.8: Various Definitions in Metric Spaces

Let X be a metric space and let $E \subset X$.

(1) **Neighborhoods**: $N_r(P) := \{q \in X : d(p,q) < (r > 0)\}$ is the neighborhood[6] of $p \in X$ of radius r. *This gives an equivalent definition for convergence of a sequence*:

$$p_n \to p \text{ if for all } \epsilon > 0 \text{ there exists } N \in \mathbb{N} \text{ such that } n \geqslant N \implies p \in N_\epsilon(p). \tag{Eq.3.1}$$

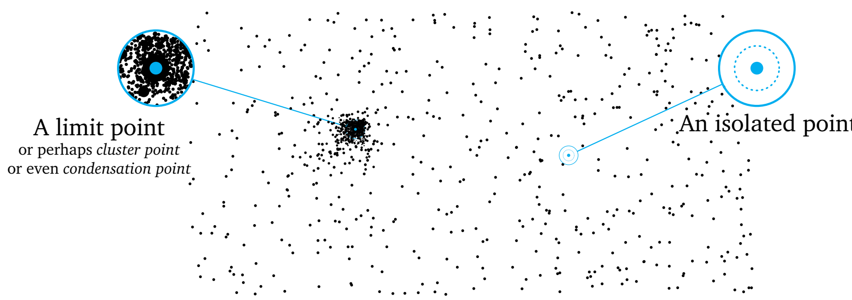
[4]Since we are also used to denote components using subscripts, we use $x^{(n)}$ to denote the elements in this sequence. An exception.
[5]We could, of course, give a more rigorous proof, completely following the (ϵ-N) language. However, that approach would significantly decrease the comprehensibility of the proof and, like I mentioned in "Chapter 0", I shall avoid that whenever possible.

(2) **Interior points**: $p \in E$ is an interior point of E if, for some $r > 0$, we have $N_r(p) \subset E$. *In other words, p is not on the "boundary" of E.*

(3) **Open sets**: we say $E \subset X$ is open *in* X if every point $p \in E$ is an interior point of E. *In other words, E has no "boundary".*

(4) **Limit points**: we say $p \in X$ is a limit point of E if there exists a sequence $(q_n) \subset E$, whose terms are not p, that converges to p.

(5) **Isolated points**: we say $p \in X$ is an isolated point if it is not a limit point of E.



(6) **Closed sets**: we say $E \subset X$ is closed if it contains all its limit points. *Heuristically, it contains all its "boundaries".* Future reference: Closed subsets of compact sets are compact

(7) **Complement**: $E^c := X - E$ is called the complement[7] of E (in X).

(8) **Perfect set**: we say E is perfect if it is closed and every point in E is a limit point. Future reference: Perfect sets are uncountable

(9) **Bounded sets**: we say E is bounded if it is contained in *some* neighborhood, i.e., there exists $p \in X$ and $r > 0$ such that $E \subset N_r(p)$. Future reference: Compact sets are bounded, Heine-Borel theorem

(10) **Closure**: $\overline{E} := E \cup \{\text{limit points of } E\}$ is called the closure of E. (*Note that this set is always closed, and a closed set is equal to its closure. Thus $\overline{\overline{E}} = \overline{E}$ for any E.*)

(11) **Dense sets**: we say E is dense in X if $\overline{E} = X$. *Using closure and limits, this means that any $x \in X$ can be approximated arbitrarily close by some $e \in E$.*

Due to limit time, we will omit some basic examples which could provide helpful insights. See Rudin's book if you are in need of them.

Lemma 3.9: Neighborhoods are Open

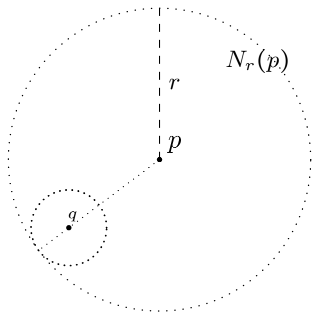
Every neighborhood $N_r(p)$ is open. Future reference: Open set condition

Proof. We need to show that every $p \in N_r(p)$ is an interior point. If we can find some neighborhood of p that is

[6]In a metric space, it is also common to define the **open ball** $B_r(p)$ or $B(p,r) := \{q \in X : d(p,q) < r\}$. The word *open* becomes clear once we define it in (3) and prove $N_r(p)$ is open in the next lemma. IMO, the word *ball* provides a more intuitive definition than *neighborhood*.

[7]Various notations for *the difference between sets* exist, for example $A \setminus B$. We will use Rudin's notation, $A - B$, to denote $\{x : x \in B \text{ but } x \notin A\}$.

entirely contained in E then we are done.



We simply need to use the triangle inequality: if $d(w,q) < r - d(p,q)$ then

$$d(w,p) \leqslant d(w,q) + d(q,p) < r - d(p,q) + d(q,p) = r. \tag{1}$$

Also notice that $r - d(p,q) > 0$ since $q \in N_r(p)$ implies $d(p,q) < r$, so the neighborhood

$$N_{r-d(p,q)}(q)$$

is well-defined. Using (1) we see that $N_{r-d(p,q)}(q) \subset N_r(p)$. This proves the claim. □

Theorem 3.10: Openness is Dual to Closedness

A is open in X if and only if A^c is closed (in X). Future reference: Closed set condition

Proof. We first prove $\Longleftarrow$. Suppose A^c is closed; we want to show that if $x \in A$ then x is an interior point of A. Clearly $x \notin A^c$, x cannot be a limit point of A^c (for it it were, closedness implies $x \in A^c$). We claim that there exists some $r > 0$ such that $N_r(x) \cap A^c = \varnothing$.
If this is false, then for each $\epsilon > 0$ we are able to find a corresponding $q_\epsilon \in A^c \cap N_\epsilon(x)$. Using this, we define

$$\epsilon_1 := 1, \quad \epsilon_2 := 1/2, \quad \ldots, \quad \epsilon_n := 1/n. \tag{Eq.3.2}$$

and we are able to construct sequence $(q_n)_{n\geqslant 1}$ such that $d(x,q_n) < 1/n$. We see that as $n \to \infty$, $d(x,q_n) \to 0$, i.e., $q_n \to x$. We have just constructed a sequence in A^c that converges to $x \notin A^c$, contradiction! Therefore there must exist some $r > 0$ such that $N_r(x) \cap A^c = \varnothing$, i.e., $N_r(x) \subset A$, and thus A is open. To justify the last step (*I have no idea why this seeming trivial equation deserves a special label*):

$$N_r(x) = N_r(x) \cap X = (N_r(x) \cap A) \cup \underbrace{(N_r(x) \cap A^c)}_{=\varnothing} \subset A. \tag{Eq.3.3}$$

Remark. *Using $\epsilon_n := 1/n$ is a very common approach for proofs, in particular when we need to construct a sequence while also letting $\epsilon \to 0$. We shall informally call this the "**proof by the $1/n$ sequence**", and we will likely encounter such method in the future again.*
Future reference: Characterization of supremum, Lebesgue's numnber lemma, convergence at a point (Heine)

We now prove $\Longrightarrow$ and suppose A is open. Assume x is a limit point of A^c. By definition there exists $(q_n) \in A^c$. We want to show that $x \in A^c$. Suppose for contradiction that $x \in A$, so there exists $r > 0$ such that $N_r(x) \subset A$. On the other hand, the (ϵ-N) definition of convergence implies that sufficiently late terms in (q_n) lie in $N_r(x)$. These points lie in both A and A^c, which is absurd! □

Example 3.11: Characterization of Suprema & Infima. Let $(X,d) := (\mathbb{R}, \|\cdot\|)$ and let $A \subset X$. Then we have the following characterization of supremum and infimum:

(1) $a = \sup A$ if and only if (a is both an upper bound and a limit point of A) or a is an isolated point of A.

(2) $b = \inf A$ if and only if (b is both a lower bound and a limit point of A) or b is an isolated point of B.

Future reference: Characterization of limsup, Theorem 8.5, connected subsets of $\mathbb{R}$

Proof. We will only prove the case for supremum; the infimum case is analogous, and we will assume that a is not an isolated point of A (if it were, the claim is trivial).

For $\Longrightarrow$, let $a = \sup A$. It is trivially true that a is an upper bound of A, so it remains to find a sequence[8] $(a_n)_{n\geqslant 1} \subset A$ that converges to a. Indeed, we can use the "$1/n$ proof" again. Let $\epsilon_1 := 1$ and let $\epsilon_n := 1/n$. By definition of supremum, there exists a_1 satisfying $a_1 > a - 1 = a - \epsilon_1$; likewise, there also there exists a_{n+1} satisfying $a_{n+1} > a - \epsilon_{n+1}$ and $a_{n+1} > a_n$ (using definition of supremum twice to guarantee that (a_n) is *strictly increasing*). It is clear that $a_n \to a$.

For the *converse* (meaning the $\Longleftarrow$ direction), suppose a is an upper bound and a limit point of A but $a \neq \sup A$. By Definition 1.6.2 a is not the least upper bound, and there exists a smaller upper bound $a - \epsilon$ for some $\epsilon > 0$. But since A is a limit point, there exists $(a_n)_{n\geqslant 1}$ converging to A, so late enough terms in this sequence is $< \epsilon$ away from a, meaning that they are created than $a - \epsilon$, our hypothesized supremum. Contradiction. ☐

Theorem 3.12: Openness & Closedness under Union and Intersection

(1) An <u>arbitrary</u> union of open sets is open, i.e., if A_α's are open then $\bigcup_{\alpha\in I} A_\alpha$ is open. *In particular the index set can be uncountable.*

(2) A <u>finite</u> intersection of open sets is open, i.e., if A_α's are open and I is a finite index set then $\bigcap_{\alpha\in I} A_\alpha$ is open. *We will show examples of countable/uncountable intersection of open sets that is closed.*

(3) Analogously, an <u>arbitrary</u> intersection of closed sets is closed, and a <u>finite</u> union of closed sets is closed. *Again, there exist examples where a countable/uncountable union of closed sets gives us an open set.*

Future reference: Compact sets are closed, intersection of nonempty nested compact sets is nonempty

Beginning of Feb. 5, 2021

Proof. (1) This follows directly from the definitions.

(2) Suppose that $A_1, \ldots, A_n$ are open and fix $x \in \bigcap_{i=1}^{n} A_i$. It follows that $x \in A_i$ for all $1 \leqslant i \leqslant n$. By definition, for all i there exists $rr_i := (i) > 0$ such that

$$N_{r_i}(x) \subset A_i.$$

Taking $r := \min\{r_1, \ldots, r_n\}$, we see that $N_r(x) \subset A_i$ for all i so it lies in the intersection. *Note that we can take the minimum because there are only finitely many sets; otherwise, the infimum may well be 0 in which case our proof becomes invalid.*

(3) The arbitrary intersection case follows from definition once again, so we will only prove the finite union case. If $A_1, \ldots, A_n$ are closed and we pick a sequence $(a_n)_{n\geqslant 1} \in \bigcup_{i=1}^{n} A_i$ that converges in the ambient space. By the *Pigeonhole principle*, there exists at least one A_i such that *infinitely many* terms of $(a_n)_{n\geqslant 1}$ lie in A_i. Therefore, since this subsequence is a sequence itself, by the closeness of A_i, it converges to some limit a.

[8] As shown by this example, it is not so interesting to construct a seequence that converges to a by simply setting its tail to be a. Unless in proofs, we take it for granted, from now on, that $(a_n) \to a$ implies $a_n \neq a$.

Since limits are unique, a must also be the limit of the mother sequence (a_n), and this shows that the union is closed.

Alternatively, this proof is a one-liner using De Morgan's law and (2):

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \overbrace{\bigcap_{i=1}^{n}}^{\text{finite}} \underbrace{A_i^c}_{\text{open}} = \text{open} \implies \bigcup_{i=1}^{n} A_i \text{ is closed.}$$ □

Example 3.13. For counterexamples of "arbitrary intersection of open sets is open" or "arbitrary union of closed sets is closed", consider the following:

$$\bigcap_{n=1}^{\infty}\left(-\frac{1}{n}, 1+\frac{1}{n}\right) = [0,1] \qquad \bigcup_{n=1}^{\infty}\left[\frac{1}{n}, 1-\frac{1}{n}\right] = (0,1).$$

We now introduce one of the most important concept (and a *really nice* property) in analysis.

2.4 Compact Sets

Definition 3.14: Open Cover & Covering Compactness

Let $K \subset X$.[9]

(1) An **open cover** of K is any collection of open sets $\{A_\alpha\}_{\alpha\in I}$ such that $K \subset \bigcup_{\alpha\in I} A_\alpha$.

(2) We say K is (**covering**) **compact** if every open cover of K contains a finite subcover, i.e., $K \subset \bigcup_{k=1}^{n} A_{\alpha_k}$.

Theorem 3.15: Compact ⇒ Closed

If K is compact then K is closed. Future reference: Heine-Borel theorem, Theorem 8.7

Proof. It suffices to show K^c is open by Theorem 3.12. Fix $p \in K^c$; we need to find some neighborhood $n_r(p)$ that is contained in K^c. Clearly

$$K = \bigcup_{q\in K}\{q\} \subset \bigcup_{q\in K} N_{d(p,q)/2}(q),$$

so by the compactness of K, there exist $q_1, \dots, q_n \in K$ such that

$$K \subset \bigcup_{i=1}^{n} N_{d(p,q_i)/2}(q_n).$$

Now we define $r := \min_{1\leqslant i\leqslant n} d(p,q_i)/2$. Since $p \notin K$, it is obvious that $r > 0$. We claim $N_r(p) \subset K^c$, since

$$(N_r(p) \cap K) \subset N_r(p) \cap \left(\bigcup_{i=1}^{n} N_{d(p,q_i)/2}(q_n)\right) = \bigcup_{i=1}^{n} \underbrace{N_r(p) \cap N_{d(p,q_i)/2}(q_n)}_{=\varnothing} = \varnothing$$

This completes the proof. □

Alternately, there is another simple proof without using K^c: if K is not closed, there exist $(a_n) \subset K$ converging to some $a \in A - K$. Consider the covering

$$U_n := \{x \in X : d(x,a) > 1/n\}, \qquad n \in \mathbb{N}$$

which indeed covers K and thus admits a finite subcover. But it cannot have one; otherwise how could $(a_n) \to a$ when the sequence is at least some positive distance away from a?

Below is the essence of compactness (in analysis) — it gives us a convergent sequence out of nowhere, like magic!

Theorem 3.16: Covering Compactness ⇔ Sequential Compactness

A set K in a metric space is compact if and only if every sequence in K has a convergent subsequence, i.e., for all $(x_n)_{n\geqslant} \subset K$, there exists a subsequence $(x_{n_k})_{k\geqslant 1}$ that converges to some $x \in K$. The second property is called **sequential compactness**. (*This theorem fails, however, in some more general topological spaces.*)

[9] It is customary to name a compact set as K.

Proof. We first show $\Longrightarrow$. Suppose for contradiction that K is covering compact but a sequence $(x_n)_{n\geqslant 1} \subset K$ has no convergent subsequence. *Immediately we know that this sequence has more than one element.* Therefore, any $y \in K$ is not the limit of $(x_n)_{n\geqslant 1}$. *In other words, (x_n) cannot be arbitrarily close to y, unless y appears in the sequence, in which case the shortest distance is 0 but the second shortest distance is not arbitrarily small.* Therefore there exists $r_y := r(y) > 0$ such that

$$N_{r_y}(y) \cap \{x_n\} = \begin{cases} \varnothing & \text{if } y \notin \{x_n\} \\ \{y\} & \text{if } y \in \{x_n\} \end{cases}$$

(where $\{x_n\}$ denotes the set of all points in the sequence (x_n)). *The second case is bothersome, but it will not affect our proof, which focuses on convergence of the entire sequence*[10]. Since

$$K = \bigcup_{y\in K} \{y\} \subset \bigcup_{y\in K} N_{r_y}(y),$$

by covering compactness the open cover on the RHS admits a finite subcover: there exist $y_1, \dots, y_n \subset K$ such that

$$K \subset \bigcup_{k=1}^{n} N_{r_{y_k}}(y_k).$$

This gives a contradiction, for the intersection between the new RHS and $\{x_n\}$ is either $\varnothing$ or $\{y\}$, and in either case there are a *lot* of elements in the sequence $(x_n)_{n\geqslant 1}$ missing in this finite cover. This proves $\Longrightarrow$.

For the converse, let K be sequentially compact and let $\bigcup_{\alpha\in I} A_\alpha$ be any open cover of A. We need to extract a finite subcover, but first, we need to prove **Lebesgue's number lemma**:

> Let A_α's be arbitrarily chosen above. If K is compact then there exists a $\delta > 0$, called the **Lebesgue number**, such that $N_\delta(x) \subset$ some A_α for all $x \in K$. *Note that the Lebesgue number depends solely on K, not our choice of open cover!*

Proof of Lebesgue's number lemma. Suppose this claim is false. We use the "$1/n$ proof": for all $n \in \mathbb{N}$, there exists $x_n \in K$ such that $N_{1/n}(x_n)$ is not a subset of any A_α. *If some $\delta > 0$ works for all $x \in K$ then smaller δ's also work; our assumption that no such δ exists means in particular that all δ's of form $1/n$ fail the condition.* Consider the sequence $(x_n)_{n\geqslant 1}$. By sequential compactness, there exists a subsequence $(x_{n_k})_{k\geqslant 1}$ converging to some $x \in K$. In particular, x belong so some open set A_{α_0} as the A_α's cover K. Since A_{α_0} is open, there exists $\epsilon > 0$ such that $N_\epsilon(x) \subset A_{\alpha_0}$. Now we have a contradiction —

On one hand, x_{n_k} is approaching x, while on the other hand, as k increases (or n_k increases), by our assumption, we can find smaller and smaller neighborhoods of x_{n_k} that belongs to no A_α, and since $N_\epsilon(x) \subset A_{\alpha_0}$, this implies that these increasingly tiny neighborhoods are not contained in $N_\epsilon(x)$. Therefore, there exists sufficiently large k such that $1/n_k < \epsilon/2$ and $d(x_{n_k}, x) < \epsilon/2$. Then, for all $x \in N_{1/n_k}(x_{n_k})$, i.e., all y with $d(x_{n_k}, y) < 1/n_k$, we have

$$d(x,y) \leqslant d(x, x_{n_k}) + d(x_{n_k}, y) < \frac{\epsilon}{2} + \frac{1}{n_k} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

so $N_{1/n_k}(x_{n_k}) \subset N_\epsilon(x) \subset A_{\alpha_0}$, contradiction indeed. END OF PROOF OF LEBESGUE'S NUMBER LEMMA

[10] Pugh's book provides a workaround: we can instead say *there exists an $r_y := r(y) > 0$ such that $a_n \in N_{r_y}(y)$ only finitely often.* Then, along with the finite cover derived later, only finitely many a_n's can appear in $\bigcup_{k=1}^{n} N_{r_{y_k}}(y_k)$, a superset of K. Contradiction.

Now we are back to the main proof. The next step is to notice that, given $\delta > 0$ (from the lemma above), there exists $x_1, ..., x_n \in K$ such that

$$K \subset \bigcup_{k=1}^{n} N_\delta(x_k). \tag{Eq.3.4}$$

(*We say K is* ***totally bounded***.) *Otherwise, fix $x_1 \in K$. Since we cannot cover K using finitely many δ-neighborhoods, in particular we cannot cover K by $N_\delta(x_1)$ only. Therefore there exists $x_2 \in K - N_\delta(x_1)$. Again, since we cannot cover K using only $N_\delta(x_1) \cup N_\delta(x_2)$, there exists $x_3 \in K - (N_\delta(x_1) \cup N_\delta(x_2))$. Inductively, we can define a sequence $(x_n)_{n\geqslant 1}$ such that*

$$x_n \in K - \bigcup_{i=1}^{n-1} N_\delta(x_i).$$

By sequential compactness, there exists a subsequence $(x_{n_k})_{k\geqslant 1}$ that converges to some $x \in K$. Therefore, there exists $M \in \mathbb{N}$ such that $d(x_{n_k}, x) < \delta/2$ for all $k \geqslant M$ (or $n_k \geqslant M$, whatever). Then, for such k's,

$$d(x_{n_k}, x_{n_{k+1}}) \leqslant d(x_{n_k}, x) + d(x, x_{n_{k+1}}) < \frac{\delta}{2} + \frac{\delta}{2} = \delta, \tag{Eq.3.5}$$

contradicting the assumption that $x_{n_{k+1}} \notin N_\delta(x_{n_k})$ (i.e., we want terms in (x_n) to be at least δ apart, whereas the convergence of (x_{n_k}) renders this impossible). This proves Equation 3.4.

We have also completed the proof of $\Longleftarrow$, as each $N_\delta(x_k)$ is in some A_α by the property of the *Lebesgue number*. Hence $\{A_\alpha\}_{\alpha\in I}$ admits a finite subcover! □

We present an immediate consequence of the above theorem:

Example 4.1: Closed Subsets of Compact Sets are Compact. If K is compact and $F \subset K$ closed, then F is compact.
Future reference: Intersection of compact sets, Heine-Borel theorem, Perfect sets are uncountable, Theorem 8.7, Dini's Theorem

Proof. Let $(x_n)_{n\geqslant 1} \subset F \subset K$. It follows from the compactness of K that there exists a subsequence $(x_{n_k})_{k\geqslant 1}$ that converges to some $x \in K$. Since F is closed it contains its limit points (Definition 3.8.6), so $x \in F$, and (x_n) admits a convergent subsequence in F, i.e., F is compact. □

Example 4.2: Compact ⇒ Bounded. If K is compact then K is bounded.
Future reference: Heine-Borel theorem

Proof. This directly follows from Equation 3.4 (right above) and the triangle inequality: let $K \subset \bigcup_{k=1}^{n} N_\delta(x_k)$ as guaranteed by (Eq.3.4). Fix x_1, and let $j \in [1, n]$ be the integer such that

$$d(x_1, x_j) = \max_{1\leqslant k\leqslant n} d(x_1, x_k).$$

Let $y \in K$ be arbitrarily chosen; it belongs to some $N_\delta(x_k)$. Thus

$$d(x_1, y) \leqslant d(x_1, x_k) + d(x_k, y) < d(x_1, x_j) + \delta < \infty,$$

so all of K is contained in $N_r(x_1)$ where $r := d(x_1, x_j) + \delta$. By Definition 3.8.9 K is bounded. □

Theorem 4.3: Intersection of Compact Sets

Let $\{K_\alpha\}_{\alpha\in I}$ be a collection of (finite, countable, or uncountable) compact sets such that any finite intersection is nonempty, i.e.,

$$\bigcap_{k=1}^{n} K_{\alpha_k} \neq \varnothing \text{ for all } n \text{ and } \{\alpha_1, ..., \alpha_n\} \subset I.$$

Then $\bigcap_{\alpha\in I} K_\alpha \neq \varnothing$. Furthermore, the intersection is compact. *Note that this is stronger than Theorem 3.12.*

Proof. Theorem 3.12 implies that $\bigcap_{\alpha\in I} K_\alpha$ is closed, and a closed subset of compact sets is compact. (*Note that this holds even if the intersection is empty – the empty set is trivially compact.*) It remains to show nonemptyness.

Suppose for contradiction that $\bigcap_{\alpha\in I} K_\alpha = \varnothing$. Taking complement gives

$$X = \Big(\bigcap_{\alpha\in I} K_\alpha\Big)^c = \bigcup_{\alpha\in I} K_\alpha^c. \tag{1}$$

Since K_α is compact and in particular closed, K_α^c is open. Fix some $\alpha_0 \in I$. Clearly $K_{\alpha_0} \subset X$, so (1)'s RHS is an open cover of K_{α_0}. By compactness, there exists a finite subcover, i.e., there exist $\{\alpha_1, ..., \alpha_n\} \subset I$ such that

$$K_{\alpha_0} \subset \bigcup_{k=1}^{n} K_{\alpha_k}^c. \tag{2}$$

Then,

$$\begin{aligned} K_{\alpha_0} \cap K_{\alpha_1} \cap ... \cap K_{\alpha_n} &\subset \Big(\bigcup_{k=1}^{n} K_{\alpha_k}^c\Big) \cap \Big(\bigcap_{k=1}^{n} K_{\alpha_k}\Big) && \text{(using (2))} \\ &= \bigcup_{k=1}^{n} \Big(K_{\alpha_k}^c \cap \underbrace{\bigcap_{j=1}^{n} K_{\alpha_j}}_{\in K_{\alpha_k}}\Big) = \bigcup_{k=1}^{n} \varnothing = \varnothing, \end{aligned}$$

contradicting the problem. This completes the proof. □

Corollary 4.4: Nested Compact Sets

Let $\{K_n\}_{n\geqslant 1}$ be a sequence of nonempty compact sets such that $K_{n+1} \subset K_n$. Then $\bigcap_{n=1}^{\infty} K_n \neq \varnothing$, i.e., *the infinite intersection of a sequence of decreasing nonempty compact sets is nonempty*.

Future reference: Perfect sets are uncountable, Dini's Theorem

Beginning of Feb. 10, 2021

Theorem 4.5: $[a, b]$ is Compact

Every closed interval $[a, b] \subset \mathbb{R}$ is compact.

Future reference: Compactness of k-cells

Proof. WLOG (*without loss of generality*) take $[a, b]$ to be $[0, 1]$. We will prove the claim using sequential compactness. Let $(x_n)_{n\geqslant 1} \subset [a, b]$ be a sequence in this interval; our goal is to find a convergent subsequence.

We define a set $\mathcal{C} := \{c \in [0, 1] : x_n < c \text{ only for finitely many } x_n's\}$. *Clearly this is a set*. Also observe that

(1) $\mathcal{C}$ is nonempty: no $x_n < 0$, so "only finitely time" trivially holds for 0, and thus $\mathcal{C}$.

(2) $\mathcal{C}$ is bounded above: clearly the sequence is bounded by 1, so any $c > 1$ is going to be greater than any x_n, and for such c's, "$x_n < c$" would happen infinitely many times.

Therefore by the LUBP property, $\mathcal{C}$ has a supremum, and we define $x := \sup \mathcal{C}$ and $x \in [0,1]$ ($0 \leqslant x$ *because* $0 \in \mathcal{C}$, *and* $x \leqslant 1$ *because* 1 *is already an upper bound*).
The remainder of the proof is to show that there exists a subsequence $(x_{n_k})_{k \geqslant 1}$ *that converges to* x. Suppose for contradiction that such subsequence does not exist. Then there exists $\epsilon > 0$ such that $x_n \in (x - \epsilon, x + \epsilon)$ only for finitely many x_n's. (*This coincides with my footnote no.11.*) Since only finitely many x_n's appear in $[a, x]$ and also finitely many x_n's appear in $[x, x + \epsilon)$, we conclude that only finitely many x_n's appear in $[a, x + \epsilon)$, so $x + \epsilon \in \mathcal{C}$, contradicting the assumption that $x = \sup \mathcal{C}$. Therefore, $[a, b]$ is compact. □

Definition 4.6: k-Cells

A set $[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_k, b_k] \subset \mathbb{R}^k$ for $a_i < b_i$, $1 \leqslant i \leqslant k$, is a ***k*-cell** in $\mathbb{R}^k$, where $\times$ denotes the *Cartesian product*. *Think of "boxes" or "k-dimensional intervals".*

Corollary 4.7: k-Cells are Compact

k-cells (in $(\mathbb{R}^k, \|\cdot\|)$) are compact.

Proof. This is a standard approach when proving (sequential) compactness: "sub-sub-sub-...-subsequences!"
Let $(x^{(n)})_{n \geqslant 1} \subset [a_1, b_1] \times \ldots \times [a_k, b_k] \subset \mathbb{R}^k$ be a sequence where $x^{(n)} = (x_1^{(n)}, x_2^{(n)}, \ldots, x_k^{(n)})$[11]. Using the compactness of $[a_1, b_1] \subset \mathbb{R}$, we are able to extract a subsequence $(x^{(n,1)})$ of $(x^{(n)})$ such that the first component of each term converges, i.e., $x_1^{(n,1)} \to x_1$ for some $x_1 \in \mathbb{R}$.
Now, looking at the second component of each term in $(x^{(n,1)})$, we immediately see that it forms another sequence, bounded by $[a_2, b_2]$, in $\mathbb{R}$. Using compactness of $[a_2, b_2]$ again, we are able to extract a sub-subsequence $(x^{(n,2)}) \subset (x^{(n,1)}) \subset (x^{(n)})$ that converges, say to some x_2, in the second component. Notice that, since $(x^{(n,2)})$ is a subsequence of $(x^{(n,1)})$, it *inherits* the convergence in the first component as well! (*This is the key fact*.) Therefore $(x^{(n,2)})$ converges in *both* the first and second component.
Iterating this process, we are able to arrive at a sub-sub-...-subsequence $(x^{(n,k)})$ such that it converges in *all* k components, i.e., it converges in $\mathbb{R}^n$ (Example 3.7)! Of course, a sub-sub-...-subsequence is still a subsequence, so we conclude that k-cells are compact. □

Theorem 4.8: Heine-Borel Theorem

In $\mathbb{R}^k$, a set K is compact if and only if it is closed and bounded. *The $\Rightarrow$ direction is true in any metric space, whereas the converse is not; counterexample: $\mathbb{N}$ equipped with discrete metric.*
Future reference: Perfect sets are uncountable, Arzelá-Ascoli Theorem

Proof. $\Longrightarrow$ follows from Theorem 3.15 and Example 4.2.

[11] I adopted the earlier notation $x^{(n)}$ here. The notation used in this proof will slightly differ from what was actually written in the lecture.

For $\Longleftarrow$, assume K is closed and bounded. By boundedness it is contained in some k-cell. *Heuristically we can of course contain a bounded set using a big box. For a rigorous proof: by definition $K \subset N_r(p)$ for some $r > 0, p \in \mathbb{R}^n$, so of course*

$$K \subset [p_1 - r, p_1 + r] \times [p_2 - r, p_2 + r] \times \ldots \times [p_k - r, p_k + r].$$

Then since a closed subset of a compact set is compact, K is compact. ☐

Theorem 4.9: Bolzano-Weierstraß Theorem[12]

Let $K \subset \mathbb{R}^n$ be closed and bounded. Then every sequence $(x_n)_{n \geqslant 1} \subset K$ has a convergent subsequence in K. Future reference: $\mathbb{R}^k$ is complete, Characterization of limsup, Theorem 5.8, minimization problems, proof of the Arzelá-Ascoli Theorem

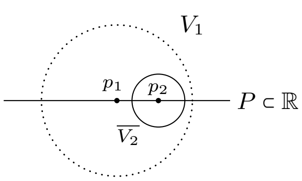
Proof. Immediate from Heine-Borel. ☐

Theorem 4.10: Nonempty Perfect Sets are Uncountable

Let $P \subset \mathbb{R}^k$ be nonempty and perfect. Then P is uncountable. *In particular $\mathbb{R}$ is uncountable.*[13]

Proof. First note that P cannot be finite, for otherwise each point would be an isolated point (in particular, no limit points, which must be limit of infinite sequences). Now, for contradiction, we suppose P is countable and let $\{p_n\}_{n \geqslant 1}$ be an enumeration of P.
Let $V_1 := N_1(p_1)$, and define V_{n+1} iteratively such that

$$\overline{V_{n+1}} \subset V_n, \qquad p_n \notin \overline{V_{n+1}}, \qquad \text{and} \qquad V_{n+1} \cap P \neq \varnothing.$$



Since P is perfect, each neighborhood V_n contains infinitely many points of P (again recall that there has to exist some sequence converging to the center of the neighborhood), so we can always define V_n's as such, and we can always find sufficiently small radius for V_{n+1} because there are only finitely many V_i's we need to take care of (in particular, to ensure there is no intersection).

Now we define $K_n := \overline{V_n} \cap P$. Notice that $\overline{V_n}$ is closed and bounded, so Heine-Borel implies $\overline{V_n}$ is compact. On the other hand P is closed by definition. Therefore, K_n is a closed subset of a compact set and is compact. In addition, each K_n is nonempty and $K_{n+1} \subset K$ as $\overline{V_{n+1}} \subset V_n \subset \overline{V_{n+1}}$. Also, by construction, $p_n \notin \overline{V_{n+1}}$ so $p_n \notin K_{n+1}$. Now consider the intersection

$$K := \bigcap_{n=1}^{\infty} K_n.$$

Is $p_1 \in K$? No, because $p_1 \notin K_2$. Is $p_2 \in K$? No, because $p_2 \notin K_3$. Is $p_k \in K$? No! Assuming $\{p_n\}_{n \geqslant 1}$ enumerates P, we conclude that

$$P \cap K = \varnothing.$$

However, a nested sequence of nonempty compact sets cannot have empty intersection. Contradiction. Therefore, P must be uncountable. ☐

Beginning of Feb. 12, 2021

[12]This theorem was initially proved in 1817, long before that of the Heine-Borel theorem (1852 by Dirichlet and 1890s by many more).

[13]We could have proved this using *Cantor's diagonalization*; the idea is similar to the infinite string of binary numnbers.

Corollary 4.11: $[a, b]$ is Uncountable

Any interval $[a, b]$ (with $a < b$ of course) is uncountable.

In fact, there exist perfect sets in $\mathbb{R}$ which do not contain any interval. One such example is the ***Cantor Set****.*[14] Below is an example of the **middle-thirds Cantor Set**. Starting with $[0, 1]$, we remove the middle $1/3$ of this segment, and then iteratively remove the middle $1/3$ of each remaining line segments. In the end we obtain the Cantor Set. This is a *compact, nonempty, perfect, totally disconnected, and uncountable* set.

We will see Cantor Set again when we talk about the Devil's Staircase / Cantor function.

[14] Siegel's 425a (which I took) spent several lectures on the Cantor Set. I am surprised that this 425a almost skips it.

Chapter 3

Sequences and Series

3.1 Convergent Sequences

Lemma 4.12: Convergent ⇒ Bounded

Convergent sequences are bounded. Future reference: Theorem 13.16, minimization problems

Proof. Let (X, d) be a metric space and suppose $(x_n)_{n\geqslant 1} \to x$ for some $x \in X$.[1] Pick, for example, $\epsilon = 1$. Then there exists $N \in \mathbb{N}$ such that all but the first N terms of (x_n) are in $N_1(x)$ [neighborhood]. It follows that the remaining finite terms cannot be arbitrarily far away from x: if we define

$$x := \max\{1, d(x_1, x), ..., d(x_N, x)\} + 1^2$$

then $x_n \in N_r(x)$ for all x_n. This completes the proof of boundedness. □

Example 4.13: Operations on Two Sequences. Let $(x_n), (y_n) \subset \mathbb{R}^k$ and $(\beta_n) \in \mathbb{R}$ be such that $(x_n) \to x$, $(y_n) \to y$, and $(\beta_n) \to \beta$. Then:

(1) Sum of limit is limit of sums: $x_n + y_n \to x + y$,

(2) (Dot) product of limit is limit of (dot) products: $x_n \cdot y_n \to x \cdot y$,

(3) (Scalar) product of limit is limit of (scalar) products: $\beta_n x_n \to \beta x$, and

(4) Convergence is preserved under exponentiation: $\beta_n^m \to \beta^m$.

Future reference: Corollary 7.4

Proof. (1) Immediate by a standard $\epsilon/2$ argument (one on x_n and the other y_n) with Δ-inequality.

[1] Now that we are familiar with the notation $(x_n)_{n\geqslant 1}$ enough, we will drop the subscript and only write (x_n) from now on.

[2] The "+1" can be replaced by any positive number; the mere purpose is to make what we want lie completely inside the *open* neighborhood, and since open neighborhoods don't contain their boundaries, we need to make the radius *slightly* larger.

(2) A classic "*add and subtract*" trick, along with Cauchy-Schwarz:

$$\begin{aligned}|x_n \cdot y_n - x \cdot y| &= |x_n \cdot y_n - x_n \cdot y + x_n \cdot y - x \cdot y| \\ &\leqslant |x_n \cdot y_n - x_n \cdot y| + |x_n \cdot y - x \cdot y| \\ &= |x_n \cdot (y_n - y)| + |(x_n - x) \cdot y| \\ &\leqslant |x_n||y_n - y| + |y||x_n - x|.\end{aligned}$$

(Here $|\cdot|$ denotes the Euclidean norm.) Notice that by the previous lemma $|x_n|$ is bounded by, say, M. Therefore, for sufficiently large n, we can have

$$|y_n - y| < \frac{\epsilon}{2M} \qquad \text{and} \qquad |x_n - x| < \frac{\epsilon}{2|y|}$$

satisfied simultaneously. Then the entire term is bounded by ϵ, completing our proof that $x_n \cdot y_n \to x \cdot y$.

(3) Similar to (2) — *add and subtract* a mixed term and then use Cauchy-Schwarz.

(4) We prove by induction. Case $m = 1$ is true by assumption. Now suppose $\beta_n^m \to \beta^m$. Notice that

$$|\beta_n^{m+1} - \beta^{m+1}| = |\beta_n^m \beta_n - \beta^m \beta|,$$

so the claim follows from the induction hypothesis and (2). □

Example 4.14: Example 1.7 Revisited. *Recall in Example 1.7 we said that there is no largest $q \in \mathbb{Q}$ such that $q^3 < 2$. Back then the only possible approach was to derive algebraic answer. Now, with the help of limits, we have a much cleaner approach, thanks to the power of convergence.*
Claim: if $p \in \mathbb{Q}$ and $p^3 < 2$ then there exists $q \in \mathbb{Q}$ with $q > p$ and $q^3 < 2$.

Proof. Since the sequence $1/n \to 0$, the sequence $(p + 1/n) \to p$. By part (4) of the previous example, taking cube preserves the convergence, so $(p + 1/n)^3 \to p^3 < 2$. This proves the claim, as the convergence implies that $(p + 1/n)^3 > p^3$ can be made arbitrarily close to p^3 for sufficiently large n. Of course, $p + 1/n$ is still a rational, and by "arbitrarily close" we mean that it can be smaller than $p^3 + (2 - p^3) = 2$. □

3.2 Cauchy Sequences

Definition 4.15: Cauchy Sequences

A sequence $(x_n) \subset X$ is **Cauchy** if the following holds:

For all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n, x_m) < \epsilon$ for all $n \geqslant N$. *Note that the criterion makes no mention of limits; "Cauchy-ness" is a property of the sequence itself.*

Lemma 4.16: Cauchy ⇒ Bounded

Any Cauchy sequence (x_n) is bounded.

Proof. This is very similar to the proof showing convergent sequences are bounded. The idea is still that we can bound the tail, and the remaining finite early terms are also bounded.
Let $k \in \mathbb{N}$ be such that $d(x_n, d_m) \leqslant 1$ for all $n, m \geqslant k$. Then, since $1 < 2$, all x_n for $n \geqslant k$ are contained in $N_2(x_k)$. Then if we simply define

$$r := \max\{2, d(x_1, x_k), \ldots, d(x_{k-1}, x_k)\} + 1$$

we see that all points are contained in $N_r(x_k)$, proving our claim. □

Lemma 4.17: Convergent ⇒ Cauchy

If (x_n) converges, it is Cauchy.

Proof. Suppose $(x_n) \to x$. Using convergence, we pick N large enough such that $d(x_n, x) < \epsilon/2$ if $n \geqslant N$. Then, for $m, n \geqslant N$, we have $d(x_n, d_m) \leqslant d(x_n, x) + d(x, x_m) < \epsilon/2 + \epsilon/2 = \epsilon$, completing the proof. □

While convergent sequences are Cauchy, the converse is not necessarily true. We need to introduce the notion of completeness to guarantee the converse, and we will present counterexamples showing that Cauchy sequences may not be convergent.

Definition 4.18: Complete Metric Spaces

Let X be a metric space. If all Cauchy sequence $(x_n) \subset X$ converge (with limits in X), we say X is **complete**. *Intuitively, there are no "holes" in this space. A complete normed vector space is called a* ***Banach space****.*

Example 4.19. **$((0,1), \|\cdot\|)$ is not Complete.** In particular, the Cauchy sequence $(1/n)$ does not converge in this space. (*It is Cauchy because* $(1/n)$ *converges in the ambient space* $(\mathbb{R}, \|\cdot\|)$, *whereas the limit* 0 *in the ambient space does not lie in* $((0,1), \|\cdot\|)$.)

Lemma 4.20: $(\mathbb{R}^k, \|\cdot\|)$ is Complete

$(\mathbb{R}^k, \|\cdot\|)$ is complete.

Proof. Let $(x_n) \subset \mathbb{R}^k$ be a Cauchy sequence. It follows that (x_n) is bounded , and thus there exists $R > 0$ such that all x_n are contained in $[-R, R]^k := [-R, R] \times \ldots \times [-R, R]$. By the Bolzano-Weierstraß Theorem there exists a subsequence (x_{n_k}) that converges to some $x \in [-R, R]^k$. *Note that we are still not done! We have found a convergent subsequence, but we need to show the convergence of the entire sequence.*

The key is to notice that:

(1) *The late terms in a Cauchy sequence are very close to each other, and*

(2) *Among these terms, some are contained in the subsequence (x_{n_k}), so these terms are also very close to the limit.*

Therefore, all late terms in the Cauchy sequence must be close to the limit of (x_{n_k}), i.e., (x_n) shares the same limit.

Now we prove the italicized portion rigorously. Fix $\epsilon > 0$. Using convergence, let $N_1 \in \mathbb{N}$ be such that $d(x_{n_k}, x) \leqslant \epsilon/2$ for all terms in the subsequence whose indices $\geqslant N_1$ (*for shorthand notation, we write "for $n_k \geqslant N_1$"*). Also, using Cauchy-ness, we can find $N_2 \in \mathbb{N}$ such that $d(x_n, x_m) < \epsilon/2$ for all $n, m \geqslant N_2$.

Now we define $N := \max(N_1, N_2)$ and pick any $n_k \geqslant N_1$ (from the subsequence). It follows that, for any $n \geqslant N$,

$$d(x_n, x) \leqslant d(x_n, x_{n_k}) + d(x_{n_k}, x) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Therefore the entire sequence converges to x, completing the proof. □

3.3 Limits & the Euler Number e

We will be temporarily done with abstract metric spaces, and we will shift our focus to $(X, d) := (\mathbb{R}, \|\cdot\|)$.

Definition 4.21: Infinite Limit of a Sequence

Let $(x_n) \subset \mathbb{R}$. If for all $M \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that $x_n > M$ for all $n \geqslant N$, we say that x_n *goes to infinity* and write $x_n \to \infty$. That $x_n \to -\infty$ is defined analogously.
Note that $(x_n \to \infty)$ *is not equivalent to* $\sup\{x_n\} = \infty$. *The* $\implies$ *is true, whereas the converse is not, unless the sequence is monotonic (see below) — for example, consider the sequence* $(2, 1, 3, 1, 4, 1, \ldots)$.

Beginning of Feb. 17, 2021

Theorem 5.1: Convergence of Monotone Sequences

Suppose $(x_n) \subset \mathbb{R}$ is **nondecreasing** (resp.[3] **nonincreasing**), i.e., $x_{n+1} \geqslant x_n$ for all n (resp. $\leqslant$), then either x_n converges to some $x \in \mathbb{R}$ or $x_n \to \infty$ (resp. $x_n \to -\infty$). In the latter case, $\sup\{x_n\} = \infty$ (resp. $-\infty$).
Future reference: Euler e, Lemma 6.2

Proof. We will only show the nondecreasing case. If $\sup\{x_n\} = \infty$, using the definition of supremum, for all $M \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that $x_N > M$. Then, since the sequence is nondecreasing, $x_n \geqslant x_N > M$ for all $n \geqslant N$, and this is precisely what $(x_n \to \infty)$ means, by definition.
Conversely, if $\sup\{x_n\} < \infty$, then (also by definition of supremum), give $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $x - \epsilon < x_N \leqslant x$. Again, since (x_n) is nondecreasing, $x - \epsilon < x_N \leqslant x_n \leqslant x < x + \epsilon$ for all $n \geqslant N$. Hence $x_n \to x$. □

Example 5.2. (A harder example) Suppose $(x_n) \subset (0, 1)$ is *any* sequence satisfying $x_n(1 - x_{n+1}) > 1/4$ for all $n \geqslant 1$. Find $\lim_{n\to\infty} x_n$ (that is, first show $\lim_{n\to\infty} x_n$ exists and then calculate it).

Solution. Note that the AM-GM inequality gives

$$\frac{x_n + (1 - x_{n+1})}{2} \geqslant \sqrt{x_n(1 - x_{n+1})} > \frac{1}{2} \implies x_n > x_{n+1}.$$

Since the sequence is bounded below by 0, the limit exists and $\geqslant 0$. Suppose $x_n \to x$. Then

$$\lim_{n\to\infty} x_n(1 - x_{n+1}) = x(1 - x) \leqslant \frac{1}{4},$$

so the only possibility is if $x(1-x) = 1/4$, i.e., $x = 1/2$. *We will show later how attaining* $1/4$ *does not violate the strictly inequality* $(x_n)(1 - x_{n+1}) > 1/4$ *later.* □

Example 5.3: the Euler Number.. *This provides one approach to define* e, *the Euler number.* Define

$$x_n := \left(1 + \frac{1}{n}\right)^n \quad \text{and} \quad y_n := \left(1 + \frac{1}{n}\right)^{n+1}.$$

[3] An abbreviation for *respectively*.

Since $1+1/n>1$, it follows that $y_n > x_n$ for all n. Using the AM-GM inequality again,

$$\left(1\cdot\left(1+\frac{1}{n}\right)^n\right)^{1/(n+1)} \leqslant \frac{\overbrace{(1+1/n)+\ldots+(1+1/n)}^{n \text{ times}}+1}{n+1} = 1+\frac{1}{n+1}.$$

where = only takes place if $1+1/n=1$, which is impossible for $n\in\mathbb{N}$. Therefore $x_n < x_{n+1}$ for all n and we say (x_n) is **strictly increasing**.

We can also show that (y_n) is **strictly decreasing**, i.e., $y_{n+1} < y_n$, using the HM-GM inequality:

$$\left(1\cdot\left(1+\frac{1}{n}\right)^{n+1}\right)^{1/(n+2)} > \frac{n+2}{1+\sum_{i=1}^{n+1} n/(n+1)} = 1+\frac{1}{n+1}.$$

Therefore, for $n>1$, we have $x_1 < x_n < y_n < y_1$. In particular (x_n) is bounded! Therefore, by Theorem 5.1 (x_n) converges, and we define the **Euler number** e to be

$$e := \lim_{n\to\infty} x_n \in \mathbb{R}. \tag{Eq.5.1}$$

Also notice that

$$\lim_{n\to\infty} y_n = \lim_{n\to\infty} x_n\left(1+\frac{1}{n}\right) = e\cdot 1 = e.$$

Future reference: Theorem 5.8, the Euler number as sum of factorials

3.4 Limit Superior & Limit Inferior

However, given an arbitrary sequence (x_n), we may or may not know its behavior, and it may or may not converge. How do we analyze such sequences as $n \to \infty$? The answer is to introduce two more definitions:

Definition 5.4: Limsup & Liminf[4]

Given $(x_n) \subset \mathbb{R}$, we set

$$E := \{y \in \mathbb{R} \cup \{\pm\infty\} : \text{ there exists a subsequence } (x_{n_k}) \text{ such that } x_{n_k} \to y\}. \quad \text{(Eq.5.2)}$$

We define

$$\limsup_{n\to\infty} x_n := \sup E \qquad \text{and} \qquad \liminf_{n\to\infty} x_n := \inf E.$$

We call these the **upper limit** (or **limit superior**) and **lower limit** (or **limit inferior**), respectively.

$\limsup$ *and* $\liminf$ *are generalizations of the notion of a limit. The best thing is that they always exist! Also notice that if* $E \subset \mathbb{R}$ *then* $\liminf$ *and* $\liminf$ *agree with* $\sup$ *and* $\inf$.

Future reference: the Squeeze Theorem

Theorem 5.5: Characterization of Limsup

$x = \limsup_{n\to\infty} x_n$ if and only if:

(1) $x \in E$ (*in other words,* $\sup E$ *is attained, and the* $\limsup$ *itself is a limit of some subsequence*), and

(2) if $(y > x) \in \mathbb{R}$, then there exists $N \in \mathbb{N}$ such that $x_n \leqslant y$ for all $n \geqslant N$ (*i.e., the tail is bounded by* y).

The characterization of $\liminf$ is characterized analogously, replacing $>$ and $\geqslant$ in (2) by $<$ and $\leqslant$.

Future reference: Corollary 5.6, Theorem 5.8, the Root Test

Proof. (Completed on 2/19 and 22.) We first show $\Rightarrow$(1). Assume $x = \limsup_{n\to\infty} x_n$. By definition $x = \sup E$. It follows that there exists $(y_m) \subset E$ converging to x. (*This could be the boring constant sequence and our proof would remain unaffected.*)

(Case 1) Suppose $x \neq \pm\infty$. Fix $\epsilon > 0$. We want to show that $x \in E$, i.e., there exists some n such that $|x_n - x| < \epsilon$. (*Then we can take* ϵ *to be smaller and smaller and obtain a subsequence* (x_{n_k}) *that becomes closer and closer to* x*, i.e., converges to* x.)

Let $y \in (y_m) \subset E$ be such that $|x - y| < \epsilon/2$. (*This is possible as we assumed* $(y_m) \to x$.) Also, pick $n \in \mathbb{N}$ such that $|x_n - y| < \epsilon/2$. (*This is also possible because* $y \in E$ *and so* y *is a limit point for some subsequence, and we simply need to pick* x_n *from this subsequence.*) Then, by triangle inequality $|x_n - x| \leqslant |x_n - y| + |y - x| < \epsilon$. This proves $x \in E$.

(Case 2) If $x = -\infty$ the claim is trivial: if so $E = \{-\infty\}$, for if there were anything else then $\sup E > -\infty$.

[4] Here is another way to think of $\limsup$ for a bounded sequence: given (x_n) and $k \in \mathbb{N}$, the supremum of $(x_k, x_{k+1}, \dots)$ exists. If we increase k, the supremum is taken over a smaller set and thus $(s_k)_{k\geqslant 1}$ defined by $s_k := \sup_{n\geqslant k}\{x_n\}$ forms a nonincreasing sequence, and by Bolzano-Weierstraß there exists a limit. The $\limsup$ is precisely *limit of supremum*, $\lim_{k\to\infty} \sup_{n\geqslant k}\{x_n\}$. The definition also holds for unbounded sequences (in which cases the values are $\pm\infty$).

(Case 3) If $x = \infty$, we want to show that there are arbitrarily large terms in (x_n), i.e., for all $M \in \mathbb{R}$, there exists $n \in \mathbb{N}$ such that $x_n > M$. Now we fix $M \in \mathbb{R}$. Let $y \in E$ be such that $y > M + 1$. (*This is possible because* $\sup E = \infty$, *so* E *contains arbitrarily large elements (or even* ∞ *itself).*) Now that $y > M + 1$ is a limit point, there exists x_n such that

$$\begin{cases} x_n > M & \text{if } y = \infty \\ |x_n - y| < 1 & \text{if } y \in \mathbb{R}. \end{cases}$$

In either case, $x_n > M$, completing our sub-proof.

Now we need to show $\Rightarrow$ (2). The case $x = \infty$ is trivial because there is no such y with $y > x$. Now we suppose x is finite and that (2) is false. Then there exists $(y > x) \in \mathbb{R}$ such that (x_n) does not have a tail that is bounded above by y. This means that, for all $N \in \mathbb{N}$, there exists $n \geqslant N$ with $x_n > y$. *In particular there are infinitely many* x_n*'s greater than* y. That is, we can extract an (infinite) subsequence (x_{n_k}) with each term $> y$.
We now show that (x_{n_k}) causes a contradiction:

(1) If (x_{n_k}) is unbounded above, then there exists a subsequence $(x_{n_{k_m}}) \to \infty$ so $\infty \in E$, and this contradicts our assumption that $x := \sup E$ is finite.

(2) If (x_{n_k}) is bounded above, (*since it is also bounded below by* y) Bolzano-Weierstraß implies it has a convergent sub-subsequence $(x_{n_{k_m}})$. This sub-subsequence converges to some $z \geqslant y > x$ so $z \in E$, again contradicting our assumption that $x = \sup E$. This concludes the proof of $\Rightarrow$(2) and thus $\implies$ entirely.

Now we show $\impliedby$. We assume that $x \in E$ and that if $y > x$ then the tail of (x_n) is bounded above by y; we want to show that $x = \sup E$. Suppose for contradiction that $x \neq \sup E$. *Notice that since* $x \in E$*, anything smaller than* x *cannot be an upper bound for* E*, so our assumption implies that it is Definition 1.6.1 — rather than 1.6.2 — that is being violated, i.e.,* x *is not an upper bound of* E. Therefore there exists $(z > x) \in E$.
Let $y \in \mathbb{R}$ be such that $y \in (x, z)$. By the theorem's second condition, there exists $N \in \mathbb{N}$ such that $x_n \leqslant y$ for all $n \geqslant N$. Therefore the tail of (x_n) is bounded above by y, and it is impossible to find any subsequence converging to z. Therefore $z \notin E$, contradiction. This concludes the proof. ☐

Corollary 5.6

(1) $\limsup_{n\to\infty} = -\infty$ if and only if $x_n \to -\infty$. *Heuristically, if all possible subsequences tend to* $-\infty$*, the main sequence must also act like this.* Similarly, $\liminf_{n\to\infty} x_n = \infty$ if and only if $x_n \to \infty$.

(2) x_n converges to some $x \in \mathbb{R}$ if and only if $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = x \in \mathbb{R}$.

Future reference: Squeeze Theorem, Example 5.12, Example 6.5, the Euler number as sum of factorials

Proof. (1) directly follows from condition (2) from Theorem 5.5. If $\limsup_{n\to\infty} x_n = -\infty$ and we take any $(x \in \mathbb{R}) > -\infty$, the Theorem implies that the tail of x_n is bounded above by x. Since x is arbitrary, $x_n \to -\infty$. Conversely, if $x_n \to -\infty$ then any subsequence tends to $-\infty$, so clearly $\limsup_{n\to\infty} -\infty = \sup\{-\infty\} = -\infty$.

(2) For $\implies$, suppose $x_n \to x$ for some x. Then any subsequence must converge to x, i.e., $E = \{x\}$. Thus

$$\liminf_{n\to\infty} x_n = \inf E = x = \sup E = \limsup_{n\to\infty} x_n.$$

For $\Longleftarrow$, fix $\epsilon > 0$. We use (2) in Theorem 5.5 again on $x + \epsilon > x$. It follows that there exists $N_1 \in \mathbb{N}$ such that $x_n < x + \epsilon$ for all $n \geqslant N_1$[5]. Similarly, we use the liminf version on $x - \epsilon < x$, and there exists $N_2 \in \mathbb{N}$ such that $x_n > x + \epsilon$ for all $n \geqslant N_2$. Thus, setting $N := \max(N_1, N_2)$, we see that

$$|x_n - x| < \epsilon \text{ for all } n \geqslant N.$$

This concludes the proof. □

Example 5.7: Examples on Liminf & Limsup.

(1) Let $x_n := (-1)^n$. It follows that the only convergent subsequence consist of -1 and 1, so $E = \{\pm 1\}$ and thus $\limsup_{n\to\infty} x_n = 1$, $\liminf_{n\to\infty} x_n = -1$.

(2) Let $(q_n)_{n\geqslant 1}$ be an enumeration of rational numbers. Then $E = \mathbb{R} \cup \{\pm\infty\}$ and

$$\limsup_{n\to\infty} q_n = +\infty \qquad \liminf_{n\to\infty} q_n = -\infty.$$

To see this, we can, for example, define a subsequence (q_{n_k}) by setting $q_{n_1} := q_1$ and $q_{n_{k+1}}$ be the first number in (q_n) that is at least $q_{n_k} + 1$ and appears after q_{n_k}. This is possible because there are infinitely many rationals $\geqslant q_{n_k} + 1$. Then this sequence has limit $+\infty$. Likewise for the opposite case.

(3) Find all limit points as well as liminf and limsup of $x_n := (2\cos(2\pi n/3))^n$.

Beginning of Feb. 19, 2021

Solution of 5.7(3). The main idea is to divide the sequence into subsequences for which we can calculate limits. For example, notice that

$$x_{3k} = 2^{3k},$$

$$x_{3k+1} = (2\cos(2\pi/3))^{3k+1} = (-1)^{3k+1} = (-1)^{k+1},$$

and

$$x_{3k+2} = (2\cos(4\pi/3))^{3k+2} = (-1)^{3k+2} = (-1)^k,$$

and these three cases account for all terms in (x_n).

The first case clearly $\to \infty$, whereas the second and third both oscillates between ± 1. It follows that the E for this sequence is $\{\pm 1, \infty\}$. It follows that

$$\limsup_{n\to\infty} x_n = \infty \qquad \text{and} \qquad \liminf_{n\to\infty} x_n = -1.$$

Theorem 5.8

If $x_n \leqslant y_n$ for all n then the same relations for lim sup and lim inf:

$$\limsup_{n\to\infty} x_n \leqslant \limsup_{n\to\infty} y_n \qquad \text{and} \qquad \liminf_{n\to\infty} x_n \leqslant \liminf_{n\to\infty} y_n.$$

In particular, if $x_n \to x$ and $y_n \to y$ and $x_n < y_n$ for all n, we have $x \leqslant y$. *The $\leqslant$ confirms Example 5.2 and the Euler number (Example 5.3), in which we indeed obtained $\leqslant$ from $<$.*

[5] Theorem 5.5 only gives $\leqslant$, but of course we make it $<$ by choosing a stronger ϵ. In ϵ-proofs, $\leqslant$ and $<$ are both fine.

Future reference: the Euler number as sum of factorials, L'Hôpital's Rule, uniformly convergent $\Leftrightarrow$ Cauchy in $\|\cdot\|_{\sup}$, Moore-Smith Theorem

Proof. We will only prove the $\limsup$ case. Let $c := \limsup_{n\to\infty} x_n$. WLOG assume $x > -\infty$ (otherwise, the claim is trivial). By condition (1) of Theorem 5.5, x is the limit of some subsequence, say (x_{n_k}). In other words, for all $\epsilon > 0$, there exists $K \in \mathbb{N}$ such that $c - \epsilon < x_{n_k} < c + \epsilon$ for all $k \geqslant K$. Notice that (y_{n_k}) is bounded below by (x_{n_k}).
We claim that (y_{n_k}) admits a subsequence $(y_{n_{k_m}})$ (*yes... triple subscript*) that converges to some $y \in \mathbb{R} \cup \{\infty\}$. (*If* (y_{n_k}) *is unbounded, set it to* ∞*; otherwise use Bolzano-Weierstraß. It cannot be* $-\infty$ *as* (x_{n_k}) *is a lower bound.*) Note that $y \geqslant c - \epsilon$ because $(y_{n_k}) > c - \epsilon$. Since ϵ is arbitrary, we get $y \geqslant c$. In particular we have found *some* limit point of (y_n) that is $\geqslant c$, so by definition $\limsup_{n\to\infty} y_n \geqslant y \geqslant \limsup_{n\to\infty} x_n$. ☐

Corollary 5.9: the "Squeeze Theorem" / "Sandwich Theorem"

If $a_n \leqslant b_n \leqslant c_n$ and $a_n \to x$, $c_n \to x$ for some $x \in \mathbb{R}$, then $b_n \to x$.
In the infinite case, if $a_n \leqslant b_n$ *and* $a_n \to \infty$*, then* $b_n \to \infty$.
Future reference: Lemma 6.4.3, Example 6.12

Proof. We can completely overkill this theorem using the previous one:

$$x = \liminf_{n\to\infty} a_n \leqslant \liminf_{n\to\infty} b_n \leqslant \limsup_{n\to\infty} b_n \leqslant \limsup_{n\to\infty} c_n = x.$$

The $=$'s are by Corollary 5.6.2, the middle $\leqslant$ by Definition 5.4 ($\inf \leqslant \sup$), and the other two $\leqslant$'s by the previous theorem. Hence $\liminf b_n = \limsup b_n$ and the other direction of Corollary 5.6.2 gives $b_n \to x$. ☐

Example 5.10. $p^{1/n} \to 1$ for all $p \to 0$.

Proof. If $p \geqslant 1$, we can define $x_n := p^{1/n} - 1 \geqslant 0$. Using binomial expansion,

$$p = (1 + x_n)^n = 1 + nx_n + \underbrace{\text{remaining terms}}_{\geqslant 0} \geqslant 1 + nx_n.$$

Therefore $0 \leqslant x_n \leqslant (p-1)/n$, and using the Squeeze Theorem we conclude $x_n \to 0$, i.e., $p^{1/n} \to 1$.
If $p < 1$ then we write $p^{1/n}$ as $1/(1/p)^{1/n}$. The denominator $\to 1$ and so $p^{1/n} \to 1/1 = 1$. (*This is a nontrivial claim and it needs some justification. Cf. PS5.8.*) ☐

Example 5.11. Compute the limit for $x_n := (4^{(-1)^n} + 2)^{1/n}$.

Solution. Observe that $(-1)^n$ can either be -1 or 1, so $4^{(-1)^n} + 2$ is either $9/4$ or 6. Therefore,

$$(9/4)^{1/n} \leqslant x_n \leqslant 6^{1/n}$$

and the Squeeze Theorem implies $x_n \to 1$. ☐

Example 5.12. Let $(a_n) \subset (0, \infty)$ be such that

$$a_{m+n} \leqslant a_n + a_m + C \qquad (\Delta)$$

for some $C \geqslant 0$ and all $m, n \in \mathbb{N}$. Show that $\lim\limits_{n\to\infty} \dfrac{a_n}{n}$ exists.

Proof. Heuristically, (Δ) *tells us* (a_n) *behaves well. In particular, if* (a_n) *is decreasing then the limit of course exists and equals* 0*. Even in the worst case where* (a_n) *is increasing, it increases at most linearly, and the constant term* C *gets killed by the ever increasing* n *in the denominator.*

We prove this by using the division algorithm. Fix $k \in \mathbb{N}$. It follows that any n can be written as $\ell k + r$ for some $\ell \in \mathbb{N}$ and $0 \leqslant r \leqslant k - 1$. Therefore,

$$a_n = a_{\ell k + r} \leqslant a_{\ell k} + a_r + C \leqslant \ldots \leqslant \ell(a_k + C) + a_r$$

(where the ... denotes iterating the inequality $a_{\ell k} \leqslant a_k + a_{(\ell-1)k} + C$). Then,

$$\frac{a_n}{n} \leqslant \frac{\ell}{n}(a_k + C) + \frac{a_r}{n} \leqslant \frac{a_k + C}{k} + \frac{a_r}{n}$$

where the second $\leqslant$ is because $n = \ell k + r \geqslant \ell k$ and then the ℓ's cancel. Since k is fixed, a_r is bounded (by $\max(a_0, \ldots, a_{k-1})$), and so $a_r/n \to 0$ as $n \to \infty$. Taking $\limsup$, we obtain

$$\limsup_{n\to\infty} \frac{a_n}{n} \leqslant \limsup_{n\to\infty} \frac{a_k + C}{k} + \frac{a_r}{n} = \frac{a_k + C}{k} \qquad \text{for all } k \in \mathbb{N}.$$

Note that here $(a_k + C)/k$ *is a constant once we prescribe a* k*.* On the other hand, if we take $\liminf$ *over* k *(not* n*)*,

$$\limsup_{n\to\infty} \frac{a_n}{n} \leqslant \liminf_{k\to\infty} \frac{a_k + C}{k} \leqslant \liminf_{k\to\infty} \frac{a_k}{k} + \liminf_{k\to\infty} \frac{C}{k} = \liminf_{k\to\infty} \frac{a_k}{k}.$$

(*It is taken for granted that* $\liminf$ *is subadditive; the proof of this requires some justification though.*) Since n, k at this point are just dummy variables, we see that the $\limsup$ and $\liminf$ of (a_n/n) must equal, and by Corollary 5.6.2 (a_n/n) converges, i.e., the limit exists.

This is an elegant proof. We first used the notion of $\limsup$ *to get an inequality involving our target of interest,* $\limsup\limits_{n\to\infty} a_n/n$*, with dependence on one variable* (k) *and then take* $\liminf$ *to get rid of* k *as well.* □

3.5 Series

Beginning of Feb. 22, 2021

Definition 6.1: Series

A **series** is a sequence of a particular form: it is a sequence consisting of *partial sums*

$$S_n := \sum_{k=1}^{n} a_k \qquad \text{for some} \qquad (a_n) \subset \mathbb{R}$$

(*or summation over a complex sequence, but we only focus on the real case*). We common use the shorthand notation $\sum a_n$. We say that $\sum a_n$ **converges** if S_n converges (i.e., if $\lim\limits_{n\to\infty} S_n = s$ for some $s \in \mathbb{R}$) and **diverges otherwise** (i.e., if $\lim\limits_{n\to\infty} S_n$ does not exist or $S_n \to \pm\infty$).

The theory of series is concerned with determining whether a given series converges, rather than finding the limit explicitly, as the latter is often too hard.

Lemma 6.2

(1) *Since $\mathbb{R}$ is complete, a series converges if and only if it is Cauchy. This gives the* ***Cauchy Convergence Criterion (CCC)***: $\sum a_n$ converges if and only if

$$\boxed{\text{For all } \epsilon > 0\text{, there exists } N \in \mathbb{N} \text{ such that } \left|\sum_{k=m}^{n} a_k\right| < \epsilon \text{ for all } m \geqslant n \geqslant N.}$$

Notice that this is exactly saying that the sequence (S_n) is Cauchy. Future reference: Theorem 6.18, Weierstraß M-Test

(2) If $\sum a_n$ converges then $a_n \to 0$. *Use (1) and take $m := n$.* Future reference: Lemma 6.4.2, the Root Test, the Ratio Test

(3) If (a_n) is nonnegative (*so S_n is nondecreasing*), then $\sum a_n$ (namely S_n) converges if and only if it is bounded (this is given directly by Theorem 5.1).

Future reference: Cauchy Condensation Test

Definition 6.3: Absolute Convergence

We say $\sum a_n$ **converges absolutely** if $\sum |a_n|$ converges, hence the word *absolutely*.

Lemma 6.4

(1) Absolute convergence implies convergence: if $\sum a_n$ converges absolutely then $\sum a_n$ converges.

One-liner proof: $\sum_{k=m}^{n} a_k \leqslant \sum_{k=m}^{n} |a_k| < \epsilon$ *by Δ-inequality and assumption on $\sum |a_n|$.*

(2) (**Comparison test**): If $|a_n| \leqslant c_n$ and $\sum c_n$ converges, then $\sum a_n$ converges absolutely (and thus converges too).

Also one-liner: $S_n = \sum_{k=1}^{n} |a_k| \leqslant \sum_{k=1}^{n} c_k < \sum_{k=1}^{\infty} c_k < \infty$ *for all n. Then use emma 6.2.3.*

Future reference: the Root Test

(3) If $a_n \geqslant b_n \geqslant 0$ and $\sum b_n$ diverges, then $\sum a_n$ diverges.

Also one-liner: $S_n = \sum_{k=1}^{n} a_k \geqslant \sum_{k=1}^{n} b_k \to \infty$ *by the Squeeze Theorem.*

Example 6.5. Consider the sequence $a_n = (-1)^n$. We have

$$S_n = \sum_{k=1}^{n} a_k = \begin{cases} 0 & n \text{ even} \\ -1 & n \text{ odd.} \end{cases}$$

This sequence diverges. (It's always oscillating between 0 and -1. Alternatively, we could use Corollary 5.6.2 and $\limsup_{n\to\infty} S_n = 0 \neq -1 = \liminf_{n\to\infty} S_n$ to say that S_n diverges.)

Example 6.6: Geometric Series. Let $q \in \mathbb{R}$. A series of form $\sum_{n=0}^{\infty} q^n$ (or starting with $n = 1$ or any n...) is called a **geometric series**. Such series converges if and only if $|q| < 1$.

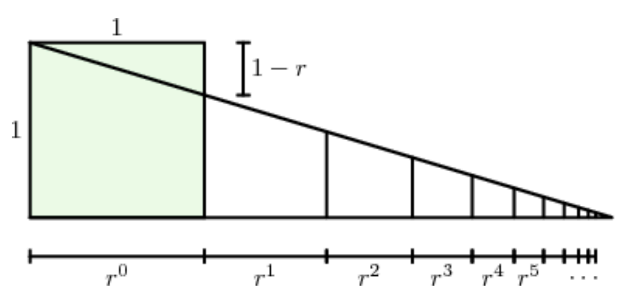
Future reference: p-series

Proof. We all know (hopefully) that

$$\sum_{k=0}^{n} q^k = \frac{1 - q^{n+1}}{1 - q}.$$

For a fixed q, $q^{n+1} \to 0$ if $|q| < 1$ so the entire series converges to $1/(1-q)$, whereas q^{n+1} is unbounded if $|q| > 1$. If $q = \pm 1$ it is clear that the series diverges as well.

Alternatively, below is a *geometric* interpretation of the *geometric* series for $|q| < 1$:



□

Beginning of Feb. 24, 2021

Theorem 6.7: "Dyadic Thinning" / Cauchy Condensation Test

Suppose (a_n) satisfies $0 \leqslant a_{n+1} \leqslant a_n$ for all n. Then

$$\sum a_n \text{ converges} \iff \sum 2^k a_{2^k} \text{ converges.}$$

We see immediately that this theorem fails without the assumption that (a_n) is nonincreasing. However, with a nonincreasing sequence, the theorem holds nicely. When I was first learning Calculus II, the proof showing that the harmonic *series* $1 + 1/2 + 1/3 + \dots$ *diverges uses precisely this trick*:

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \left(\frac{1}{2} + \frac{1}{3}\right) + \left(\frac{1}{4} + \dots + \frac{1}{8}\right) + \dots = \sum_{n=0}^{\infty} \left(\frac{1}{2^n} + \dots + \frac{1}{2^{n+1}-1}\right) > \sum_{n=0}^{\infty} 1 = \infty.$$

Proof. We define $S_n := \sum_{n=1}^{k} a_k$ and $T_k := \sum_{m=0}^{k} 2^m a_{2^m}$. We will show that

$$\begin{cases} S_n \leqslant T_k & \text{for } n \leqslant 2^k \quad (1) \\ T_k \leqslant 2S_n & \text{for } n > 2^k. \quad (2) \end{cases}$$

If so, S_n is bounded $\Leftrightarrow$ T_k is. (S_n bounded $\Rightarrow$ T_k bounded by (2) and the converse by (1).)
Since a_k and $a^m a_{2^m}$ are nonnegative, Lemma 6.2.3 (first and third $\Leftrightarrow$) would then imply

$$\sum a_n = S_n \text{ converges } \Leftrightarrow S_n \text{ is bounded } \Leftrightarrow T_k \text{ is bounded } \Leftrightarrow T_n = \sum 2^k 2_{2^k} \text{ converges}$$

which would complete the proof. Below we show (1) and (2).
For (1), if $n \leqslant 2^k$, since a_k is nonnegative, we have $S_n \leqslant S_{2^k}$. Thus,

$$\begin{aligned} S_n &\leqslant S_{2^k} \\ &= a_1 + (a_2 + a_3) + (a_4 + \dots + a_7) + \dots + (a_{2^{k-1}} + \dots + a_{2^k-1}) + a_{2^k} \\ &\leqslant 2^0 \cdot a_{2^0} + 2 \cdot a_{2^1} + 2^2 \cdot a_{2^2} + \dots + 2^{k-1} a_{2^{k-1}} + a_{2^k} \\ &\leqslant \dots + 2^k a_{2^k} = T_k. \end{aligned}$$

Conversely, for (2), if $n > 2^k$,

$$\begin{aligned} S_n &= a_1 + a_2 + (a_3 + a_4) + (a_5 + \dots + a_8) + \dots + (a_{2^{k-1}+1} + \dots + a_{2^k}) + (a_{2^k+1} + \dots + a_n) \\ &\geqslant 2^{-1} \cdot a_1 + 2^0 \cdot a_2 + 2^1 a_{2^2} + \dots + 2^{2k-1} a_{2^k} + 0 \\ &\geqslant \sum_{m=0}^{k} 2^{m-1} a_{2^m} = \frac{T_k}{2}. \end{aligned}$$

□

Example 6.8: Convergence of *p*-Series. A series of form $\sum 1/n^p$ (indexed over n) is called a ***p*-series**. A p-series converge if and only if $p > 1$.[6]
Future reference: the Root Test, Example 6.14, Example 13.10

Proof. The case where $p \leqslant 0$ is trivial as $1/n^p \geqslant 1$ and the series clearly diverges.
If $p > 0$, since the terms are all nonnegative, the previous theorem states that

$$\sum_{n=1}^{\infty} \frac{1}{n^p} \text{ converges } \iff \sum_{k=0}^{\infty} \frac{2^k}{2^{kp}} \text{ converges.}$$

Notice that the RHS is a geometric series[!] Using Example 6.6, it converges if and only if the ratio

$$\frac{2^{k+1}}{2^{(k+1)p}} \cdot \frac{2^{kp}}{2^k} = 2^{1-p} < 1.$$

This implies that the series converges if and only if $1 - p < 0$, i.e., $p > 1$. □

[6] Rudin took a completely different approach than the one I learned, which used the integral test. Mind-blowing for me!

Example 6.9: Characterization of the Euler Number.

$$\sum_{n=0}^{\infty} \frac{1}{n!} = e,$$

where ! denotes factorial and $0! := 1$.

As previous mentioned, the limit

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

exists, and we defined it to be e, the Euler number. We didn't know precisely what that limit is, but now we do.

Proof. *This proof requires clear manipulations of* $\liminf$ *and* $\limsup$. First we define

$$b_n := \left(1+\frac{1}{n}\right)^n \qquad \text{and} \qquad S_n := \sum_{k=0}^{n} \frac{1}{k!}.$$

Using binomial theorem, for each n, we have

$$\begin{aligned}
b_n &= \frac{1}{n^n}(n+1)^n = \frac{1}{n^n}\sum_{k=0}^{n}\binom{n}{k}n^{n-k}1^k = \sum_{k=0}^{n}\frac{1}{n^k}\frac{n!}{k!(n-k)!} \\
&= 1 + \sum_{k=1}^{n}\frac{1}{n^k}\frac{1}{k!}\underbrace{n(n-1)\ldots(n-k+1)}_{k \text{ terms}} \\
&= 1 + \sum_{k=1}^{n}\frac{1}{k!}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right) \\
&\leqslant 1 + \sum_{k=1}^{n}\frac{1}{k!} = \sum_{k=0}^{n}\frac{1}{k!} = S_n.
\end{aligned}$$

As shown previously, b_n converges and so $\liminf\limits_{n\to\infty} b_n$ equals the limit e by Corollary 5.6.2. On the other hand, since $b_n \leqslant S_n$ for all n, by Theorem 5.8 we have

$$\liminf_{n\to\infty} b_n = e \leqslant \liminf_{n\to\infty} S_n. \tag{1}$$

The other direction $\geqslant$ is slightly harder. If we fix $m \in \mathbb{N}$ and take $n \geqslant m$, we have

$$\begin{aligned}
b_n &= \frac{1}{n^n}(n+1)^n = 1 + \sum_{k=1}^{n}\frac{1}{k!}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right) \\
&\geqslant 1 + \sum_{k=1}^{m}\frac{1}{k!}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)
\end{aligned}$$

Now we take $\limsup$ on both sides (on n). Keeping m fixed and letting $n \to \infty$, each term $(n-j)/n$ converges to 1, and so does their finite product. Then,

$$b_n \geqslant 1 + \sum_{k=1}^{m}\frac{1}{k!} \text{ for all } n \implies \liminf_{n\to\infty} b_n = e \geqslant 1 + \sum_{k=1}^{m}\frac{1}{k!} = \sum_{k=0}^{m}\frac{1}{k!}.$$

Finally, letting $m \to \infty$ and taking $\limsup$ over m we get $e \leqslant \limsup\limits_{m\to\infty} b_m$. This, with (1), completes the proof. □

3.6 The Root Test & the Ratio Test

In this section we present some convenient methods to decide whether a series is convergent.

Theorem 6.10: the Root Test

Let (a_n) be a sequence. We define its **exponential growth rate** to be $\alpha := \limsup_n \sqrt[n]{a_n}$.

(1) If $\alpha < 1$ then $\sum a_n$ converges absolutely (and thus converges).

(2) If $\alpha > 1$ then $\sum a_n$ diverges.

(3) If $\alpha = 1$ the root test is inconclusive: for example $\sum 1/n$ diverges whereas $\sum 1/n^2$ converges (to $\pi^2/6$).

Proof. (1) If $\alpha < 1$, we can take $\beta \in (\alpha, 1)$ and compare (a_n) with the geometric sequence (β_n) with $\beta_n := \beta^n$. By the characterization (2) of lim sup, there exists $N \in \mathbb{N}$ such that $\sqrt[n]{a_n} \leqslant \beta$ for all $n \geqslant N$. In other words, $a_n \leqslant \beta^n$. The comparison test (6.4.2) gives the convergence of (a_n) (we can discard finitely many early terms as they have no effect on convergence — only the tail has).

(2) Using the characterization (1) of lim sup, $\alpha > 1$ is the limit of some $|a_{n_k}|^{1/(n_k)}$. In particular, there exists $N \in \mathbb{N}$ such that $|a_{n_k}|^{1/(n_k)} > 1$ for all $k \geqslant N$. Therefore, for such a_{n_k}'s we have $|a_{n_k}| > 1^{n_k}$, and this violates the CCC and/or Lemma 6.2.2.[7]

□

Theorem 6.11: the Ratio Test

Let (a_n) be a sequence.

(1) If $\limsup_{n\to\infty} |a_{n+1}/a_n| < 1$ then $\sum a_n$ converges absolutely.

(2) If for some $N \in \mathbb{N}$, $|a_{n+1}/a_n| \geqslant 1$ for all $n \geqslant N$, then $\sum a_n$ diverges. [8]

Proof. (1) Let $\alpha := \limsup_{n\to\infty} |a_{n+1}/a_n|$. We can pick $\beta \in (\alpha, 1)$. Similar to the reasoning above, there exists large $N \in \mathbb{N}$ such that $|a_{n+1}| \leqslant \beta |a_n|$ for all $n \geqslant N$. In particular,

$$|\alpha_{n+m}| \leqslant \beta^m |\alpha_n| \implies \sum_{k=N}^{\infty} |a_k| \leqslant |a_n| \sum_{k=0}^{\infty} \beta^k < \infty.$$

(2) For all $n \geqslant N$, we have

$$|a_{n+1}| \geqslant |a_n| \geqslant \ldots \geqslant |a_N|,$$

so $a_n \nrightarrow 0$. By Lemma 6.2.2 $\sum a_n$ diverges. □

Beginning of Feb. 26, 2021

[7] Alternatively, we could again pick $\beta \in (1, \alpha)$. There exists a subsequence (a_{n_k}) with $|a_{n_k}|^{1/(n_k)}$ converging to α. The tail of such sequence $> \beta$ and comparison test once gives divergence.

[8] A weaker but more symmetric version goes as follows: if $\liminf_{n\to\infty} |a_{n+1}/a_n| > 1$ then the series diverges.

Remark. Just a remark on limit points coming out of nowhere: the set $\{-1, 1\}$ has no limit point (no sequence other than $(1, 1, \ldots)$ converges to 1 and likewise for -1) whereas $\{-1, 1\}$ are both limit points of the sequence $a_n := (-1)^n$.

Example 6.12: Root Test vs. Ratio Test. Consider the sequence $(1/2, 1, 1/8, 1/4, 1/32, 1/16, \ldots)$ where we *multiply by* 2 *then divide by* 8 for successive terms. It is obvious that

$$\limsup_{n\to\infty} |a_{n+1}/a_n| = 2 \qquad \text{and} \qquad \liminf_{n\to\infty} |a_{n+1}/a_n| = \frac{1}{8}.$$

Since $\limsup > 1$ and $\liminf < 1$, the ratio test tells us nothing. On the other hand, since

$$\begin{cases} a_{2n+1} = 1/2^{2n+1} \\ a_{2n} = 1/2^{2n-2}, \end{cases} \implies \frac{1}{2^n} \leqslant a_n \leqslant \frac{4}{2^n},$$

and both sides converges to 0, by the Squeeze Theorem we also have

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} = \limsup_{n\to\infty} \sqrt[n]{|a_n|} = \liminf_{n\to\infty} \sqrt[n]{|a_n|} = \frac{1}{2}.$$

This shows that the exponential growth rate of $(a_n) < 1$ and thus the series converges. *It is also very easy for us to verify this:* $(a_1 + a_2, a_3 + a_4, \ldots)$ *forms a geometric sequence with ratio* $1/8$.

Theorem 6.13: Root Test vs. Ratio Test

$$\liminf_{n\to\infty} |a_{n+1}/a_n| \leqslant \liminf_{n\to\infty} \sqrt[n]{|a_n|} \qquad \text{and} \qquad \limsup_{n\to\infty} |a_{n+1}/a_n| \geqslant \limsup_{n\to\infty} \sqrt[n]{|a_n|}.$$

In particular, this shows that (root test works) ⇒ (ratio test works), whereas the converge may not hold, as seen in the above example. Thus the root test is stronger than the ratio test in some cases. However, the root test is harder to apply (ratios are much easier to compute). Of course, both may fail for some series: for example consider the p*-series with* $p \in (0, 1]$.

Example 6.14. Let $a, b > 0$. Then $\sum \frac{n^{2n}}{(n+a)^{n+b}(n+b)^{n+a}}$ converges if and only if $a + b > 1$.

Proof. Notice that both the root test and the ratio test fails in this example. We therefore resort to the comparison test. Notice that

$$\frac{n^{2n}}{(n+a)^{n+b}(n+b)^{n+a}} = \frac{1}{(n+a)^b(1+a/n)^n(n+b)^a(1+b/n)^n}.$$

Since $\lim_{n\to\infty} (1 + a/n)^n = e^{1/a}$ and $\lim_{n\to\infty} (1 + b/n)^n = e^{1/b}$ (*take the* a^{th} *power of both sides and we recoved precisely the limit definition of* e) are both finite, it suffices to bound the other two terms, making sure they stay finite as $n \to \infty$. Since the powers of n in the denominator is $a + b$, the main idea here is to find constants C_1, C_2 such that

$$\frac{C_1}{n^{a+b}} \leqslant \frac{1}{(n+a)^b(n+b)^a} \leqslant \frac{C_2}{n^{a+b}}.$$

If we can do so, since we know exactly when $\sum 1/n^{a+b}$ converges (i.e., if and only if $a + b > 1$), our proof would be finished.

On one hand, by assumption $a, b > 0$, so $(n+a)^b > n^a$ and $(n+b)^a > n^b$. On the other hand, for $n > 1$ we have $a < na$, so

$$(n+a)^b < (n+na)^b = (1+a)^b n^b \qquad \text{and} \qquad (n+b)^a < (1+b)^a n^a.$$

Since a, b are prescribed, $(1+a)^b$ and $(1+b)^a$ can be treated as constants! We have therefore found proper lower and upper bounds for our fraction of interest, completing the proof that $\sum \ldots$ converges if and only if $a+b>1$. ☐

Example 6.15. Suppose (a_n) is a strictly positive sequence. Then

$$\sum a_n \text{ converges} \implies \sum \sqrt{a_n a_{n+1}} \text{ converges.}$$

The proof is a one-liner using AM-GM:

$$\sqrt{a_n a_{n+1}} \leqslant \frac{a_n + a_{n+1}}{2}.$$ ☐

The converse is not true in general. Consider

$$a_n := \begin{cases} 1 & n \text{ odd} \\ 1/n^4 & n \text{ even} \end{cases} \implies \sqrt{a_n a_{n+1}} = \frac{1}{n^2} \text{ or } \frac{1}{(n+1)^2} \leqslant \frac{1}{n^2}.$$

Example 6.16.

(1) $\sum(\sqrt{n+1} - \sqrt{n})$ diverges:

$$\sqrt{n+1} - \sqrt{n} = \frac{1}{\sqrt{n+1} + \sqrt{n}} \geqslant \frac{1}{2\sqrt{n+1}}.$$

When seeing difference of square roots, it is very common to take the "conjugate":

$$\sqrt{a} - \sqrt{b} = \frac{(\sqrt{a} - \sqrt{b})(\sqrt{a} + \sqrt{b})}{\sqrt{a} + \sqrt{b}} = \frac{a-b}{\sqrt{a} + \sqrt{b}.}$$

Same thing for fractions involving complex numbers $a+bi$, in which case we multiply by $a-bi$.

(2) $\sum(\sqrt[n]{n} - 1)^n$ converges: $\sqrt[n]{n}$ cries out for the root test, and we have

$$\limsup_{n\to\infty}((\sqrt[n]{n} - 1)^n)^{1/n} = \limsup_{n\to\infty} \sqrt[n]{n} - 1 = 0.$$

That $\limsup_{n\to\infty} \sqrt[n]{n} = 1$ is usually shown using binomial expansion and squeeze theorem, but here is a quick alternate way using AM-GM:

$$n^{1/n} \leqslant \frac{\overbrace{n + \ldots + n}^{n-2 \text{ times}} + \sqrt{n} + \sqrt{n}}{n} = \frac{n-2}{n} + \frac{2}{\sqrt{n}} \to 1.$$

3.7 Power Series

Example 6.17: Power Series & Radius of Convergence. Given a **power series** $\sum c_n x^n$, we let $\alpha := \limsup_{n\to\infty} \sqrt[n]{|c_n|}$ and define $R := 1/\alpha$ to be its **radius of convergence**. Then, $\sum c_n x^n$ converges absolutely if $|x| < R$ and diverges if $|x| > R$. When $|x| = R$ no assertion can be made. *The proof directly follows from the root test as the power n and the n^{th} power cancel each other and we can take x outside.*
If $\alpha = 0$ we define the radius of convergence to be $+\infty$, and if $\alpha = \infty$ we define the radius of convergence to be 0.
Future reference: Infinite Taylor series

*For example, the series $\sum n^n x^n$ has $\alpha = \infty$ and $R = 0$ so it diverges for all nonzero x. The series $\sum x^n/n!$ has $\alpha = 0$ and $R = \infty$ so it converges (to e^x) for any x [we'll talk about the **Taylor expansion** of e^x later]. Finally, $\sum x^n$ has $\alpha = R = 1$.*

3.8 Absolute vs. Conditional Convergence

Question. Are there convergent series that are not absolutely convergent, i.e., $\sum a_n$ converges but not $\sum |a_n|$?
Answer. Yes!

Theorem 6.18

Suppose $(a_n), (b_n)$ are such that

(1) $S_n := \sum_{k=1}^{n} a_k$ is a bounded sequence,

(2) $b_n \geqslant b_{n+1} \geqslant 0$, i.e., b_n is nonnegative and nonnegative, and

(3) $b_n \to 0$.

Then $\sum a_n b_n$ converges.
Future reference: Alternating Series Test

Proof. (Completed on 3/1.) Let $(a_n), (b_n)$ be sequences that satisfy (1), (2), and (3). Let $\epsilon > 0$ be given. *To show the convergence of $\sum a_n b_n$, one approach is by using the CCC. Therefore, our goal is to show that the partial sums*

$$\Big|\sum_{k=m}^{n} a_k b_k\Big| < \epsilon \text{ for } m, n \text{ large}.$$

By (1), there exists some $M \in \mathbb{R}$ such that $|S_n| \leqslant M$ for all $n \in \mathbb{N}$. By (3), there exists $N \in \mathbb{N}$ such that $b_n < \epsilon/2M$ for $n \geqslant N$. We claim that this N is precisely the one satisfying the CCC as stated above: notice that

$$\begin{aligned}\sum_{k=m}^{n} a_k b_k = \sum_{k=m}^{n} (S_k - S_{k-1}) b_k &= \sum_{k=m}^{n} S_k b_k - \sum_{k=m}^{n} S_{k-1} b_k \\ &= \sum_{k=m}^{n-1} S_k (b_k - b_{k+1}) + S_n b_n - S_{m-1} b_m,\end{aligned}$$

where each term is nonnegative. Also, using the bound M on S_n, we have

$$\begin{aligned}\Big|\sum_{k=m}^{n} a_k b_k\Big| &\leqslant M\Big(b_n + b_m + \sum_{k=m}^{n-1} (b_k - b_{k+1})\Big) \\ &= M(b_n + b_m + (b_m - b_n)) \\ &= 2M b_m \leqslant 2M b_N < \epsilon.\end{aligned}$$

This completes the proof. □

Corollary 6.19: Leibniz Criterion / Alternating Series Test

Suppose $b_n \geqslant b_{n+1} \geqslant 0$, i.e., b_n is nonnegative and nonnegative, and $b_n \to 0$. Then $\sum(-1)^n b_n$ converges. *One-liner proof: take (a_n) with $a_n := (-1)^n$ and then use Theorem 6.18.*

Example 6.20. The series $\sum(-1)^n/n^\alpha$ converges for any $\alpha > 0$. *This generalizes the convergence of p-series which only holds for $p > 1$.*

We conclude the chapter with one last definition, which follows naturally from our discussion of p-series with $p < 1$:

Definition 6.21: Conditional Convergence

We call a sequence that diverges but converges absolutely a **conditionally convergent** sequence.

Chapter 4

Continuity

Beginning of March 1, 2021

In this chapter we will mainly be focus on functions between two metric spaces abstractly, whereas in the next chapter we will focus on integrals of functions on $\mathbb{R}$. We have seen the "ϵ-N" language in the previous chapters; we will start to see "ϵ-δ" now. Unless otherwise indicated, we let $(X, d_X), (Y, d_Y)$ be two metric spaces.

4.1 Limits of Functions

Definition 7.1: Cauchy Definition of Convergent at a Point

Let f be a function that maps $E \subset X$ into Y and let p be a limit point of E. We say $f(x)$ **converges to $q \in Y$ as $x \to p \in X$**, and we write $\lim_{x \to p} f(x) = q$, if the following holds:

Given $\epsilon > 0$, there exists $\delta > 0$ such that $d_Y(f(x), q) < \epsilon$ whenever $0 < d_X(x, p) \leqslant \delta$ and $x \in E$.

In particular, if $(X, d_X) = (\mathbb{R}, \|\cdot\|)$ and $p < \infty$ then the definition reduces to the following:[1]

Given $\epsilon > 0$, there exists $\delta > 0$ such that $d_Y(f(x), q) < \epsilon$ whenever $x \in (p - \delta, p) \cup (p, p + \delta)$.

Future reference: Monotonic functions

Remark. p is a limit point of E but it needs not to be in E. $\lim_{x \to p} f(x) \neq f(p)$ is possible even for $p \in E$. Also, the strict inequality $0 < d_X(x, p)$ implies the definition is meaningless for isolated points.

Definition 7.2: Heine Definition of Convergence at a Point

$\lim_{x \to p} f(x) = q$ if and only if, for every sequence $(p_n) \subset E$ with $p_n \neq p$, we have $\lim_{n \to \infty} f(p_n) \to q$.
Future reference: Lemma 11.19

[1] The definition given in lecture was "$f(x) \to q$ as $x \to p^-$ (i.e., $x \uparrow p$) if ... for all $x \in (p - \delta, p)$", but I believe mine is a more precise analogue. The lecture, however, also gives one condition for $x \to +\infty$: $f(x) \to q$ as $x \to \infty$ if for all $\epsilon > 0$ there exists $M \in \mathbb{R}$ such that $d_Y(f(x), q) < \epsilon$ whenever $x \geq M$ and $x \in E$.

Proof. We first show $\Longrightarrow$. Let $\epsilon > 0$ be given. By assumption, there exists $\delta > 0$ such that $d_Y(f(x), q) < \epsilon$ whenever $d_X(x, p) < \delta$. Now we let (p_n) be any sequence satisfying the *Heine definition*. Since $(p_n) \to p$, there exists $N \in \mathbb{N}$ such that $d_X(p_n, p) < \delta$ for all $n \geqslant N$. Therefore, for these late enough p_n's, $d_Y(f(p_n), q) < \epsilon$. Since ϵ is arbitrary, this is precisely what it mean for $f(p_n)$ to converge to q.

Conversely, suppose for contradiction that $\lim_{x\to p} f(x) \neq q$. Taking the negation of the definition (*which states that whenever x is close enough to p, $f(x)$ is close enough to $f(p)$*), we see that *there will always be some point x close enough to p while $f(x)$ and $f(p)$ are certain distance apart.* Putting this into mathematical language, there exists $\epsilon > 0$ such that, for all $\delta > 0$, there exists x such that $d_X(x, p) < \delta$ but $d_Y(f(x), q) > \epsilon$.

We again construct a "$1/n$ sequence." Let $\delta_n := 1/n$. By what is stated above, to each δ_n there corresponds an x_n with $d_X(x_n, p) < 1/n$ but $d_Y(f(x), q) > \epsilon$. It is clear that $(x_n) \to p$, but on the other hand $f(p_n)$ is *not* covering to q. This contradicts the RHS of the $\Leftrightarrow$ statement, which was our assumption. Hence $\lim_{x\to p} f(x) = q$. □

> **Remark**. Note that in the proof of $\Longleftarrow$, our construction may fail if we do not require $d_X(x, p)$ to be strictly bigger than 0 (take the *boring sequence* for instance). Therefore it is necessary that we require $0 < d_X(x, p)$ in the Cauchy definition.

From the Heine definition, we see that the limit of a function at a point, should it exist, is equal to the limit of a sequence $(f(p_n))$. *Therefore, we naturally have the following results*:

Corollary 7.3

Limits of a function at a point, should they exist, are unique (by Lemma 3.5).

Corollary 7.4

Suppose $(Y, d_Y) = (\mathbb{R}, \|\cdot\|)$ and $\lim_{x\to p} f(x) = A$, $\lim_{x\to p} g(x) = B$ for $A, B \in \mathbb{R}$. Then (by Example 4.13)

$$\lim_{x\to p}(f+g)(x) := \lim_{x\to p}(f(x)+g(x)) = A+B, \lim_{x\to p}(fg)(x) := \lim_{x\to p}(f(x)g(x)) = AB$$

and

$$\lim_{x\to p}(f/g)(x) := \lim_{x\to p}(f(x)/g(x)) = A/B \qquad \text{if } B \neq 0.$$

Future reference: Corollary 7.6

Now, having defined the limit of a function at a point, we can move on and define continuity.

4.2 Continuous Functions

Definition 7.5: "ϵ-δ" Continuity

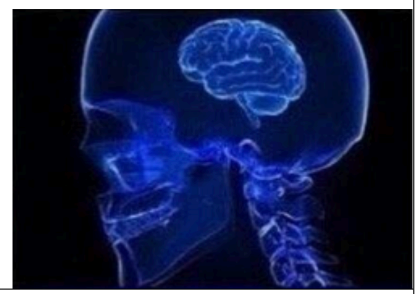
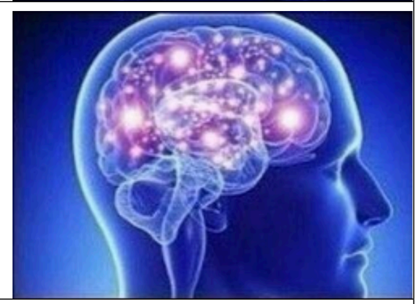
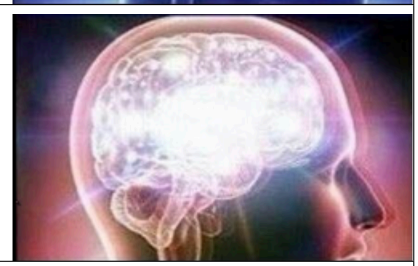
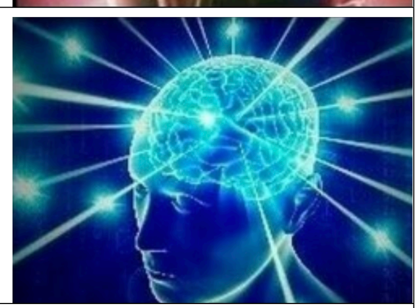
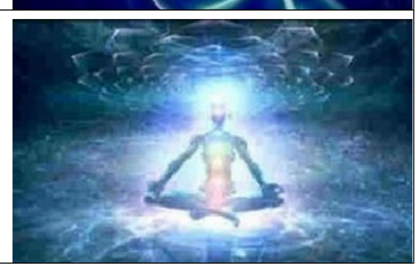
We say a function $f : (E \subset X) \to Y$ is **continuous at $p \in E$** if:

Given $\epsilon > 0$, there exists $\delta > 0$ such that $d(f(x), f(p)) < \epsilon$ whenever $d(x, p) < \delta$ and $x \in E$.

In other words, $\lim_{x\to p} f(x) = f(p)$ *if p is a limit point of E. (If p is an isolated point then f is trivially continuous at p: there exists a sufficiently small δ-neighborhood of p in which p is the only point. Then of course* $d(f(p), f(p)) = 0 < \epsilon$ *for any ϵ[!]) This definition holds for both isolated points too, so we don't need to require* $0 < d(x, p)$.

We say f is **continuous on E** if it is continuous at every $p \in E$. Future reference: Theorem 8.5

Remark. In particular, f needs to be defined at p. Unlike in previous definitions where p, a limit point, may or may not be in E, here we begin by picking a point in the domain. For example, $f(x) := 1/x$ defined on $\mathbb{R} - \{0\}$ is continuous, so "f is not continuous at 0" is false because it is not defined at 0 a priori.

What you were taught in...	A function is continuous if...	[2]
Middle school	If you can draw it without picking up your pencil	
High school	If it's like x^2	
Pre-calculus	If it does not have any holes or jumps	
"Calculus"	If for each c $\lim_{x \to c} f(x) = f(c)$	
Intro analysis (425a)	If for all $\epsilon > 0$ and all c there exists $\delta > 0$ such that $\lvert f(x) - f(c) \rvert < \epsilon$ whenever $\lvert x - c \rvert < \delta$	

[2] Borrowed from Jay Cummings, *Real Analysis: A Long Form Mathematics Textbook*, p.155.

Corollary 7.6

If $(Y, d_Y) = (\mathbb{R}, \|\cdot\|)$ and f, g are continuous functions from (X, d_X) to Y, then $f+g, fg, f/g$ are all continuous (f/g needs the additional assumption that $g(x) \neq 0$ for all $x \in X$). This is a direct result from Corollary 7.4.

Theorem 7.7: Composition of Continuous Functions

Let X, Y, Z be metric spaces. If $f : (E \subset X) \to Y$ is continuous at $p \in E$ and $g : (f(E) \subset Y) \to Z$ is continuous at $f(p)$, then the **composition**, written $g \circ f$ [i.e., $(g \circ f)(x) := g(f(x))$], is continuous at p.

Beginning of March 5, 2021

Proof. We fix $\epsilon > 0$ and let $p \in E$ be given. By continuity of g, since $f(p)$ is in its domain, there exists $\delta_1 > 0$ such that $d_z[g(y), g(f(p))] < \epsilon$ whenever $d_Y(y, f(p)) < \delta_1$ and $y \in f(E)$. In particular, $y \in f(E)$ means $y = f(x_0)$ for some $x \in X$.
Again, using continuity of f, and "treating δ_1 as the ϵ", there exists $\delta_2 > 0$ such that $d_Y(f(x_0), f(x)) < \delta_1$ whenever $d_X(x, x_0) < \delta_2$ and $x \in X$. Then for such δ_2,

$$d_X(x,p) < \delta_2 \implies d_Y(f(x), f(p)) < \delta_1 \implies d_Z[(g \circ f)(x), (g \circ f)(p)] < \epsilon.$$

□

Theorem 7.8: Continuity: Open Set Condition

A function $f : X \to Y$ is continuous if and only if the preimage of any open set $V \subset Y$ is open in X, i.e.,

$$V \subset Y \text{ open } \implies f^{-1}(V) \subset X \text{ open}.$$

*We define the **preimage** $f^{-1}(V)$ as $\{x \in X : f(x) \in v\}$. This is not to be confused with inverse of f, which may not exist, whereas $f^{-1}(V)$ always exists. They agree if and only if f is bijective: see Theorem 8.6.*
Future reference: Continuity & compactness, continuity & connectedness

Proof. We first show $\implies$. Assuming f is continuous, $V \subset Y$ open, and $x \in f^{-1}(V)$, we need to show that there exists a $\delta > 0$ such that $N_\delta(x) \subset f^{-1}(V)$ [i.e., showing $f^{-1}(V)$ is open]. Indeed, since V is open and $f(x) \subset V$, there exists some $\epsilon > 0$ such that $N_\epsilon(f(x)) \subset V$. Using the ϵ-δ definition of continuity, there indeed exists a $\delta > 0$ such that

$$d_X(x,p) < \delta \text{ and } p \in X \implies d_Y(f(x), f(p)) < \epsilon.$$

In particular, this shows that if $d_X(x,p) < \delta$ then $f(p) \in N_\epsilon(f(x))$, so $f(p) \in V$ and $p \in f^{-1}(V)$. Therefore $N_\delta(x) \subset f^{-1}(V)$, which completes the proof of $\implies$.
For $\impliedby$, suppose f satisfies the open set condition. Fix $\epsilon > 0$. We want to find $\delta > 0$ such that

$$d_X(x,y) < \delta \text{ and } y \in X \implies d_Y(f(x), f(y)) < \epsilon.$$

Since $V := N_\epsilon(f(x))$ is a neighborhood of $f(x)$, it is open by Lemma 3.9. By assumption $f^{-1}(V)$ is open in X. Of course, $x \in f^{-1}(V)$ as $f(x) \in N_\epsilon(f(x))$. Thus, by openness of $f^{-1}(V)$ there exists $\delta > 0$ such that

$$d_X(x,x') < \delta, \text{ i.e., } x' \in N_\delta(x) \implies x' \in f^{-1}(V) \implies d_Y(f(x'), f(x)) < \epsilon.$$

This is precisely the ϵ-δ definition of continuity. □

Corollary 7.9: Continuity: Closed Set Condition

$f : X \to Y$ is continuous if the preimage of any closed set is closed.
Future reference: Theorem 8.7, Dini's Theorem

Proof. Recall that openness is dual to closedness. With a little bit of computation, if $C \subset Y$ is closed then

$$f^{-1}\underbrace{(C^c)}_{open} = \left(f^{-1}(C)\right)^c$$

which implies $f^{-1}(C)$ is closed. □

Example 7.10. Let $(Y, d_Y) := (\mathbb{R}^k, \|\cdot\|)$, and we define $f : (E \subset X) \to Y$ by its components:

$$f(x) = (f_1(x), \ldots, f_k(x)).$$

Then f is continuous if and only if each f_i is continuous.

Proof. The proof is somewhat similar to that of Example 3.7. Both directions are given by

$$|f_i(x) - f_i(y)| \leqslant \|f(x) - f(y)\| = \left(\sum_{i=1}^{k} |f_i(x) - f_i(y)|^2\right)^{1/2}$$

(where $|\cdot|$ denotes absolute value in $\mathbb{R}$ and $\|\cdot\|$ denotes the standard Euclidean norm on $\mathbb{R}^k$). *If*

$$|f_i(x) - f_i(y)| < \frac{\epsilon}{\sqrt{k}} \qquad \text{for all } i,$$

then $\|f(x) - f(y)\| < \epsilon$. This means that if each component is convergent and if we want to ensure $\|f(x) - f(y)\| < \epsilon$, by the second $\leqslant$ we can just pick some sufficiently small δ such that $|f_i(x) - f_i(y)| < \epsilon/\sqrt{k}$ for all i. Conversely, if we want to ensure $|f_i(x) - f_i(y)| < \epsilon$ for all i [note that this would give convergence of all components], it suffices to ensure $\|f(x) - f(y)\| < \epsilon$ by the first $\leqslant$. □

Beginning of March 8, 2021

Example 7.11. Let $X := \mathbb{R}^n$ and $Y = \mathbb{R}$. Then the **coordinate functions** $\varphi_i(x) := x_i$ are continuous.

Example 7.12. The function $f(x) := 1/x$ (defined on $\mathbb{R} - \{0\}$) is continuous.

Proof using Cauchy definition. We fix $x \in \mathbb{R} - \{0\}$ and $\epsilon > 0$. We want to find $\delta > 0$ whose exact conditions will be specified later.[3] We have

$$|f(x) - f(y)| = \left|\frac{1}{x} - \frac{1}{y}\right| = \frac{|y - x|}{|xy|}.$$

In order to bound this from above, we want to bound $|xy|$ from below. Since

$$|y| \leqslant |x| + |y - x| \geqslant |x| - \delta,$$

[3] Just like the "$\epsilon/\sqrt{k}$ proof" involved in the component-wise convergence example, here we think "backwards" by doing the computations first and choosing the appropriate δ appropriately later.

we can choose δ smaller than $|x|/2$ so that $|y| \geqslant |x|/2$, and

$$\frac{|y-x|}{|xy|} \leqslant \frac{2}{|x|^2} \cdot |y-x|.$$

If it so happens that $|y-x| < \epsilon \cdot |x|^2/2$ we are done. Therefore,

$$|y-x| < \min(|x|/2, \epsilon|x|^2/2) \implies |f(x) - f(y)| < \epsilon.$$ □

Proof using Heine definition. First notice that every x in the domain $\mathbb{R} - \{0\}$ is a limit point. Now pick x and let $(x_n) \subset \mathbb{R} - \{0\}$ be any sequence that converges to x. Then $1/x_n \to 1/x$ and this completes the proof. □

Example 8.1.

(1) Any **monomial** (i.e., $x_1^{n_1} x_2^{n_2} ... x_k^{n_k}$) on $\mathbb{R}^k$ is continuous.

(2) Any polynomial is continuous on $\mathbb{R}^k$.

(3) Any **rational function**,, i.e., any function of the following form,

$$\frac{(x_1-a_1)^{n_1} ... (x_k-a_k)^{n_k}}{(x_1-b_1)^{m_1} ... (x_k-b_k)^{m_k}}.$$

is continuous on $\mathbb{R}^k - (\{x_1 = b_1\} \cup ... \cup \{x_k = b_k\})$.

(4) **Exponential functions** a^x is continuous for all $a > 0$. [*Will be proved later*]

Remark. *When studying continuity of a function, we only care about what happens in the domain of f. What happens in $X - E$ has no effect on our discussion on continuity.* Since E itself is also a metric space, it is more convenient to simply treat f as a mapping between metric spaces rather than mapping on a subset. In the future if we say $f : X \to Y$, we assume that f is defined on all of X.

4.3 Continuity and Compactness

Theorem 8.2: Continuous Mapping Preserves Compactness

If $f : X \to Y$ is continuous and if X is compact, then $f(X)$ is also compact.
Future reference: Theorem 8.7, Example 8.8

Proof. The main idea of this proof is to extract a finite subcover of $f(X)$. Let $\{A_\alpha\}_{\alpha\in I}$ be an arbitrary open cover of $f(X)$. We consider the pre-images $\{f^{-1}(A_\alpha)\}_{\alpha\in I}$. Since f is continuous, each $f^{-1}(A_\alpha)$ is also open by the open set condition. Notice that $\{f^{-1}(A_\alpha)\}_{\alpha\in I}$ covers X by definition of preimage (*for any* $x \in X$, $f(x)$ *is in the image so it lies in some* A_α *as* $\{A_\alpha\}_{\alpha\in I}$ *covers* $f(X)$.) Since X is compact it admits a finite subcover

$$X \subset f^{-1}(A_{\alpha_1}) \cup \ldots \cup f^{-1}(A_{\alpha_k}).$$

Notice that this implies $f(X) \subset A_{\alpha_1} \cup \ldots A_{\alpha_k}$. (*Indeed, if* $A \subset B$ *then* $f(A) \subset f(B)$ *and* $f(A \cup B) = f(A) \cup f(B)$.) Hence we have extracted a finite subcover of $f(X)$, and this completes the proof. □

Definition 8.3: Bounded Functions

We say a function $f : X \to \mathbb{R}^k$ is **bounded** if there exists $M > 0$ such that $\|f(x)\| \leqslant M$ for all $x \in X$. *Sometimes we say* $f(M)$ *is bounded as a shorthand notation.*

Lemma 8.4: Continuous Functions on Compact Sets are Bounded

If $f : X \to \mathbb{R}^k$ is continuous and if X is compact, then f is bounded.
One-liner proof: by the previous theorem $f(X)$ *is compact and therefore bounded.*
Future reference: Theorem 11.3, $(C(K), \|\cdot\|_{\sup})$ is complete, Weierstraß Approximation Theorem

Theorem 8.5: Continuous Functions on Compact Sets Attain Bounds

If $f : X \to \mathbb{R}$ is continuous and X is compact, then f *attains* its maximum and minimum[!] In other words, there exists $p, q \in X$ such that

$$f(p) = \sup_{x\in X} f(x) \qquad \text{and} \qquad f(q) = \inf_{x\in X} f(x).$$

This is, of course, false if X *is not compact: for example* $f(x) := x$ *on* $(0,1)$ *has supremum* 1 *and infimum* 1 *but clearly neither is attained.*
Future reference: Rolle's Theorem, Darboux Property, continuous functions are R-S integrable, minimization problems

Proof. We will only prove the case for supremum; the other is analogous. We define

$$M := \sup_{x\in X} f(x) = \sup f(X).$$

By the characterization of supremum, there exists a sequence $(y_n) \subset f(X) \subset \mathbb{R}$ that converges to M. It follows that there exists a sequence (x_n) such that $f(x_n) = y_n$.

A sequence in a compact set? It follows naturally that we should extract a convergent subsequence. Let (x_{n_k}) be that such subsequence and suppose $x_{n_k} \to p$ for some $p \in X$. By "Heine continuity", since f is continuous, $\lim_{k\to\infty} f(x_{n_k}) = f(p)$.

It remains to notice that $y_{n_k} = f(x_{n_k})$: since the mother sequence $(y_n) \to M$, so does the convergent subsequence (y_{n_k}). This means precisely that $(y_{n_k}) = (f(x_{n_k})) \to M$, so $f(p) = M$ by the uniqueness of limits. This completes our proof. □

Theorem 8.6: Inverse of a Bijection

If $f : X \to Y$ is a bijection then $f^{-1} : Y \to X$ exists and is also a bijection. *After all, f is a bijection so each $x \in X$ uniquely corresponds to some $y \in Y$. Then we can simply define f^{-1} by mapping y back to x for each $x \in X$. That f is bijective (in particular, surjective) implies f^{-1} is defined on all of Y.* Also, for such cases, $f^{-1}(V)$ and "the preimage of V" are equivalent, whereas in general they are not, as discussed in open set condition.

Theorem 8.7: Compactness, Bijection, & Inverses

If X is compact and $f : X \to Y$ a continuous bijection (*note that $Y = f(X)$ is also compact*), then the inverse mapping f^{-1} is also continuous. *However, the theorem is false if we omit the assumption that X is compact; an example will be shown in the next example.*

Proof. *We will use the closed set condition for this proof.* Let $G \subset X$ be closed (in X); we will show that $(f^{-1})^{-1}(G) = f(G)$, the set which f^{-1} maps to G, is closed (in Y).
This is in fact immediate: notice that G is a closed subset of a compact metric space so it is compact itself (Example 4.1). Then, since f is continuous, $f(G)$ is compact (Theorem 8.2). In particular, this implies $f(G) \subset Y$ is closed (Theorem 3.15). The claim follows. □

Beginning of March 10, 2021

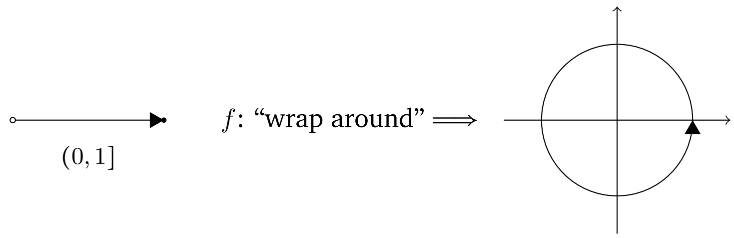
Example 8.8. This example shows the necessity of X's compactness in the previous theorem. Consider $X := [0,1)$[4] and $Y := S^1 = \{z \in \mathbb{C} : |z| = 1\}$. We claim that

$$f(x) := (\cos 2\pi x, \sin 2\pi x)$$

is a bijection but it does not admit a continuous inverse.

Proof. The diagram below visualizes what f and f^{-1} do. The main idea of the proof is to consider a sequence of points on the circle approaching the point $(1,0)$ from below; these points get mapped to the right side of $[0,1)$ via f^{-1} whereas $(1,0)$ gets mapped to the leftmost $0 \in [0,1)$. We see that f^{-1} does not preserve sequential continuity and is therefore not continuous.

[4] The example given in lecture uses $X := (0,1]$ but I found that drawing the diagram for $[0,1)$ is easier so... Of course, both examples hold.



□

Alternate Proof. Notice that Y is compact:

$$Y = \overline{B(1)} \cap B(1)^c$$

where $B(1)$ denotes the open ball of radius 1. This shows Y is the intersection of two closed balls and is therefore closed. It is obviously bounded, so Heine-Borel gives compactness.

If f^{-1} is compact then $f^{-1}(Y) = X$ is compact by Theorem 8.2, but $[0, 1)$ is not. Therefore f^{-1} is not continuous. □

4.4 Uniform Continuity

Now we present a "stronger" continuity:

Definition 8.9: Uniform Continuity

We say $f : X \to Y$ is **uniformly continuous** if:

Given $\epsilon > 0$, there exists $\delta > 0$ such that $d(f(x), f(y)) < \epsilon$ whenever $d(x, y) < \delta$.[5]

Note that this time our δ needs to work for all $x, y \in X$, whereas previously in the "ordinary" continuity, we can pick δ after the point of interest x is given.

Example 8.10. $f(x) := x^2$ from $[0, \infty)$ to $[0, \infty)$ is continuous but not uniformly continuous. *It is a polynomial and so it is indeed continuous.*

Proof. Heuristically, f becomes increasingly steep, so given ϵ, the corresponding δ needs to be smaller and smaller as x gets larger and larger. Therefore we cannot find some δ that works for all $x > 0$.
Suppose for contradiction that it is indeed uniformly continuous. We fix $\epsilon = 1$. By assumption there exists $\delta > 0$ such that

$$|x - y| < \delta \implies |x^2 - y^2| < \epsilon.$$

Now we just need to find sufficiently large x that brings a contradiction. We take $x := 1/\delta$ and let $y := x + \delta$. A simple calculation suggests that

$$|x^2 - y^2| = |(x - y)(x + y)| = \delta(x + y) > 2\delta x > 1 = \epsilon.$$

Contradiction. Hence f is not uniformly continuous. □

Theorem 8.11: Continuous and Compact Domain ⇒ Uniformly Continuous

If $f : X \to Y$ is continuous and X is compact, then f is uniformly continuous.
This shows us the power of compactness — just because the domain is compact, we are able to "upgrade" continuity into a much stronger form.
Future reference: Example 8.14, continuous functions are R-S integrable, Weierstraß Approximation Theorem

Proof. Suppose for contradiction that f is not uniformly continuous. This means that there exists some $\epsilon > 0$, there does not exist "the δ that works for all points in X". Therefore, for such $\epsilon > 0$, for all $\delta > 0$ we can find $x, y \in X$ such that $d(x, y) < \delta$ but $d(f(x), f(y)) \geqslant \epsilon$.
Let $\delta_n := 1/n$. It follows that for each n, there exists $x_n, y_n \in X$ with $d(x_n, y_n) \geqslant< \epsilon$ and $d(f(x_n), f(y_n)) \geqslant \epsilon$. Since $(x_n), (y_n)$ form two sequences in X, using sequential compactness, we can extract a convergent subsequence of (x_n), and based on the y_n's corresponding to this subsequence, we can further construct a sub-subsequence, whose indices we label as n_k, such that both (x_{n_k}) and (y_{n_k}) converge. Suppose $(x_{n_k}) \to x$ and $(y_{n_k}) \to y$. By triangle inequality

$$d(x, y) \leqslant d(x, x_{n_k}) + d(x_{n_k} + y_{n_k}) + d(y_{n_k}, y)$$

[5] We will drop the cumbersome subscripts $d_Y(f(x), f(y))$ and $d_X(x, y)$ when the context is clear, i.e., $f(x) \in Y$ and $x \in X$.

where the first and third term tends to 0 by convergence and the second is bounded by $1/n$. As $n \to \infty$, $d(x, y) \to 0$, so $x = y$. By continuity of f, we have $f(x_{n_k}) \to f(x)$ and $f(y_{n_k}) \to f(x)$. *It then follows that late terms from two sequences can be made arbitrarily close to the limit, and at the same time they must be arbitrarily close to each other as well. This gives a contradiction*. Using triangle inequality,

$$\epsilon \leqslant d(f(x_{n_k}), f(y_{n_k})) \leqslant d(f(x_{n_k}), f(x)) + d(f(x), f(y_{n_k}))$$

where the RHS tends to 0. Therefore $\epsilon \leqslant 0$, contradicting our assumption on $\epsilon > 0$. Hence f is uniformly continuous.

□

Definition 8.12: Lipschitz Continuity

We say $f : (E \subset R^n) \to \mathbb{R}^m$ is **Lipschitz** if there exists a **Lipschitz constant** $L > 0$ such that

$$\|f(x) - f(y)\| \leqslant L\|x - y\| \qquad \text{for all } x, y \in E.$$

If $f : \mathbb{R} \to \mathbb{R}$, then that the slope/derivative at each point on the graph is always bounded by $\pm L$.

Lemma 8.13: Lipschitz ⇒ Uniformly Continuous

Any Lipschitz function is uniformly continuous. *Proof: take ϵ/L.*
Future reference: FTC part 1

Example 8.14: Uniformly Continuous ⇏ Lipschitz. This example shows that being Lipschitz is a *stronger* condition than being uniformly continuous. For example $f : [0, 1] \to [0, 1]$ by $x \mapsto \sqrt{x}$ is uniformly continuous but not Lipschitz.

Proof. f is uniformly continuous because $[0, 1]$ is compact and $\sqrt{x}$ is clearly continuous (we used Theorem 8.11 here). We now show that f is not Lipschitz.
Suppose for contradiction that f is Lipschitz; in particular f is Lipschitz at $x = 0$. Then there exists $L > 0$ with

$$|f(y) - f(x)| = |f(y)| = \sqrt{y} \leqslant L|y - x = Ly \qquad \text{for all } y \in [0, 1].$$

In particular, for all nonzero y we have $\sqrt{y} \leqslant Ly$, so dividing by $\sqrt{y}$ gives $1 \leqslant L\sqrt{y}$ for all $y \in (0, 1]$. Letting $x \to 0$, the LHS remains 1 whereas the RHS $\to 0$, and clearly $1 \leqslant 0$ gives a contradiction. Hence f is not Lipschitz. □

Remark. We conclude this section by a summary of the "hierarchy of continuities":

$$\begin{aligned} \text{Lipschitz} &\implies \text{Uniformly Continuous} \implies \text{Continuous} \\ \text{Continuous} &\not\Longrightarrow \text{Lipschitz} : f(x) := \sqrt{x} \text{ from } [0, 1] \to [0, 1] \\ \text{Continuous} &\not\Longrightarrow \text{Uniformly Continuous} : f(x) := x^2 \text{ from } [0, \infty) \to [0, \infty) \end{aligned}$$

4.5 Continuity and Connectedness

Definition 8.15: Separated, Connected, & Path-Connected Sets

Let $A, B \subset X$.

(1) We say A, B are **separated** if $A \cap \overline{B} = \overline{A} \cap B = \varnothing$.

For the closure part, think of two tangent balls, one open and one closed. By definition they are not separated; however, if they are further apart and no longer tangent, then they are indeed separated.

(2) We say A is **connected** if $A \neq C \sqcup D$ for any separated and nonempty sets C, D.[6] *Heuristicaly, A consists of one piece.*

(3) We say A is **path-connected** if for all $a, b \in A$, there exists a continuous function $f : [0,1] \to A$ such that $f(0) = a$ and $f(1) = b$. *In other words, any two points in A can be connected by a (continuous) path.*

Example 8.16. If $c \in (a,b)$ then $A := [a,b] - \{c\}$ is disconnected. Indeed, A is the union of two separated intervals $[a,c)$ and $(c,b]$.

Example 8.17: Closed Topologist's Curve is Not Path-Connected. The closed topologist's curve is connected but *not* path-connected!
Future reference: Corollary 9.2, Example 9.12

Proof sketch. The proof is omitted by this course, but since my 425a included this as a HW problem, I will briefly provide an outline below. Notation: let S be the closed topologist's curve, $S_1 :=$ the graph of $\sin(1/x)$, and $S_0 :=$ the line segment $\{0\} \times [-1,1]$. It follows that $S = S^+ \cup S_0$ by definition.

(1) S is the closure of S^+. Since S^+ is connected so is S. (If $\overline{E} = A \sqcup B$ for separated A, B, we can write $E = (A \cap E) \cup (B \cap E)$. Since E is connected, WLOG assume $A \cap E = E$ and $B \cap E = \varnothing$. Then $E \subset A$ and thus $\overline{E} \subset \overline{A}$. Since A, B are separated, $B \cap \overline{E} \subset B \cap \overline{A} = \varnothing$, completing the proof.)

(i) It is clear that $S^+ \subset S$. Also, for every $x \in S_0$, draw a horizontal line and we see that there indeed exists a sequence on S^+ converging to x. Hence $S^+ \cup S_0 \subset S$.

(ii) Any limit point of S^+ must have nonnegative x-coordinate and y coordinate in $[-1,1]$. Besides limit points in $\{0\} \times [-1,1]$, if it is in $(x_0, y_0) \in (0,\infty) \times [-1,1]$, then some sequence $(x_n, \sin(1/x_n))$ converges to (x_0, y_0). In particular $(x_n) \to x_0$ and by continuity of $\sin(1/x)$ we also have $\sin(1/x_n) \to \sin(1/x_0)$, i.e., the only possible limit points with positive x-coordinate are those already in S. This shows the other inclusion and thus $S = \overline{S^+}$.

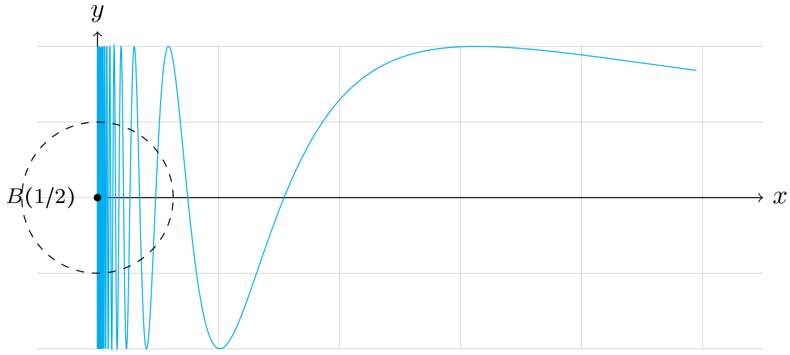
(2) S is not path-connected; in particular no path exists between $p \in S_0$ and $x \in S^+$.

Suppose S is path connected. Pick $p := (0,0)$ any $x \in S^+$, and there should exist a continuous $f : [0,1] \to S$. In particular, we let $\epsilon := 1/2$ and there should exist some δ such that

$$|f(t) - f(0)| < \epsilon \textit{ whenever } |t - 0| < \delta.$$

[6] The symbol $\sqcup$ stands for disjoint union. Pugh defines *clopen* to be closed and open, and in his definition a set is connected if it does not have a proper clopen subset. A set that is not connected is **disconnected**, and we will adopt this definition too.

This is "clearly" impossible if we look at the graph: the topologist's curve keeps jumping in and out of the circle with radius $1/2$ centered at the origin. With a bit of effort we can show this rigorously, but since Ożański's focus isn't on this example I shall omit it.



(Analogously, we can generalize the case $p = (0,0)$ to any $p \in S_0$.) These conclude the proof. □

Theorem 8.18: Connected Subsets of $\mathbb{R}$

If $A \subset \mathbb{R}$ then A is connected if and only if for all $x, y \in A$, $(z \in (x,y) \Rightarrow z \in A)$. *Put informally, A is connected if and only if it is an closed/open/clopen interval.*
Future reference: Corollary 9.2

Proof. We first show $\implies$. Suppose for contradiction that A is connected and $x, y \in A$, but there exists $z \in (x,y)$ with $z \notin A$. Then immediately we see that we can write A as

$$A = (A \cap (-\infty, z)) \sqcup (A \cap (z, +\infty)).$$

By assumption neither set is empty as $x < z < y$. On the other hand, these two sets are also separated. Hence A is disconnected, contradiction.
For $\impliedby$, suppose A is not connected but for all $x, y \in A$, $(z \in (x,y) \Rightarrow z \in A)$. Then we can write $A = C \sqcup D$ where C, D are nonempty and separated. Pick $x \in C$ and $y \in D$. Clearly they are different points; WLOG assume $x < y$. We consider the interval $[x,y]$ and define $z := \sup([x,y] \cap C)$. *This set contains x and is therefore nonempty; it is clearly bounded above by y. Thus the supremum exists.*
By Example 3.11, z is a limit point of $[x,y] \cap C$. Therefore $z \in \overline{[x,y] \cap C} = [x,y] \cap \overline{C}$. In particular $z \in \overline{C}$. By the assumption that $\overline{C} \cap D = \varnothing$ we see $z \notin D$. Since $y \in D$ and $z \in [x,y]$, we see that $z < y$ (*the equality cannot be attained*). *We will show that either $z \in C$ or $z \notin C$ gives contradiction.*
If $z \notin C$, then since $x \in C$ and $x \leqslant z$ we see $x < z$. Then by assumption $z \in (x,y)$ and so $z \in A$, but we have said $z \notin C$ and $z \notin D$, contradiction.
If $z \in C$ then $z \notin \overline{D}$ since C, D are separate. *In particular, no sequence in D converges to z, i.e., no points in D can be arbitrarily close to z.* Therefore there exists $\epsilon > 0$ such that $(z - \epsilon, z + \epsilon) \cap (D) = \varnothing$. (*In particular we focus on the $(z, z+\epsilon)$ half of that interval.*) In particular, since $z < y$ as well, there exists $z_1 \in (z,y) \cap (z, z+\epsilon)$. By assumption such $z_1 \notin D$. Also notice that such z_1 is also not in C: it is in $[a,b]$ and it is greater than z, the supremum of $[a,b] \cap C$ so it *must not* be in C. Then $z_1 \in (z,y)$ where $z, y \in A$ but $z_1 \notin A$, contradiction. □

Lemma 9.1: Path Connected ⇒ Connected

If $A \subset X$ (X being any metric space) is path connected then A is connected.

Proof. We suppose for contradiction that A is path connected but not connected. This by definition means that there exist nonempty, separated C, D such that $C \sqcup D = A$. We take $x \in C$ and $y \in D$. By path-connectedness there exists a continuous $f : [0,1] \to A$ with $f(0) = x, f(1) = y$. Then,

$$[0,1] = [0,1] \cap f^{-1}(A) = [0,1] \cap \left(f^{-1}(C \sqcup D)\right) = \left([0,1] \cap f^{-1}(C)\right) \sqcup \left([0,1] \cap f^{-1}(D)\right).$$

Notice that $f(x) = 0$ implies $0 \in f^{-1}(C)$, so $0 \in [0,1] \cap f^{-1}(C)$. Likewise $1 \in [0,1] \cap f^{-1}(D)$. On the other hand, $f^{-1}(C)$ and $f^{-1}(D)$ are separated:

$$C, D \text{ separated} \implies \overline{C} \cap D = \varnothing \implies \overline{f^{-1}(C)} \cap f^{-1}(D) \subset f^{-1}(\overline{C}) \cap f^{-1}(D) = \varnothing$$

and likewise $\overline{f^{-1}(D)} \cap f^{-1}(C) = \varnothing$. (*We have taken it for granted that $\overline{f^{-1}(C)} \subset f^{-1}(\overline{C})$: indeed, $f^{-1}(C) \subset f^{-1}(\overline{C})$ so taking the closure does not affect $\subset$. The converse, however, is false: see here.*[7]) But then we see that $[0,1]$ admits a separation, contradiction. Therefore A must be connected. □

Corollary 9.2: Path-Connected in $\mathbb{R}$ ⇔ Connected in $\mathbb{R}$

In $\mathbb{R}$, a set is path-connected if and only if it is connected. *Notice that the closed typologist's curve shows that this claim is invalid in $\mathbb{R}^2$!*

Proof. The forward direction $\implies$ follows from the previous lemma.
Now we show the converse. Let $A \subset \mathbb{R}$ be connected and let $x, y \in A$. It follows from Theorem 8.18 that every point in (x, y) is also in A. It remains to notice that the map $f : [0,1] \to A$ defined by

$$f(t) = (1-t)x + ty$$

is continuous and satisfies $f(0) = x, f(1) = y$ (*we parametrized the line segment between x and y*). □

Theorem 9.3: Continuous Mapping Preserves Connectedness

If $f : X \to Y$ is continuous and $E \subset X$ connected, then $f(E)$ is connected.

Proof. Suppose for contradiction that $f(E)$ is not connected. This means that there exist nonempty, separated C, D such that $f(E) = C \sqcup D$. Notice that

$$E = f^{-1}(E) = E \cap F^{-1}(E) = \underbrace{\left(E \cap f^{-1}(C)\right)}_{=:G} \sqcup \underbrace{\left(E \cap f^{-1}(D)\right)}_{=:H}.$$

Since C and D are nonempty, there exists $y \in E$ and $y = f(x)$ for some $x \in E$. Then $x \in f^{-1}(C) \cap E$, so G is nonempty. Likewise, H is nonempty. Now it remains to show that G and H are separated (which contradicts E's connectedness).
We will show that $\overline{G} \cap H = \varnothing$; the other argument is also valid. Since $G = E \cap f^{-1}(C)$, in particular we have $G \subset f^{-1}(C) \subset f^{-1}(\overline{C})$. Since $\overline{C}$ is closed and f continuous, the open set condition asserts that $f^{-1}(\overline{C})$ is also closed. Hence $G \subset f^{-1}(\overline{C})$ also implies $\overline{G} \subset f^{-1}(\overline{C})$. Therefore

$$f(\overline{G}) \subset f\left(f^{-1}(\overline{C})\right) \implies f(\overline{G}) \subset \overline{C}. \tag{1}$$

[7] Thus the lecture made a small logical loophole here by directly claiming $f^{-1}(\overline{C}) = \overline{f^{-1}(C)}$

In addition, since H is clearly a subset of $f^{-1}(D)$, we have

$$f(H) \subset f\left(f^{-1}(D)\right) \implies f(H) \subset D. \tag{2}$$

Since C and D are separated by assumption, $\overline{C} \cap D = \varnothing$ and so $f(\overline{G}) \cap f(H) = \varnothing$ by using (1) and (2). This means that $f(\overline{G})$ and $f(H)$ share no common element, so indeed we have $\overline{G} \cap H = \varnothing$, contradiction. □

We now present a powerful application of connectedness and continuity:

Theorem 9.4: Intermediate Value Theorem (IVT) / the Bolzano Theorem

Let $f : [a, b] \to \mathbb{R}$ be continuous. If for some interval $[a, b]$ we have $f(a) \leqslant f(b)$ [resp. $f(a) \geqslant f(b)$], then for any $c \in [f(a), f(b)]$, there exists $x \in [a, b]$ with $f(x) = c$.
Future reference: Example 9.5, Darboux Property, Theorem 11.1, MVT for integrals

Proof. Since $[a, b]$ is connected (by Theorem 8.18) and d is continuous, $f([a, b])$ is also connected (by the previous theorem). Clearly $f(a), f(b) \in f([a, b])$, so by using Theorem 8.18 once more we see that if $c \in [f(a), f(b)]$ we must have $c \in f([a, b])$, i.e., $c = f(x)$ for some $x \in [a, b]$. □

Example 9.5: Applications of the IVT.

(1) If $f : [0, 1] \to [0, 1]$ is continuous, then f has a **fixed point**, i.e., there exists $x_0 \in [0, 1]$ with $f(x_0) = x_0$. This is also the $\mathbb{R}^1$ case of the *Brouwer Fixed-Point Theorem*.

Proof. Take $g(x) := f(x) - x$ a continuous mapping with $g(0) = f(0) \geqslant 0$ and $g(1) = f(1) - 1 \leqslant 0$. Using IVT there exists $x_0 \in [0, 1]$ with $g(0) = f(x) - x = 0$, completing our proof. Note that $[0, 1]$ can be easily replaced by $[a, b]$; the claim and the proof follows almost identically. □

(2) There exists some $x \in [0, 4]$ such that $2^x = \sqrt{x} + 2$. *Proof. Take $f(x) := 2^x - \sqrt{x} - 2$. Then f is continuous (take 2^x to be continuous for granted now... we will prove this later) with $f(0) < 0 < f(4)$. Therefore $f(x) = 0$ for some $x \in (0, 4)$, completing the proof.* □

Remark. *I would personally like to add a remark here which connects to the pathological examples in Example 9.12.* A differentiable function's derivative cannot have a jump discontinuity. Otherwise, using IVT on the derivative on a small interval containing that jump, we obtain a contradiction.
However, this doesn't mean the derivative of a differentiable function must be continuous — our pathological examples will show that they can indeed fail to be continuous.

4.6 Monotonicity & Discontinuity

Beginning of March 17, 2021

Theorem 9.6: Monotonic Functions & One-Sided Limits

If $f : (a, b) \to \mathbb{R}$ is **monotonic** (i.e., increasing, decreasing, nonincreasing, or nondecreasing), then the *one-sided limits* always exist, i.e.,

$$f(x_-) := \lim_{t \to x_-} f(t) \qquad \text{and} \qquad f(x_+) := \lim_{t \to x_+} f(t)$$

exist for all $x \in (a, b)$.

If f is nondecreasing, then

$$\sup_{t \in (a,x)} f(t) = f(x_-) \leqslant f(x) \leqslant f(x_+) = \inf_{t \in (x,b)} f(t) \tag{1}$$

and in addition

$$f(x_+) \leqslant f(y_-) \qquad \text{for} \qquad x < y. \tag{Eq.9.1}$$

We can make analogous statements if f is nonincreasing.

Proof. Define $A := \sup_{t \in (a,x)} f(t)$ and $B := \inf_{t \in (x,b)} f(t)$. Suppose f is nondecreasing; this means that $f(t) \leqslant f(x)$ for all $t \in (a, x)$, so $f(x)$ is an upper bound for $\{f(t) : t \in (a, x)\}$. In particular, since A is the *least* upper bound, $A \leqslant f(x)$.
Now we show that $A = f(x_-)$. Let $\epsilon > 0$ be given. Our goal is to show that the supremum A is arbitrarily close to $f(x_-)$. Since $A - \epsilon$ cannot be an upper bound of $\{f(t) : t \in (a, x)\}$, there must exist some $x - \delta \in (a, x)$ with

$$A - \epsilon < f(x - \delta) \leqslant A \underbrace{\leqslant A + \epsilon}_{\text{trivially}}.$$

Since f is nondecreasing and every $f(t)$ for $t \in (x - \delta, x)$ is bounded above by A, we have

$$A - \epsilon < f(t) \leqslant A \underbrace{\leqslant A + \epsilon}_{\text{trivially}} \Leftrightarrow |f(t) - A| < \epsilon \qquad \text{for all } t \in (x - \delta, x).$$

This is precisely the definition of one-sided limit as shown in the footnote of Definition 7.1. We now have $A = f(x_-)$ and $A \leqslant f(x)$, and the other half of the equation follows analogously.
Equation 9.1 is obtained by applying (1) and monotonicity, which allows us to extend the set on which supremum and infimum are taken:

$$f(x_+) = \inf_{t \in (x,b)} f(t) = \inf_{t \in (x,y)} f(t) \leqslant \sup_{t \in (x,y)} f(t) = \sup_{t \in (a,y)} f(t).$$

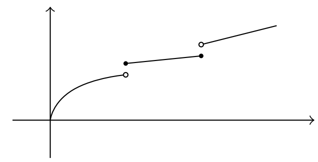
□

Corollary 9.7

Let $f : (a, b) \to \mathbb{R}$ be nondecreasing (*in this example we allow a, b to take values of $\pm\infty$*).

(1) If f is **discontinuous** (not continuous) at some $x \in (a, b)$ then f has a **jump** at x, i.e., $f(x_-) < f(x_+)$.

For exampmple, the graph below has two points of discontinuity, both of which are jumps.



(2) f has at most countably many points of discontinuity.

(*The claim for a nonincreasing function is similar.*)

Proof. (1) If $f(x_-)$ is not strictly less than $f(x_+)$ then by the preceding theorem $f(x_-) = f(x_+)$. Therefore $\lim_{t\to x} f(t) = f(x)$ so f is continuous at x. Contradiction.

(2) *Main idea: notice that if $x < y$ are two discontinuity points, then we already know $f(x_-) < f(x_+)$ and $f(y_-) < f(y_+)$. By the preceding theorem we also know $f(x_+) < f(y_-)$, so using everything together we see that $(f(x_-), f(x_+))$ and $(f(y_-), f(y_+))$ are two disjoint open intervals. In particular this means that any two different discontinuity points correspond to two disjoint open intervals.*

From the italicized text we see that there exists a one-to-one correspondence between x, a discontinuity point, and $(f(x_-), f(x_+))$, an open interval. Since each open interval contains *some* rationals and rational numbers are countable, there are at most countably many disjoint open intervals in $\mathbb{R}$. This means that there are at most countably many discontinuity points, proving our claim. □

Chapter 5

Differentiation

5.1 Derivatives

First, some very basic definitions that we are all familiar with...

Definition 9.8: Differentiable Functions & Derivatives

Let $f : [a,b] \to \mathbb{R}$. If

$$\lim_{h\to x} \frac{f(x+h) - f(x)}{h} \qquad \text{or equivalently } \lim_{y\to x} \frac{f(y) - f(x)}{y - x}$$

exists and is finite for $x \in [a,b]$, we say f is **differentiable at x** and say the **derivative** of f at x, written $f'(x)$, equals that limit. *Note that the derivative at endpoints a, b may exist.*

If f is differentiable at every $x \in E$ for $E \subset [a,b]$ then we say f is **differentiable on E**; if it's differentiable on all of $[a,b]$ then it is **differentiable** (on $[a,b]$).

Lemma 9.9: Differentiable ⇒ Continuous

If f is differentiable at x then it's continuous at x.

Future reference: Higher order derivatives

Proof. If $y \to x$, then $|f(y) - f(x)| = \dfrac{|f(y) - f(x)|}{|y - x|} \cdot |y - x| = \underbrace{f'(x)}_{\in\mathbb{R}} \cdot |y - x| \to 0.$ □

Lemma 9.10

Let $f, g; [a,b] \to \mathbb{R}$ be functions that are differentiable at $x \in [a,b]$. Then

(1) $(f+g)'(x) = f'(x) + g'(x)$.

(2) (**Leibniz product rule**) $(fg)'(x) = f'(x)g(x) = f(x)g'(x)$.

(3) (**Quotient rule**) $(f/g)'(x) = [f'(x)g(x) - f(x)g'(x)]/g^2(x)$ if $g(x) \neq 0$.

Proof of the quotient rule. Adding and subtracting (*recall this is a classic trick when dealing with limits of sequences*) $f(x)g(x)$ in the numerator, we obtain

$$\frac{f(y)}{g(y)} - \frac{f(x)}{g(x)} = \frac{g(x)(f(y) - f(x)) - f(x)(g(y) - g(x))}{g(y)g(x)}. \qquad (\Delta)$$

As $y \to x$, the denominator $g(y)g(x) \to g^2(x)$. On the other hand

$$f(y) - f(x) \to f'(x)(y-x) \qquad \text{and} \qquad g(y) - g(x) \to g'(x)(y-x)$$

as $y \to x$, so dividing (Δ) by $y - x$ yields

$$(f/g)'(x) = \lim_{y\to x} \frac{(f/g)(y) - (f/g)(x)}{y-x} = \frac{g(x)f'(x) - f(x)g'(x)}{g^2(x)},$$

as claimed. ☐

Theorem 9.11: Chain Rule

Suppose $f : [a,b] \to \mathbb{R}$ is differentiable at some x and $g : I \to \mathbb{R}$ is differentiable at $f(x) \in I$.[1]Then

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

Heuristically (after all, we haven't justified the rigorousness of differential operator), if we write $h := g \circ f$ *then the chain rule states that* $\frac{\mathrm{d}h}{\mathrm{d}x} = \frac{\mathrm{d}h}{\mathrm{d}f}\frac{\mathrm{d}f}{\mathrm{d}x}$. *The chain rule "cancels out" the* $\mathrm{d}f$. *Of course, this can get more crazy: for example a trample nested composite function gives* $\frac{\mathrm{d}h}{\mathrm{d}x} = \frac{\mathrm{d}h}{\mathrm{d}g}\frac{\mathrm{d}g}{\mathrm{d}f}\frac{\mathrm{d}f}{\mathrm{d}x}$.

Proof. By definition we want to compute $\frac{g(f(y)) - g(f(x))}{y-x}$. Note that by using the definition of derivative of g at $f(x)$, we have

$$g'(f(x)) = \lim_{f(y)\to f(x)} \frac{g(f(y)) - g(f(x))}{f(y) - f(x)},$$

so in particular the error function

$$\text{err}(z) := \begin{cases} \dfrac{g(z) - g(f(x))}{z - f(x)} - g'(f(x)) & \text{if } z \neq f(x) \\ 0 & \text{if } z = f(x) \end{cases}$$

tends to 0 as $z \to f(x)$. Now it remains to notice that

$$g(f(y)) - g(f(x)) = (f(y) - f(x))\left[g'(f(x)) + \text{err}(f(y))\right].$$

As $y \to x$, continuity of f implies $f(y) \to f(x)$, and so $\text{err}(f(y)) \to 0$, and we obtain

$$\frac{g(f(y)) - g(f(x))}{y-x} \to \frac{f(y) - f(x)}{y-x} g'(f(x)) = g'(f(x))f'(x).$$ ☐

Example 9.12: Examples of Derivatives.

(1) Some very basic examples: $f(x) = c \Rightarrow f' \equiv 0$. The derivative of x^n is nx^{n-1}; the exponential function e^x is invariant after taking derivative (*and it is the only function that is equal to its own derivative*); $\sin(x)$ has derivative $\cos(x)$; and $\cos(x)$ has derivative $-\sin(x)$.

(2) The topologist's curve is defined by $f(x) = \sin(1/x)$. Thus

$$f'(x) = -\frac{\cos(1/x)}{x^2},$$

[1] From now on, when saying "$f : [a,b] \to \mathbb{R}$ is continuous", we will use the notation $f \in C([a,b]) = C([a,b];\mathbb{R})$. This means that f is a Continuous function with domain $[a,b]$ and codomain $\mathbb{R}$. We will see these in details later when we talk about space of continuous functions.

so the topologist's sine curve is differentiable on $(0,\infty)$. It is not defined at $x=0$, but it cannot be extended to a continuous function including 0 in its domain, either. (*The main idea is still to use the diagram drawn in Example 8.17.*)

(3) More pathological examples involving $\sin(1/x)$:

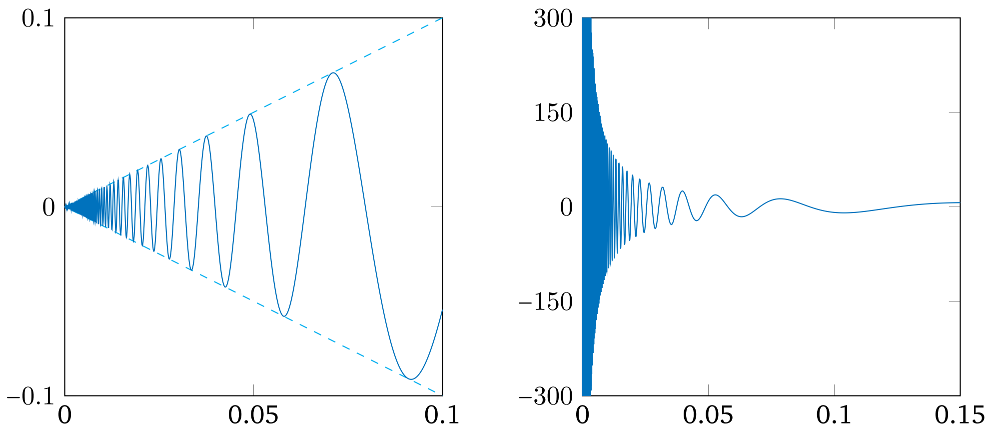
(i) If we define

$$f(x) := \begin{cases} x\sin(1/x) & x \neq 0 \\ 0 & \text{otherwise,} \end{cases}$$

it is clear that $f'(x) = \sin(1/x) - 1/x\cdot\cos(1/x)$ for $x \neq 0$, so x is differnetiable on $(-\infty,0)\cup(0,\infty)$. Moreover, since f is bounded between $y=x$ and below by $y=-x$ (since $-1 \leqslant \sin(\cdot) \leqslant 1$), from the graph below it is clear that $\lim_{x\to 0} f(x)$ exists and equals 0. Hence setting $f(x) := 0$ at $x=0$ makes this a continuous function on all of $\mathbb{R}$. However, this function is *not* differentiable at $x=0$:

$$\frac{f(y)-f(0)}{y-0} = \frac{f(y)}{y} = \sin(1/y)$$

which does *not* converge as $y \to 0$. (*Look at the topologist's curve again and take the ball $B(1/2)$, for example.*) Below are the graphs for f and f'.



(ii) If we give this function enough **decay** (this one decays quadratically while the previous one decays linearly) and set

$$f(x) := \begin{cases} x^2\sin(1/x) & x \neq 0 \\ 0 & \text{otherwise,} \end{cases}$$

then we obtain a differentiable function! That f is differentiable on $\mathbb{R}-\{0\}$ is clear; also,

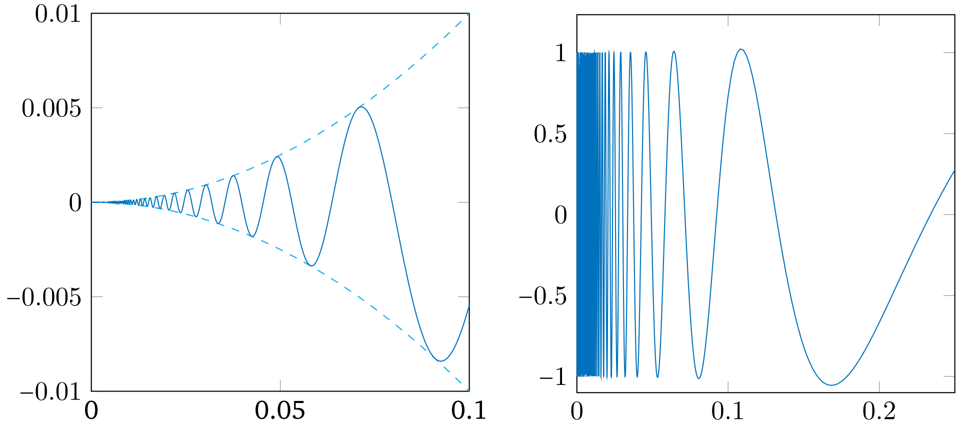
$$\frac{f(y)-f(0)}{y-0} = y\sin(1/y)$$

which converges to 0 as $y \to 0$. Therefore we simply set $f'(0) := 0$.

However, the derivative of this function is not continuous, even though it doesn't have a jump discontinuity. (This connects with my added remark on the IVT.) For nonzero x, the derivative is

$$f'(x) = 2x\sin(1/x) - \cos(1/x).$$

Yet another example showing how sometimes our intuition lead us astray. See graphs below. (f on the left; f' on the right.) [2]



Future reference: Darboux Property

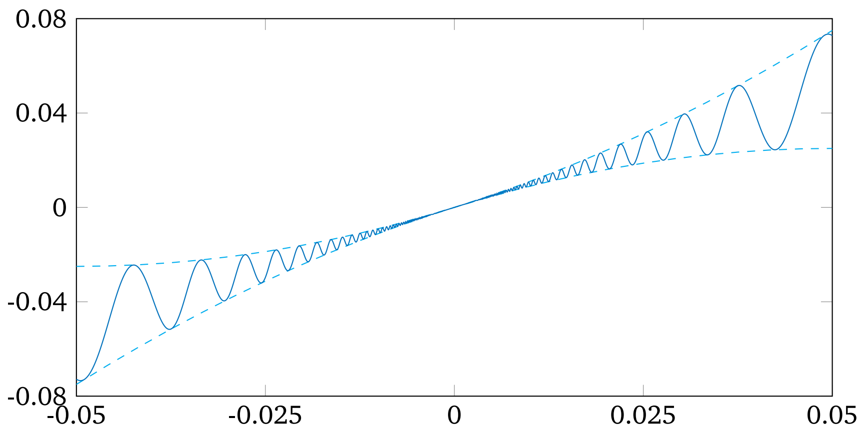
[2]The italicized remark is borrowed from Pugh's *Real Mathematical Analysis*, p.156.

5.2 The Mean Value Theorems

Lemma 9.13

If a real-valued function f is differenetiable at $x \in [a,b]$ and if $f'(x) > 0$, then there exists a $\delta > 0$ such that $f(y) < f(x) < f(z)$ for all $y, z \in [a,b] \cap (x-\delta, x+\delta)$ with $y < x < z$.
Future reference: Darboux Property

Put informally, if $f'(x) > 0$ then locally all points on the right side of x (i.e., $> x$) have greater function values, whereas locally all points on the left side of x have smaller function values.
This does not imply f is locally increasing. A counterexample[3]*is by defining $f(x) := x + 2x^2 \sin(1/x)$ at $x \neq 0$ and $f(x) := x$ at $x = 0$. This function is always positive for $x > 0$ [this is actually true, but at least it is bounded below by $x - 2x^2$ which is positive on $(0, 1/\sqrt{2})$] and always negative for $x < 0$, yet we are well aware how it oscillates like crazy near 0. The coefficient 2 guarantees that f is never locally increasing on any (nondegenerate) interval containing 0. Below is the graph of a more dramatic version, using $f(x) := x + 10x^2 \sin(1/x)$:*



Proof. Since the derivative $f'(x) > 0$, we can fix $\epsilon \in (0, f'(x)/2)$. The existence of $f'(x)$ implies that

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x),$$

so in particular there exists $\delta > 0$ such that

$$\left| \frac{f(x+h) - f(x)}{h} - f'(x) \right| < \epsilon \qquad \text{for} \qquad |h| < \delta.$$

In other words,

$$hf'(x) - |h|\epsilon < f(x+h) - f(x) < hf'(x) + |h|\epsilon.$$

If $h > 0$ then $f(x+h) - f(x) \geqslant hf'(x) - h\epsilon = h(f'(x) - \epsilon) > 0$, so $f(z) > f(x)$ for all $z \in (x, x+\delta)$.
Likewise, if $h < 0$ then $f(x+h) - f(x) < hf'(x) + (-h)\epsilon = h(f'(x) - \epsilon) < 0$, so $f(z) < f(x)$ for any $z \in (x-\delta, x)$. ☐

Theorem 9.14: "Local Optimum Theorem"

If f attains a local maximum (resp. minimum) at $x \in (a,b)$ and if f is differentiable at x,then $f'(x) = 0$.

[3] This is PS9 problem 2.

Proof. This is immediate by the previous lemma. For example, if $f'(x) > 0$ then the value of f on right-side neighborhood $(x, x+\delta)$ is larger than $f(x)$, contradicting f's local maximality. □

Theorem 9.15: Rolle's Theorem

If $f : [a,b] \to \mathbb{R}$ is continuous, if f is differentiable on (a,b)[4], and if $f(a) = f(b)$, then there exists $x \in (a,b)$ such that $f'(x) = 0$.
Future reference: Taylor's Theorem

Proof. Let $x_1, x_2 \in [a,b]$ be points where f attains maximum and minimum on $[a,b]$, respectively (recall that a continuous real-valued function on a compact set attains maximum and minimum by Theorem 8.5).

(1) If x_1, x_2 are both at endpoints, i.e., if $x_1, x_2 \in \{a,b\}$, then f must be a constant function (*the maximum and the minimum agree, so $f(x_2) = f(x) = f(x_1)$ for every x*). The claim holds in this case.

(2) If (at least) one of x_1, x_2 lives inside (a,b) (i.e., not an endpoint), then the previous theorem asserts that $f' = 0$ at that point. This completes the proof. □

Theorem 9.16: Generalized MVT / Cauchy's MVT

If $f, g \in C([a,b]) \cap D((a,b))$, i.e., f, g are continuous on $[a,b]$ and differentiable on (a,b), then there exists $x \in (a,b)$ such that

$$(f(b) - f(a))\, g'(x) = (g(b) - g(a))\, f'(x).$$

Future reference: L'Hôpital's Rule

Proof. *Given what we have just learned, we want to relate the equation with* 0. We set

$$h(t) := (f(b) - f(a))g(t) - (g(b) - g(a))\, f(t).$$

Notice that here $f(b) - f(a), g(b) - g(a)$ serve as constants. It follows that

$$h(a) = f(b)g(a) - f(a)g(a) - g(b)f(a) + g(a)f(a) = h(b),$$

and it is clear that h is a continuous function. Therefore, using Rolle's Theorem, there exsts $x \in (a,b)$ with $h'(x) = 0$. It remains to notice that $h'(x)$ has exactly the form we are looking for. □

Corollary 9.17: Mean Value Theorem (MVT) / Lagrange Theorem

If $f \in C([a,b]) \cap D((a,b))$, then there exists $x \in (a,b)$ such that

$$\frac{f(b) - f(a)}{b - a} = f'(x).$$

Future reference: FTC part 2

Proof. Take $g := x$ and use the previous theorem. □

[4]We write the shorthand notation $f \in D(a,b)$ to say f is differentiable on (a,b) from now on.

Corollary 9.18: Derivative & Monotonicity

Suppose f is differentiable on (a, b). Then:

(1) $f' \geqslant 0$ if and only if f is nondecreasing,

(2) $f' \equiv 0$ if and only if f is conatant, and

(3) $f' \leqslant 0$ if and only if f is nonincreasing.

(*We write $f' \geqslant 0$ as a shorthand notation for $f'(x) \geqslant 0$ for all x; likewise for $f' \leqslant 0$.*)

Proof. The $\implies$ direction for all three claims follow from MVT: for example, if $f' \geqslant 0$ then for $x_1 > x_1$,

$$f(x_2) - f(x_1) = (x_2 - x_1) f'(x) \text{ for some } x \in (x_1, x_2),$$

and since both $x_2 - x_1 > 0$ and $f'(x) \geqslant 0$ we see f is nondecreasing. The other two are analogous.
For $\impliedby$, we again use (1) for example. If f is nondecreasing then

$$\frac{f(y) - f(x)}{y - x} \geqslant 0$$

for all $y > x$. Since the limit exists by the assumption that f is differentiable, the limit, i.e., $f'(x)$, must also be nonnegative. This proves the claim. □

Theorem 9.19: IVT for Derivatives / Darboux Property

Suppose $f \in D([a, b])$, i.e., f is differentiable on $[a, b]$, and suppose $\lambda \in (f'(a), f'(b))$, then there exists $x \in (a, b)$ such that $f'(x) = \lambda$.
Notice that this is highly analogous to the IVT for continuous functions. However, as we have shown in the pathological examples, the derivative of a differentiable function may fail to be continuous. This theorem guarantees that the IVT still holds in those situations.

Proof. We put $g(t) := f(t) - \lambda t$. Since $\lambda \in (f'(a), f'(b))$, evaluating the derivative of g at endpoints gives

$$\begin{cases} g'(a) = f'(a) - \lambda < 0 \\ g'(b) = f'(b) - \lambda > 0 \end{cases} \implies \exists_{\delta > 0} \text{ such that } \begin{cases} g(t) < g(a) \text{ for } t \in (a, a + \delta) \\ g(t) < g(b) \text{ for } t \in (b - \delta, b). \end{cases}$$

(The $\Rightarrow$ is given by Lemma 9.13.) In addition, since g is continuous and $[a, b]$ compact, it attains its mimum at some $x \in (a, b)$ (the above equations already ruled out the possibility that the minimum is on endpoints). Thus we have $g'(x) = f'(x) - \lambda = 0$, i.e., $f'(x) = \lambda$, which completes our proof. □

5.3 L'Hôpital's Rule

Beginning of March 22, 2021

Theorem 9.20: L'Hôpital's Rule

Suppose that $f, g \in D((a,b))$, $-\infty \leqslant a < b \leqslant +\infty$ [*note that we allow a, b to be $\pm\infty$*] and that $g(x) \neq 0$ for all $x \in (a,b)$. Suppose the quotient of derivative

$$\frac{f'(x)}{g'(x)} \to A \quad \text{as} \quad x \to a.$$

If $f(x) \to 0$ and $g(x) \to 0$ as $x \to a$, or if $g(x) \to \infty$ as $x \to a$ [*and no assumption on f needs to be made*] then

$$\frac{f(x)}{g(x)} \to A \text{ as } x \to a, \text{ just like } \frac{f'(x)}{g'(x)}.$$

The claim also holds if $x \to b$ or if $g(x) \to -\infty$, and the proofs are analogous.

In particular, if $f(x)/g(x) \to [0/0]$ or $[\pm\infty/\pm\infty]$, the theorem gives a result that we all know. This theorem shows that it sufficiently to have only the denominator $\to \pm\infty$.

Future reference: Taylor's Theorem of Peano form

Proof. Note that since a is the left endpoint of (a,b), $x \to a$ means $x \to a_+$. Therefore we are only interested in analyzing the behavior of f/g as $x \to a$ from the right. The proof will follow from the following claims:

Claim (i): Suppose $-\infty \leqslant A < \infty$ and let q be any real number with $A < q$. Then there exists $c > a$ such that

$$\frac{f(x)}{g(x)} < q \quad \text{if} \quad x \in (a,c).$$

Claim (ii): Suppose $-\infty < A \leqslant \infty$ and let $\tilde{q}$ be any real number with $A > \overline{q}$. Then there exists $\tilde{c} > a$ such that

$$\frac{f(x)}{g(x)} > \tilde{q} \quad \text{if} \quad x \in (a,\tilde{c}).$$

Once we have these claims, if $A = \infty$ we use Claim (ii), if $A = -\infty$ we use Claim (i), and if $A \in \mathbb{R}$ we fix $\epsilon > 0$, use both claims, and pick the smaller c and obtain an interval, on which the quotient is always $< \epsilon$ away from A. □

Proof of Claim (i). We will only prove the first claim; the second is highly analogous. We begin by picking $r \in (A,q)$. Since $f'(x)/g'(x) \to A$, there exists $d \in (a,b)$ such that $f'(x)/g'(x) < r$ for all $x \in (a,d)$. (*The quotient needs to approach A so on some interval it needs to be $< r$.*)

If we pick any two $x, y \in (a,d)$ with $y < x$, using Generalized MVT on (y,x), there exists $t \in (y,x) \subset (a,d)$ with

$$\frac{f(x)-f(y)}{g(x)-g(y)} = \frac{f(y)-f(x)}{g(y)-g(x)} = \frac{f'(t)}{g'(t)} < r \tag{Δ}$$

[*where the $<$ is because $t \in (a,d)$*].

Case (a): Suppose $f(y), g(y) \to 0$ as $y \to a$. Since the only restriction imposed on y is that $a < y < x$, we are allowed to *push* y to a as close as we want. Letting $y \to a$, we get

$$\lim_{y\to a} \frac{f(y)-f(x)}{g(y)-g(x)} = \frac{f(x)}{g(x)} \leqslant r \tag{1}$$

where $<$ becomes $\leqslant$ when taking limits. Note that the only restriction imposed on x is that $a < y < x < d$, and as $y \to a$ we simply need $a < x < d$. In other words (1) holds for all $x \in (a,d)$.

Case (b): Suppose that $g(y) \to \infty$ as $y \to a$. *The trick is to multiply* (Δ) *by* $(g(y) - g(x))/g(y)$ *and obtain*

$$\frac{f(y) - f(x)}{g(y)} < \frac{r(g(y) - g(x))}{g(y)} \implies \frac{f(y)}{g(y)} < r - r\frac{g(x)}{g(y)} + \frac{f(x)}{g(y)} \quad \text{for all } x, y \in (a, d).$$

Once again, once x is picked, we can simply fix (and ignore) it and *push* y towards a. In particular, since $-rg(x) + f(x)$ has a fixed value,

$$-r\frac{g(x)}{g(y)} + \frac{f(x)}{g(y)}$$

can be made arbitrarily close since $g(y) \to \infty$ as $y \to a$. In particular we can find $c > a$ sufficiently close to a such that

$$\frac{f(y)}{g(y)} < r - r\frac{g(x)}{g(y)} + \frac{f(x)}{g(y)} < q \quad \text{for all } y \in (a, c). \tag{2}$$

The trick here helped us to obtain the quotient we want and also allowed us to bound it by q, both of which weren't possible if we simply tried to use the approach in Case (a).

(1) and (2) together complete the proof of Claim (i). □

Example 10.1. $\lim_{x\to 0} \frac{\sin x}{x} \overset{H}{=} \lim_{x\to 0} \frac{\cos x}{1} = 1$, where $\overset{H}{=}$ denotes the use of L'Hôpital's rule.

Example 10.2. Let $f \in D((0, \infty))$ be such that $\lim_{x\to\infty} (af(x) + 2\sqrt{x}f'(x)) = L \in \mathbb{R}$. Find $\lim_{x\to\infty} f(x)$.

Solution. "Note that"

$$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} \frac{e^{a\sqrt{x}}f(x)}{e^{a\sqrt{x}}}.$$

We chose this exponential because of the special form of its derivative. Since the denominator $\to \infty$, we can invoke L'Hôpital's rule and obtain

$$\lim_{x\to\infty} f(x) \overset{H}{=} \lim_{x\to\infty} \frac{e^{a\sqrt{x}}(f'(x) + af(x)/2\sqrt{x})}{ae^{a\sqrt{x}}/(2\sqrt{x})} = \lim_{x\to\infty} \frac{1}{a}(af(x) + 2\sqrt{x}f'(x)) = \frac{L}{a}.$$

5.4 Taylor's Theorem

Definition 10.3: Higher Order Derivatives, D^n, & C^n

For $f : [a,b] \to \mathbb{R}$ and $n \geqslant 2$, we define

$$f^{(n)} := (f^{(n-1)})' = \lim_{h\to 0} \frac{f^{(n-1)}(x+h) - f^{(n-1)}(x)}{h},$$

assuming that $f^{(n-1)}$ is differentiable. We define $f = f^{(0)}$ and $f' = f^{(1)}$.
If $f^{(n)}$ exists, we say f is $\boldsymbol{n}^{\text{th}}$ **order differentiable** and write $f \in D^n((a,b))$.
If $f^{(n)}$ is continuous, then $f^{(0)}, \ldots, f^{(n-1)}$, and we say f is $\boldsymbol{n}^{\text{th}}$ **order continuously differentiable** or f is C^n, written $f \in C^n([a,b])$.
Note that if $f \in D^n((a,b))$ then this automatically implies $C^{n-1}((a,b))$ since differentiability $\Rightarrow$ continuity. However, this does not necessarily imply $f \in C^{n-1}([a,b])$. The behaviors at endpoints may be wild!

Theorem 10.4: Taylor's Theorem

Suppose $f \in C^{n-1}([a,b]) \cap D^n((a,b))$, i.e., $f^{(n-1)}$ is continuous and $f^{(n)}$ exists for every $t \in (a,b)$ and suppose $n \in \mathbb{N}$. Then, for all $x, x_0 \in [a,b]$ there exists a ξ between x, x_0 such that

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \frac{f^{(n)}(\xi)}{n!}(x - x_0)^n.$$

*The first summation is called the **Taylor's expansion** (or **Taylor series**) of order $n-1$ at x_0. For convenience, we define this sum to be $P_{x_0}(x)$. If $x_0 = 0$, the expansion is often called the **Maclaurin expansion**. The second term is called the remainder in the **Lagrange form**.*
Future reference: Taylor's Theorem of Peano form, Taylor's Theorem (summary)

Note that if we take $n = 1$ then this is simply the MVT: $f(x) = f(x_0) + f'(\xi)(x - x_0)$ for some ξ between x, x_0.
Soon we will show what is special about $P_{x_0}(x)$: it is, in some sense, the best approximation of f by a polynomial of degree $\leq n-1$ subject to the condition that both functions agree at x_0. It resembles f "closely enough" by having the same derivatives at x_0 up to order $n-1$ and it is also the only polynomial of degree $\leqslant n-1$ that makes the remainder "small enough". We will discuss this more in-depth very soon.

Proof. If $x = x_0$ the claim is trivial.
If $x \neq x_0$, we define $M := (f(x) - P_{x_0}(x))/(x - x_0)^n$ [*a number*]. We want to show that there exists ξ between x and x_0 (i.e., in (x, x_0) if $(x < x_0)$ and in (x_0, x) if $x > x_0$) such that $M = f^{(n)}(\xi)/n!$.
We define

$$g(y) := f(y) - P_{x_0}(y) - M(y - x_0)^n$$

for $y \in (a,b)$. Since $P_{x_0}(y)$ does *not* have n^{th} order derivative (its derivatives only exist up to order $n-1$) and M is merely a constant if we take derivative of $M(y - x_0)^n$ with respect to y, we obtain

$$g^{(n)}(y) = f^{(n)}(y) - n!M. \tag{Δ}$$

On the other hand, if we evaluate g and its derivatives at x_0, it is clear that the "0$^{\text{th}}$ order derivative"

$$g(x_0) = f(x_0) - \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x_0 - x_0)^{(k)} - M(x_0 - x_0)^n = f(x_0) - f(x_0) = 0. \tag{1}$$

In addition, for $m \leqslant n-1$,

$$g^{(m)}(y) = f^{(m)}(y) - \sum_{k=m}^{n-1} \frac{f^{(k)}(x_0)}{(k-m)!}(y-x_0)^{k-m} - Mn(n-1)\dots(n-m+1)(y-x_0)^{n-m},$$

so if $y = x_0$, the only term that gets subtracted is $f^{(k)}(x_0)(y-x_0)^{k-m}/(k-m)!$ where $k = m$, and this evaluates to $f^{(m)}(x_0)0^0/0! = f^{(m)}(x_0)$. Along with (1), we see

$$g^{(m)}(x_0) = 0 \quad \text{for} \quad m = 0, 1, \dots, n-1. \tag{2}$$

To conclude the proof, also notice that

$$g(x) = f(x) - P_{x_0}(x) - M(y-x_0)^n = f(y) - P_{x_0}(y) - (f(x) - P_{x_0}(x)) = 0$$

directly by construction of M. Therefore, by Rolle's Theorem,

$$\begin{aligned}
[\text{Rolle} : g(x) = g(x_0) = 0] &\implies \text{there exists } x_1 \text{ between } x, x_0 \text{ such that } g'(x_1) = 0 \\
[\text{Rolle: } g'(x_0) = g'(x_1) = 0] &\implies \text{there exists } x_2 \text{ between } x_1, x_0 \text{ such that } g^{(2)}(x_2) = 0 \\
&\implies \dots \\
[\text{Rolle: } g^{(n-1)}(x_0) = g^{(n-1)}(x_{n-1}) = 0] &\implies \text{there exists } x_n \text{ between } x_{n-1}, x_0 \text{ such that } g^{(n)}(x_n) = 0.
\end{aligned}$$

(Δ) suggests that $g^{(n)}(x_n) = 0$, so of course $f^{(n)}(x_n) = n!M$, i.e., $M = f^{(n)}(x_n)/n!$, and we are done. □

Beginning of March 24, 2021

Example 10.5.

(1) The Taylor series of order $n-1$ at 0 of the exponential function e^x is given by

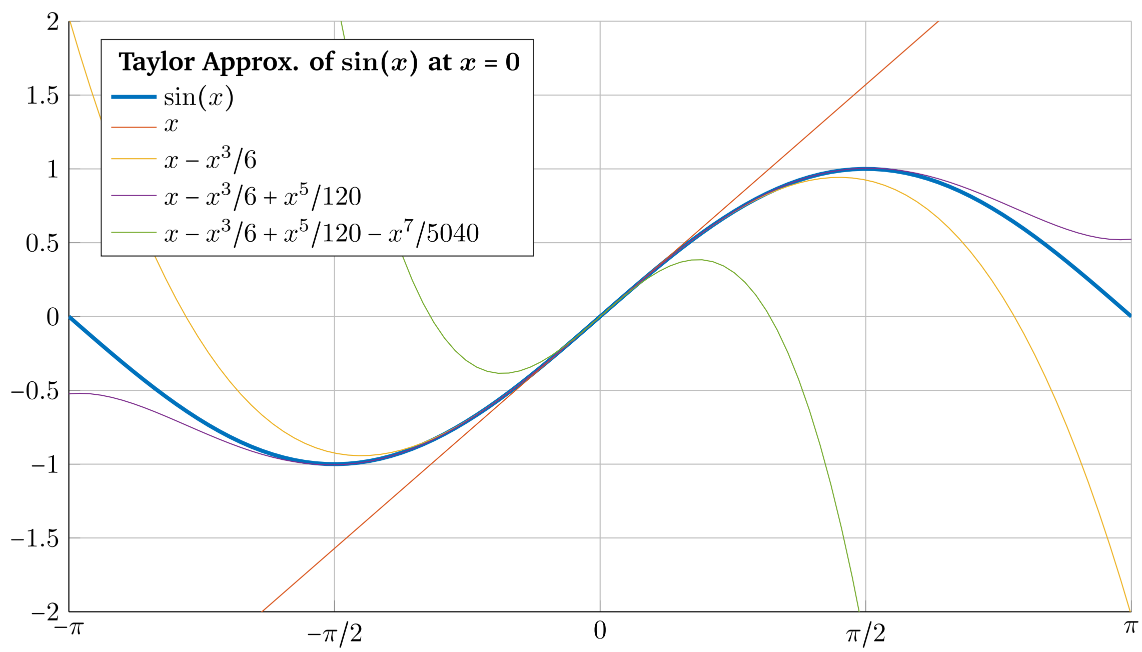
$$\begin{aligned}
e^x &= f(0) + f'(0)(x-0) + \frac{f''(0)}{2}(x-0)^2 + \dots + \frac{f^{(n-1)}(0)}{(n-1)!}(x-0)^{n-1} + \frac{e^\xi}{n!}x^n \\
&= 1 + x + \frac{x^2}{2} + \dots + \frac{x^{n-1}}{(n-1)!} + \frac{e^\xi}{n!}x^n \qquad \text{for some } \xi.
\end{aligned}$$

Note that this is also the Maclaurin series for e^x.

(2) The Taylor series / Maclaurin series for $\sin(x)$ of order $n-1$ is given by

$$\sin(x) = \sum_{k=0}^{n-1} \frac{(-1)^k}{(2k+1)!}x^{2k+1} + \frac{(\sin(x))^{(n)}\big|_{x=\xi}}{(2n+1)!}x^{2n+1} \qquad \text{for some } \xi.$$

Note that for this one the even order derivatives vanish because $\sin(0) = 0$. Below is the graph of $\sin(x)$ and some of its Taylor approximations on $[-\pi, \pi]$. We see that the approximation improves as the order increases.



(3) The $(n-1)^{\text{th}}$-order Taylor expansion of $1/x$ at 1 is given by

$$\frac{1}{x} = \sum_{k=0}^{n-1}(-1)^k(x-1)^k + \frac{(-1)^n}{\xi^{1+n}}(x-1)^n \qquad \text{for some } \xi.$$

Question. Can we write *infinite* series, e.g., write $\sin(x)$ as $\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)!}x^{2k+1}$?

Answer. Sometimes. Wait for the section on infinite Taylor series. *Infinity can be tricky and sometimes counterintuitive examples arise. We will need to discuss more on convergence before being able to characterize this question.*

While we have previously used the Lagrange form for the remainder in Taylor's Theorem, we can also write Taylor expansion with another form of the remainder:

Definition 10.6: Big $\mathcal{O}$ and Little o Notations

Let h be a (real-valued) function. We define

(1) $\mathcal{O}(h) :=$ any function of h such that there exists $c > 0$ such that $|\mathcal{O}(h)| < c|h|$ for sufficiently small h.

(2) $o(h) :=$ any function of h such that $o(h) = f(h)h$ for some f such that $f(h) \to 0$ as $h \to 0$.

In particular, if $f(h) = o(h)$ then $f(h) = \mathcal{O}(h)$, meaning that little o is stronger than big $\mathcal{O}$.

To make things more concrete, $f \in \mathcal{O}(h)$ if $\limsup_{x\to\infty} f(x)/h(x) < \infty$, and $f \in h(h)$ if $\limsup_{x\to\infty} f(x)/h(x) = 0$. (Usually this is just fine if we take limit rather than limit superior, but some special functions, e.g., $f(x) := x \pmod 5$, the notion of limit becomes meaningless.)

In the former case, f "grow in at most the same rate" as h: the asymptotic growth of f is no faster than h's [it could be slower though]; in the latter, f is "eventually negligible when compared to h": the asymptotic growth of f is strictly slower than h's. Some examples:

(1) $x^2 \in \mathcal{O}(x^2)$, $x^2 \in \mathcal{O}(100x^2)$, *and* $x^2 \in (3x^2+2x+1)$, *but none of these hold if we replace* $\mathcal{O}$ *by* o.

(2) $x^2 \in o(x^3)$, $x^2 \in o(x!)$, *and* $x^2 \in (2^x)$, *and all holds if we place* o *by* $\mathcal{O}$ *since* $o(x^2) \subset \mathcal{O}(x^2)$.

And... below is a fuller list of related notations:

Notation	Limit Definition	*Asymptotic* Intuition	Example(s)
$f \in o(h)$	$\limsup_{x\to\infty} \lvert f(x)/h(x)\rvert = 0$	f grows strictly slower than h	$x \in o(x^2)$
$f \in \mathcal{O}(h)$	$\limsup_{x\to\infty} \lvert f(x)/h(x)\rvert < \infty$	f grows at most as fast as h	$x \in \mathcal{O}(x^2)$, $x^2 \in \mathcal{O}(x^2)$
$f \in \Theta(h)$	$\limsup_{x\to\infty} \lvert f(x)/h(x)\rvert \in \mathbb{R}^+$	f grows at the same rate as h	$x^2 \in \Theta(x^2)$
$f \in \Omega(h)$	$\limsup_{x\to\infty} \lvert f(x)/h(x)\rvert > 0$	f grows at least as fast as h	$x^2 \in \Omega(x^2), x^3 \in \Omega(x^2)$
$f \in \omega(h)$	$\limsup_{x\to\infty} \lvert f(x)/h(x)\rvert = \infty$	f grows strictly faster than h	$x^3 \in \omega(x^2)$

Theorem 10.7: Taylor's Theorem with Remainder of Peano Form

If $f^{(n)}(x_0)$ exists then

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + o\left((x-x_0)^{n-1}\right).$$

This is stronger than Lagrange form Taylor's Theorem as the only thing it requires is n^{th} *order differentiability atone point rather than on an entire interval. Also notice that the Lagrange form remainder*

$$\frac{f^{(n)}(\xi)}{n!}(x-x_0)^n \quad \text{satisfies} \quad \lim_{x\to x_0} \frac{f^{(n)}(\xi)(x-x_0)^n/n!}{(x-x_0)^{n-1}} = \lim_{x\to x_0} \frac{f^{(n)}(\xi)(x-x_0)}{n!} = 0,$$

so this theorem implies the previous one.

Taylor's Theorem (summary)

Proof. Like before, we use $P_{x_0}(x)$ to denote the $(n-1)^{\text{th}}$-order Taylor series of f at x_0. We want to show that

$$\lim_{x\to x_0} \frac{f(x)-P_{x_0}(x)}{(x-x_0)^{n-1}} = 0.$$

First notice that as $x \to x_0$, since f is continuous at x_0 (because ${}^{(n)}(x_0)$ exists), $f(x) \to f(x_0)$, and so does $P_{x_0}(x)$, as $(x-x_0)^k$ vanishes for all k but $k=0$, in which case $f^{(0)}(x_0)$ is simply $f(x_0)$. Therefore the numerator $\to 0$ and clearly the denominator $\to 0$ too. We get an indeterminate form $0/0$, so we invoke L'Hôpital's Rule.

$$\lim_{x\to x_0} \frac{f(x)-P_{x_0}(x)}{(x-x_0)^{n-1}} \overset{H}{=} \lim_{x\to x_0} \frac{f'(x) - \sum_{k=1}^{n-1}\left[f^{(k)}(x_0)/(k-1)!\right](x-x_0)^{k-1}}{(n-1)(x-x_0)^{n-2}}.$$

Once again, as $x \to x_0$, since f' is continuous at x_0, $f'(x) \to f'(x_0)$. For the sum, all terms but $f^{(1)}(x_0)(x-x_0)^{1-1} = f'(x_0)$ survives, so the numerator is $f'(x_0) - f'(x_0) = 0$. Clearly the denominator $\to 0$, so we invoke L'Hôpital's Rule again (all the way until we cannot):

$$\lim_{x\to x_0} \frac{f(x)-P_{x_0}(x)}{(x-x_0)^{n-1}} \overset{H}{=} \dots \overset{H}{=} \lim_{x\to x_0} \frac{f^{(n-2)}(x) - \overbrace{\left[f^{(n-2)}(x_0) + f^{(n-1)}(x_0)(x-x_0)\right]}^{(n+2)^{\text{th}} \text{ derivative of } P_{x_0}(x)}}{(n-1)(n-2)\dots\cdot 3\cdot 2\cdot (x-x_0)}$$

$$\overset{H}{=} \lim_{x\to x_0} \frac{f^{(n-1)}(x) - f^{(n-1)}(x_0)}{(n-1)!} \tag{1}$$

Recall the assumption states that $f^{(n)}(x_0)$ exists. That is,

$$\lim_{x \to x_0} \frac{f^{(n-1)}(x) - f^{(n-1)}(x_0)}{x - x_0} = f^{(n)}(x_0)$$

exists! Therefore, if we multiply and divide (1) by $(x - x_0)$ we obtain

$$\ldots = \lim_{x \to x_0} \frac{f^{(n-1)}(x) - f^{(n-1)}(x)}{(n-1)!(x - x_0)} \cdot (x - x_0) = \lim_{x \to x_0} \frac{f^{(n)}(x_0)}{(n-1)!}(x - x_0) = 0,$$

completing the proof. [*Notice that the existence of $f^{(n)}(x_0)$ is stronger than a sufficient condition; all we needed was that $f^{(n-1)}$ is continuous at x_0. If so, we are immediately done after arriving at (1).*] □

We have just shown rigorously that the remainder "is small enough" compared to our expansion, as we once hypothesized when first introducing the Lagrange form Taylor's Theorem. Now we prove the uniqueness of such approximations subject to the constraint of order.

Lemma 10.8: Uniqueness of Local Expansions

If

$$f(x) = a_0 + a_1(x - x_0) + \ldots + a_n(x - x_0)^n + o\left((x - x_0)^n\right)$$

and

$$f(x) = b_0 + b_1(x - x_0) + \ldots + b_n(x - x_0)^n + o\left((x - x_0)^n\right),$$

then $a_i = b_i$ for all $i \in [0, n]$.
In particular, if f admits a Taylor expansion (it may or may not!), then it is the Taylor expansion.

Remark. One thing to keep in mind is that (differentiable $\Leftrightarrow$ expansion exists) is a false statement. This lemma shows that $\Rightarrow$ holds, i.e., if f is differentiable then its expansion is the Taylor expansion. However, it is not true that if f admits an expansion of the above form then it is the Taylor expansion. For example,

$$\lim_{x \to 0} \frac{x^3 \cos(1/x)}{x^2} = \lim_{x \to 0} x \cos(1/x) = 0$$

so $f(x) := x^3 \cos(1/x) \in o(x^2)$, and it admits a second-order expansion

$$f(x) = 0 + 0(x - x_0) + 0(x - x_0)^2 + o\left((x - x_0)^2\right)$$

at $x_0 = 0$. However we can also verify that $f''(x_0)$ does not exist, so $f''(x_0)/2$ makes no sense. This is PS10.3.

Proof. When we see two polynomials of similar forms that are equal, it is natural to consider subtraction:

$$0 = (a_0 - b_0) + (a_1 - b_1)(x - x_0) + \ldots + (a_n - b_n)(x - x_0)^n + o\left((x - x_0)^n\right).$$

(Notice that the difference between two $o\left((x - x_0)^n\right)$ functions are still $o\left((x - x_0)^n\right)$. This can be easily verified directly using the limit definition.) Letting $x \to x_0$, all but the first term $(a_0 - b_0)$ tends to 0. Therefore taking the limit gives $(a_0 - b_0) = 0$. Now we no longer have $a_0 - b_0$ so we can divide both sides by $x - x_0$, obtaining

$$0 = (a_1 - b_1) + (a_2 - b_2)(x - x_0) + \ldots + (a_n - b_n)(x - x_0)^{n-1} + o\left((x - x_0)^{n-1}\right).$$

(It can also be easily proven that if we divide an $o\left((x - x_0)^n\right)$ function by $(x - x_0)$ we obtain an $o\left((x - x_0)^{n-1}\right)$ function.) Letting $x \to x_0$ again we obtain $a_1 = b_1$. Doing this inductively, we obtain $a_i = b_i$ for all i, completing the proof. □

Chapter 6

The Riemann-Stieltjes Integral

Finally, Calculus I done rigorously! Prior to this point, we've always been told that a (Riemann) integral $\int_a^b f$ is the "area under the curve on $[a, b]$" (this is called the **undergraph**). But how do we justify (or modify) this rigorously? Also, we know continuous functions are integrable [or do we?], but what about other functions?

6.1 Riemann and Riemann-Stieltjes Integrals

Definition 10.9: Partition

A **partition** P of $[a, b]$ is a finite, nondecreasing sequence $(x_k)_{k=0}^n$ of points such that

$$a = x_0 \leqslant x_1 \leqslant \ldots \leqslant x_n = b.$$

We also set $\Delta x_i := x_i - x_{i-1}$.

Remark. This is a generalization of the "dividing $[a, b]$ into equal subintervals" which we encountered in first course in Calculus. Here we still divide $[a, b]$ into the disjoint union of many $[x_{i-1}, x_i]$'s, but we generalize it by allowing the intervals to have different lengths.

Definition 10.10: The Riemann Integral

Suppose $f : [a, b] \to \mathbb{R}$ is bounded. Given a partition P, we define

$$M_i := \sup_{[x_{i-1}, x_i]} f \qquad m_i := \inf_{[x_{i-1}, x_i]} f$$

and the upper and lower finite Riemann sums

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

We set the **upper** and **lower Riemann integrals** to be

$$\overline{\int_a^b} f \, \mathrm{d}x := \inf_P U(P, f) \qquad \text{and} \qquad \underline{\int_a^b} f \, \mathrm{d}x := \sup_P L(P, f),$$

where the supremum and infimum are taken over all $U(P, f)$ and all $L(P, f)$, where P is a partition, respectively. If

$$\overline{\int_a^b} f \, \mathrm{d}x = \underline{\int_a^b} f \, \mathrm{d}x$$

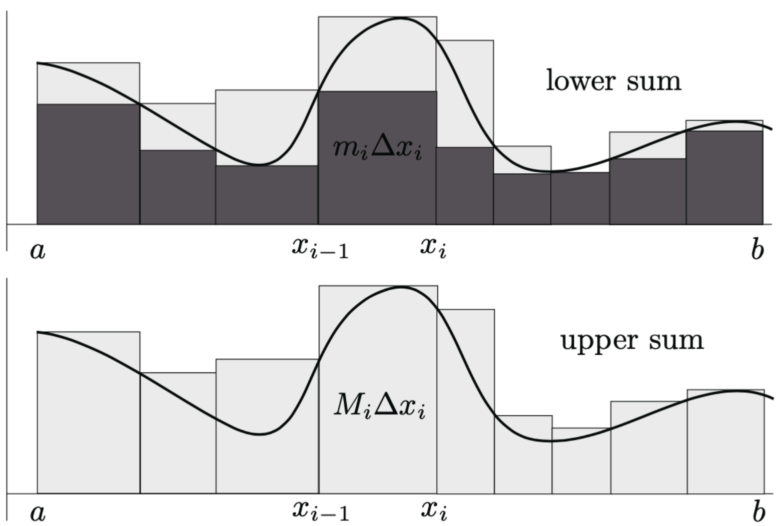
we define the **Riemann integral of f over $[a, b]$** to be

$$\int_a^b f \, \mathrm{d}x := \overline{\int_a^b} f \, \mathrm{d}x = \underline{\int_a^b} f \, \mathrm{d}x,$$

we say f is **Riemann integrable**, and we write $f \in \mathfrak{R}$ (or $f \in \mathfrak{R}([a, b])$).
Future reference: FTC part 1

This is a long and abstract definition. What it really says connects closely to our "informal" Riemann integration learned in "Calculus" earlier. The upper sums and lower sums are analogous to the ones we have learned, the only difference bowing that that the partition need not to be evenly spaced. We will soon see that as we refine the partition (i.e., making the intervals smaller by adding more points to P), the upper integral decreases and the lower integral increases; after all, they should be better approximations of what we think $\int_a^b f$ is. In Pugh's language, the length of the largest interval in P is called the ***mesh*** *of P, written* mesh P*, so an integrable function should satisfy*

$$\lim_{\text{mesh } P \to 0} \overline{\int_a^b} f \, \mathrm{d}x = \lim_{\text{mesh } P \to 0} \underline{\int_a^b} f \, \mathrm{d}x = \int_a^b f \, \mathrm{d}x.$$



From Pugh, *Real Mathematical Analysis*, p.167

Beginning of March 26, 2021

Definition 10.11: The Riemann-Stieltjes Integral

Let α be an increasing function on $[a, b]$. [*In particular, this implies α is bounded on $[a, b]$, as for all $x \in [a, b]$ we have $\alpha(a) \leqslant \alpha(x) \leqslant \alpha(b)$.*]
The construction of the Riemann-Stieltjes integral is highly similar to that of the Riemann integral. Suppose $f : [a, b] \to \mathbb{R}$ is bounded. Given a partition P, we define M_i and m_i just like above, and we define

$$\Delta\alpha_i := \alpha(x_i) - \alpha(x_{i-1}).$$

For the upper and lower finite Riemann sums, we replace Δx_i's by $\Delta\alpha_i$'s and obtain

$$U(P,f,\alpha) := \sum_{i=1}^{n} M_i \Delta\alpha_i \qquad \text{and} \qquad L(P,f,\alpha) := \sum_{i=1}^{n} m_i \Delta\alpha_i.$$

Once again, we set

$$\overline{\int_a^b} f \,\mathrm{d}\alpha := \inf_P U(P,f,\alpha) \qquad \text{and} \qquad \underline{\int_a^b} f \,\mathrm{d}\alpha := \sup_P L(P,f,\alpha).$$

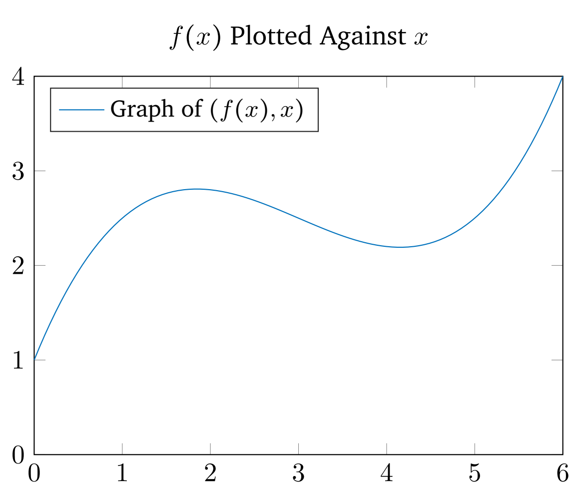
If they agree, we define the **Riemann-Stieltjes integral**[1]

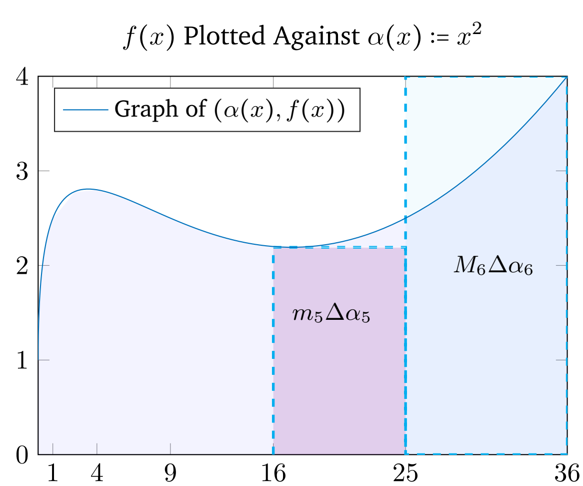
$$\int_a^b f \,\mathrm{d}\alpha := \overline{\int_a^b} f \,\mathrm{d}\alpha = \underline{\int_a^b} f \,\mathrm{d}\alpha, \qquad \text{(Eq.10.1)}$$

we say f is **Riemann-Stieltjes integrable** with respect to α, and we write $f \in \mathfrak{R}(\alpha)$ (i.e., it is Riemann integrable with respect to α).

Immediately, we see that the R-S integral is a generalization of the Riemann integral, as the latter is a special case of the R-S integral where $\alpha(x) = x$.

There are multiple geometric approaches to interpret the R-S integral, for example this article or this post (which also cites the same article). Equivalently, from a different (but essentially the same) perspective, we can also view the R-S integral as "area under the curve", where the curve is given by $(\alpha(x), f(x))$ instead of $(x, f(x))$. Below is an example of a polynomial f and $\alpha(x) := x^2$: heuristically, if we replace the x-axis by the "$\alpha(x)$-axis", the R-S integral reduces to the Riemann integral.





There are many applications of the R-S integral, some of which we have already encountered: the line integrals of form $\mathbf{F}\cdot\mathrm{d}\mathbf{r}, \mathbf{F}\cdot\mathrm{d}\mathbf{S}$ *from Calculus III are both examples.*

Question. For what kind of functions does (Eq.10.1) hold?

We will first examine the behaviors of $U(p,f,\alpha)$ and $L(P,f,\alpha)$.

[1] We will use "R-S integral" as an abbreviation of "Riemann-Stieltjes integral" from now on.

Definition 10.12: Refinement

A partition P^* is called a **refinement** of P if $P \subset P^*$. (*Some intervals produced by P get further subdivided into smaller intervals, hence the word "refinement."*)
We say P^* is a **common refinement** of P_1 and P_2 if $P^* = P_1 \cup P_2$.
(*Of course, the elements in P^* are ordered and increasing by the definition of a partition.*)

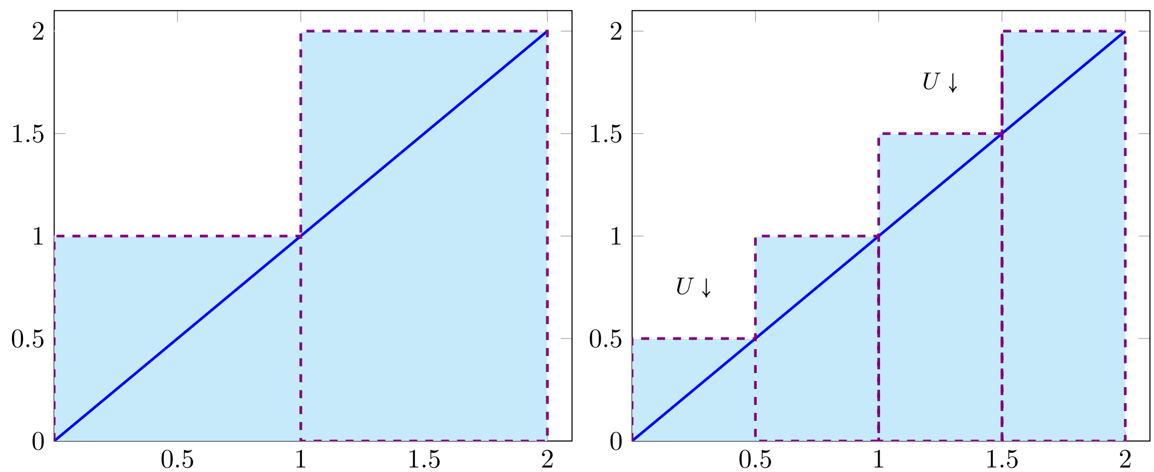
Theorem 10.13

If P^* is a refinement of P, then

$$L(P, f, \alpha) \leqslant L(P^*, f, \alpha) \qquad \text{and} \qquad U(P, f, \alpha) \geqslant U(P^*, f, \alpha).$$

Heuristically, this shows that refinements make Riemann sums closer to the actual integral, should it exist.
Future reference: Characterization of Riemann-Stieltjes integrability, Corollary 10.16

Main idea of the proof, informally illustrated by one diagram:



Proof. We will show $L(P, f, \alpha) \leqslant L(P^*, f, \alpha)$ only.

(1) Suppose that the refinement P^* only differs from P by one point, i.e., $P^* - P = \{x_*\}$ for some $x_* \in (a, b)$. We write $P := \{x_1, \dots, x_n\}$, and we pick $k \in \{1, \dots, n\}$ such that $x_{k-1} < x_* < x_k$ (*so the extra x_* is in $[x_{k-1}, x_k]$*). Then, by definition,

$$\begin{aligned}
L(P, f, \alpha) &= \sum_{i \neq k} m_i \Delta\alpha_i + m_k \Delta\alpha_k \\
&= \sum_{i \neq k} m_i \Delta\alpha_i + \inf_{[x_{k-1}, x_k]} f \cdot (\alpha(x_k) - \alpha(x_{k-1})) \\
&= \sum_{i \neq k} m_i \Delta\alpha_i + \inf_{[x_{k-1}, x_k]} f \cdot [(\alpha(x_k) - \alpha(x_*)) + (\alpha(x_*) - \alpha(x_{k-1}))] \\
&\leqslant \sum_{i \neq k} m_i \Delta\alpha_i + \inf_{[x_{k-1}, x_*]} f \cdot (\alpha(x_k) - \alpha(x_*)) + \inf_{[x_*, x_k]} f \cdot (\alpha(x_*) - \alpha(x_{k-1})) \\
&= L(P^*, f, \alpha).
\end{aligned}$$

This is basically the same idea as shown in the diagram; we focus on the interval in which a new point is added by refinement; if we divide this interval into two, then the two new infima we obtain cannot be smaller than the original infimum, and so the new lower Riemann sum we obtain cannot be smaller.

(2) If P^* and P differ more than one points, we simply need to apply case (1) multiple times, removing one point in $P^* - P$ each time, and eventually obtain our desired result. *Recall that P^* is itself a partition, so it can only have finitely many points. Therefore the difference is finite, meaning we only need to apply (1) finitely many times. This is valid.* □

Theorem 10.14: Lower Integral ⩽ Upper Integral

$$\underline{\int_a^b} f \, \mathrm{d}\alpha \leqslant \overline{\int_a^b} f \, \mathrm{d}\alpha.$$

Future reference: Characterization of Riemann-Stieltjes integrability

Proof. *Heuristically, the lower integrals are taken over infima whereas the upper integrals are taken over suprema.* We let P_1 and P_2 be any partitions of $[a, b]$ and define $P^* := P_1 \cup P_2$. Then,

$$L(P_1, f, \alpha) \leqslant L(P^*, f, \alpha) \leqslant U(P^*, f, \alpha) \leqslant U(P_2, f, \alpha),$$

where the first and third ⩽ are given by the previous theorem on refinement and the second because $m_i \leqslant M_i$ for each i corresponding to an element in the partition P^*. Thus

$$L(P_1, f, \alpha) \leqslant U(P_2, f, \alpha). \tag{1}$$

Since P_1 is arbitrarily chosen with respect to P_2 (i.e., given any P_2, $U(P_2, f, \alpha)$ is an upper bound for all the lower sums, so it is in particular no smaller than the least upper bound), we are allowed to let the argument on the LHS vary and take the supremum of all such lower sums and obtain

$$\sup_{P_1} L(P_1, f, \alpha) \leqslant U(P_2, f, \alpha). \tag{2}$$

On the other hand, since P_2 is also arbitrarily chosen, (2) holds for any P_2, so taking the infimum over all partitions gives us

$$\underline{\int_a^b} f \, \mathrm{d}\alpha = \sup_{P_1} L(P_1, f, \alpha) \leqslant \inf_{P_2} U(P_2, f, \alpha) = \overline{\int_a^b} f \, \mathrm{d}\alpha,$$

as claimed. *The key in this proof is that P_1, P_2 are chosen arbitrarily.* □

Theorem 10.15: Criterion for R-S Integrability

A function f is R-S integrable, i.e., $\underline{\int_a^b} f \, \mathrm{d}\alpha = \overline{\int_a^b} f \, \mathrm{d}\alpha$, if and only if

$$\text{for all } \epsilon > 0, \text{ there exists a partition } P \text{ such that } U(P, f, \alpha) - L(P, f, \alpha) < \epsilon. \qquad \text{(Eq.10.2[2])}$$

Future reference: Corollary 10.16, continuous functions are R-S integrablem Theorem 11.1, Theorem 11.3, FTC part 2

Proof. We first show $\Longrightarrow$. Let $\epsilon > 0$ be given. By the definitions of supremum and infimum, there exists a partition P_1 such that

$$U(P_1, f, \alpha) \leqslant \overline{\int_a^b} f \, \mathrm{d}\alpha + \frac{\epsilon}{2}, \tag{1}$$

[2]Not an equation though, but whatever...

since the upper integral is the infimum of these upper sums. Similarly there exists P_2 with

$$L(P_2, f, \alpha) \geqslant \underline{\int_a^b} f \, d\alpha - \frac{\epsilon}{2}. \tag{2}$$

We define a refinement $P^* := P_1 \cup P_2$. It follows that

$$U(P^*, f, \alpha) - \frac{\epsilon}{2} \leqslant U(P_1, f, \alpha) - \frac{\epsilon}{2} \leqslant \int_a^b f \, d\alpha \leqslant L(P_2, f, \alpha) + \frac{\epsilon}{2} \leqslant L(P^*, f, \alpha) + \frac{\epsilon}{2},$$

where the first and last $\leqslant$ are by Theorem 10.13 and the middle two by (1), (2), and the assumption that $\int_a^b = \underline{\int_a^b} = \overline{\int_a^b}$. Hence $U(P^*, f, \alpha)$ and $L(P, f, \alpha)$ are $\leqslant \epsilon$ apart (and since ϵ is arbitrary of course we can make the difference $< \epsilon$).

For the converse, we also let $\epsilon > 0$ (*and we want to show that* $\overline{\int_a^b} - \underline{\int_a^b} < \epsilon$ *for any* ϵ.) This is immediate, as now we have the assumption that there exists P satisfying (the first $\leqslant$ is by Theorem 10.14, independent of ϵ)

$$0 \leqslant \overline{\int_a^b} f \, d\alpha - \underline{\int_a^b} f \, d\alpha = \inf_{\tilde{P}} U(\tilde{P}, f, \alpha) - \sup_{\tilde{P}} L(\tilde{P}, f, \alpha)^3 \leqslant \underbrace{U(P, f, \alpha)}_{\leqslant \sup L} - \underbrace{L(P, f, \alpha)}_{\geqslant \inf L} < \epsilon. \qquad \square$$

Corollary 10.16

Let $\epsilon > 0$. Suppose Equation 10.2 holds for some partition $P = \{x_0, \dots, x_n\}$. Then,

(1) Equation 10.2 also holds for any refinement P^* of P. (*Immediate by Theorem 10.13 and Theorem 10.15. The upper sum decreases whereas the lower sum increases, so the difference is still* $< \epsilon$.)

(2) Let $\{s_0, \dots, s_n\}$ and $\{t_0, \dots, t_n\}$ be two sequences of points. If for all i we have $s_i, f_i \in [x_{i-1}, x_i]$ then

$$\sum_{i=1}^{n} |f(s_i) - f(t_i)| \Delta\alpha_i < \epsilon.$$

(*Also immediate since* $|f(s_i) - f(t_i)|$ *cannot exceed* $M_i - m_i$*, so the sum* $\leqslant U(P, f, \alpha) - L(P, f, \alpha) < \epsilon$.)

(3) If $f \in \mathfrak{R}(\alpha)$ and $\{t_0, \dots, t_n\}$ are such that $t_i \in [x_{i-1}, x_i]$, then

$$\left| \sum_{i=1}^{n} f(t_i) \Delta\alpha_i - \int_a^b f \, d\alpha \right| < \epsilon.$$

Proof. Since $m_i \leqslant f(t_i) \leqslant M_i$ *we have* $L(P, f, \alpha) \leqslant \sum_{i=1}^{n} f(t_i)\Delta\alpha_i \leqslant U(P, f, \alpha)$. *Also, since* $L \leqslant \int_a^b \leqslant U$ *by definitions of supremum and infimum, putting everything together, we have*

$$-\epsilon < L(P, f, \alpha) - U(P, f, \alpha) \leqslant \sum_{i=1}^{n} f(t_i)\Delta\alpha_i - \int_a^b f \, d\alpha \leqslant U(P, f, \alpha) - L(P, f, \alpha) < \epsilon. \qquad \square$$

Future reference: FTC part 2

Beginning of March 29, 2021

[3] I hope the notation $\tilde{P}$ here doesn't cause more confusion than all the mess already going on here. I chose to write $\tilde{P}$ rather than P for the mere purpose to show that here $\tilde{P}$ is like a "dummy variable" in integration when we take $\inf U$ and $\sup L$.

Theorem 10.17: Continuous Functions are R-S Integrable

If $f \in C([a,b])$ then $f \in \mathfrak{R}(\alpha)$ on $[a,b]$ (for all α).
Future reference: Future reference: Example 11.9

Proof. Let $\alpha > 0$ be given and fix $\epsilon > 0$. We can pick a sufficiently small $\eta > 0$ such that $\eta(\alpha(b) - \alpha(a)) < \epsilon$. Since f is continuous on a compact domain, Theorem 8.11 states that it is uniformly continuous. Therefore there exists a $\delta > 0$ such that

$$|f(x) - f(y)| < \eta \qquad \text{whenever} \qquad |x - y| < \delta, x, y \in [a,b].$$

Now we pick a partitioxn $P = \{x_0, \ldots, x_n\}$ such that $\Delta x_i < \delta$ for all i, i.e., mesh $P < \delta$. It follows that $M_i - m_i < \eta$ for all i. Also, since f is continuous and $[x_{i-1}, x_i]$ compact, M_i and m_i are attained. Therefore,

$$U(P,f,\alpha) - L(P,f,\alpha) = \sum_{i=1}^{n}(M_i - m_i)\Delta\alpha_i \leqslant \eta\sum_{i=1}^{n}\Delta\alpha_i = \eta(\alpha(b) - \alpha(a)) < \epsilon.$$

By Theorem 10.15 we are done, as $U(P,f,\alpha)$ and $L(P,f,\alpha)$ can be arbitrarily close. □

Theorem 11.1

If f is monotonic on $[a,b]$ and $\alpha \in C([a,b])$ (and nondecreasing by definition), then $f \in \mathfrak{R}(\alpha)$.

Proof. We will only prove the case where f is nondecreasing. Let $\epsilon > 0$ be given. For sufficiently large n (whose exact condition we will specify later), we can define a partition $P := \{x_0, \ldots, x_n\}$ of $[a,b]$ such that $\{\alpha(x_0), \ldots, \alpha(x_n)\}$ forms an arithmetic sequence, i.e.,

$$\Delta\alpha_i = \frac{\alpha(b) - \alpha(a)}{n}$$

for all i. (Such partition exists because α is continuous and we can use the IVT.) Then, since f is nondecreasing, the supremum of f on each interval is attained at the right endpoint whereas the infimum is attained at the left endpoint. Hence $M_i - m_i = f(x_i) - f(x_{i-1})$, and

$$\begin{aligned} U(P,f,\alpha) - L(P,f,\alpha) &= \sum_{i=1}^{n}\Delta\alpha_i\left[f(x_i) - f(x_{i-1})\right] \\ &= \frac{\alpha(b) - \alpha(a)}{n}\sum_{i=1}^{n}\left[f(x_i) - f(x_{i-1})\right] \\ &= \frac{(\alpha(b) - \alpha(a))(f(b) - f(a))}{n}. \end{aligned} \tag{1}$$

We see that n needs to be sufficiently large such that (1) $< \epsilon$ in the first place. Then the claim follows from Theorem 10.15.

As a side note: if α is strictly increasing, then the choice of partition is unique. □

Theorem 11.2

If f is bounded on $[a,b]$ and has only *finitely many* points of discontinuity $y_1, \ldots, y_k$ on $[a,b]$ and α is continuous at y_i, then $f \in \mathfrak{R}(\alpha)$. *Note that we do not require α to be continuous on all of $[a,b]$; it only needs to be continuous at points where f is not.*

We will omit the proof, but the main idea is that, even at a point of discontinuity, since f is bounded, the jump is finite. Then if we have sufficiently small $\Delta\alpha$, we can make the difference between $M_i\Delta\alpha_i$ and $m_i\Delta\alpha_i$ arbitrarily *small. Therefore U and L can be made arbitrarily small, proving the R-S integrability of f. For a full proof, see Rudin, Theorem 6.10.*

Theorem 11.3: Continuous Images of R-S Integrable Functions are R-S Integrable

Suppose that $f \in \mathfrak{R}(\alpha)$ on $[a,b]$ and $f(x) \in [m,M]$ for all $x \in [a,b]$ (i.e., R-S integrable with respect to α and bounded). If $\varphi \in C([m,n])$, then the composite $\varphi \circ f \in \mathfrak{R}(\alpha)$.
Future reference: Theorem 11.5

Proof. Let $\epsilon > 0$ be given. Since φ is uniformly continuous (continuous on compact domain), there exists a $\delta \in (0,\epsilon)$ [*this means that we can WLOG take δ to be smaller than ϵ; if for some $\delta \geqslant \epsilon$ the condition holds, then the same condition clearly holds for a smaller δ*] such that

$$|\varphi(x) - \varphi(y)| < \epsilon \qquad \text{whenever} \qquad |x-y| < \delta \text{ and } x,y \in [m,M]. \tag{1}$$

By Theorem 10.15, since $f \in \mathfrak{R}(\alpha)$, there exists a partition P such that

$$U(P,f,\alpha) - L(P,f,\alpha) < \delta^2 \tag{2}$$

Since we want to show the R-S integrability of $\varphi \circ f$ with respect to α, it is natural that we define

$$M_i^* := \sup_{[x_{i-1},x_i]} \varphi \circ f \qquad \text{and} \qquad m_i^* := \inf_{[x_{i-1},x_i]} \varphi \circ f$$

where the x_i's come from the partition P. This leads to

$$U(P,\varphi \circ f,\alpha) - L(P,\varphi \circ f,\alpha) = \sum_{i=1}^{n} (M_i^* - m_i^*)\Delta\alpha_i. \tag{3}$$

For each index i, we either have $M_i - m_i < \delta$ or $M_i - m_i \geqslant \delta$ (not starred, just M_i, m_i).[4] Thus (3) becomes

$$U(P,\varphi \circ f,\alpha) - L(P,\varphi \circ f,\alpha) = \sum_{M_i - m_i < \delta} (M_i^* - m_i^*)\Delta\alpha_i + \sum_{M_i - m_i \geqslant \delta} (M_i^* - m_i^*)\Delta\alpha_i.$$

(i) For every i in the first term, since $f(x), f(y) \in [m_i, M_i]$ for all $x, y \in [x_{i-1}, x_i]$, we have $|f(x) - f(y)| < \delta$. Using (1) on the above inequality we obtain

$$|\varphi(f(x)) - \varphi(f(y))| < \epsilon \qquad \text{for all } x,y \in [x_{i-1},x_i].$$

Therefore, taking supremum on LHS then infimum on RHS give

$$\begin{aligned} \varphi(f(x)) < \varphi(f(y)) + \epsilon &\implies \sup_{x \in [x_{i-1},x_i]} \varphi(f(x)) \leqslant \varphi(f(y)) + \epsilon \\ &\implies \sup_{x \in [x_{i-1},x_i]} \varphi(f(x)) \leqslant \inf_{y \in [x_{i-1},x_i]} \varphi(f(y)) + \epsilon \\ &\implies M_i^* \leqslant m_i^* + \epsilon. \end{aligned}$$

Therefore the first term $\leqslant \epsilon \sum_{i=1}^{n} \Delta\alpha_i = \epsilon(\alpha(b) - \alpha(a))$.

[4]Dividing a sum into two parts as such and bounding each of them separately is a very useful trick when trying to obtain a bound of something. We will encounter this again in the Weierstraß Approximation Theorem later.

(ii) For the second term, there is not much option for us. Fortunately, since φ is continuous on a compact domain $[m, M]$, it is in particular bounded (Lemma 8.4). If we set $K := |\sup\varphi|$ it is clear that M_i, m_i are both in $[-K, K]$, so $(M_i^* - m_i^*)$ is bounded by $2K$. Since

$$2K = \frac{2K\delta}{\delta} \leqslant \frac{2K(M_i - m_i)}{\delta},$$

the second term is bounded by

$$\sum_{M_i - m_i \geqslant \delta}(M_i^* - m_i^*)\Delta\alpha_i \leqslant \frac{2K}{\delta} \cdot \sum_{M_i - m_i \geqslant \delta}(M_i - m_i)\Delta\alpha_i \leqslant \frac{2K}{\delta}\sum_{i=1}^{n}(M_i - m_i)\Delta\alpha_i < 2K\delta$$

where the last $<$ is by (2). *This explains why we wanted to bound $U - L$ by δ^2 in (2). As $1/\delta$ can get very large, we need something much smaller — δ^2 in this case — to counter it.*

Summarizing what we computed in cases (i) and (ii),

$$U(P, \varphi \circ f, \alpha) - L(P, \varphi \circ f, \alpha) < \epsilon(\alpha(b) - \alpha(a) + 2K).$$

Since the coefficient $\alpha(b) - \alpha(a) + 2K$ is some constant, $U - L$ can be made arbitrarily small, and so $\varphi \circ g \in \mathfrak{R}(\alpha)$ by Theorem 10.15. Done! □

6.2 Properties of the Riemann-Stieltjes Integral

Lemma 11.4

(1) The sum of two R-S integrable functions is R-S integrable, and so is any scalar multiple of a R-S integrable function, i.e.,

$$f_1, f_2 \in \mathfrak{R}(\alpha) \implies f_1 = f_2 \in \mathfrak{R}(\alpha)$$

and

$$f \in \mathfrak{R}(\alpha) \text{ and } c \in \mathbb{R} \implies cf \in \mathfrak{R}(\alpha).$$

In addition,

$$\int_a^b (f_1 + f_2)\, \mathrm{d}\alpha = \int_a^b f_1\, \mathrm{d}\alpha + \int_a^b f_2\, \mathrm{d}\alpha \qquad \text{and} \qquad \int_a^b cf\, \mathrm{d}\alpha = c\int_a^b f\, \mathrm{d}\alpha.$$

(In other words, the set of R-S integrable functions forms a vector space.)

(2) If $f_1 \leqslant f_2$ on $[a,b]$ (i.e., $f_1(x) \leqslant f_2(x)$ for all $x \in [a,b]$), then

$$\int_a^b f_1\, \mathrm{d}\alpha \leqslant \int_a^b f_2\, \mathrm{d}\alpha,$$

if they exist.

(3) If $f \in \mathfrak{R}(\alpha)$ on $[a,b]$ and $c \in [a,b]$, then

$$\int_a^b f\, \mathrm{d}\alpha = \int_a^c f\, \mathrm{d}\alpha + \int_c^b f\, \mathrm{d}\alpha,$$

and in particular $f\big|_{[a,c]}, f\big|_{[c,b]}$ (i.e., f restricted to these two intervals) are both R-S integrable.

(4) If $f \in \mathfrak{R}(\alpha)$ on $[a,b]$ and $|f(x)| \leqslant M$ (i.e., bounded by $[-M,M]$), then

$$\left|\int_a^b f\, \mathrm{d}\alpha\right| \leqslant M(\alpha(b) - \alpha(a)).$$

(5) If $f \in \mathfrak{R}(\alpha_1)$ and $f \in \mathfrak{R}(\alpha_2)$ on $[a,b]$, then $f \in \mathfrak{R}(\alpha_1 + \alpha_2)$, and $f \in R(c\alpha_1)$ for any $c > 0$ (positive c because we need $c\alpha_1$ to be nondecreasing). In addition,

$$\int_a^b f\, \mathrm{d}(\alpha_1 + \alpha_2) = \int_a^b f\, \mathrm{d}\alpha_1 + \int_a^b f\, \mathrm{d}\alpha_2 \qquad \text{and} \qquad \int_a^b f\, \mathrm{d}(c\alpha_1) = c\int_a^b f\, \mathrm{d}(\alpha_2).$$

(6) If c is a constant (function), then $\int_a^b c\, \mathrm{d}\alpha = c(\alpha(b) - \alpha(a))$.

Future reference: Example 11.9, FTC part 1, integration by parts, MVT for integrals, proof of the Picard-Lindelöf Theorem

Theorem 11.5

If $f, g \in \mathfrak{R}(\alpha)$ on $[a,b]$, then so far fg and $|f|$. Also,

$$\left|\int_a^b f\, \mathrm{d}\alpha\right| \leqslant \int_a^b |f|\, \mathrm{d}\alpha.$$

For the inequality, think heuristically about the triangle inequality but applied to a continuum.

Future reference: FTC part 1, integration by parts, uniform convergence & R-S integrals, proof of the Picard-Lindelöf Theorem

Beginning of March 31, 2021

Proof. Notice that $(f(x)+g(x))^2-(f(x)-g(x))^2=4f(x)g(x)$, so

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

Therefore, by the previous lemma, $f+g, f-g\in\mathfrak{R}(\alpha)$. Then using $\varphi(t):=t^2$ from Theorem 11.3, we see that $(f+g)^2,(f-g)^2\in\mathfrak{R}(\alpha)$. Finally, using the previous lemmas a few more times we get $fg\in\mathfrak{R}(\alpha)$.
For $|f|$, simply take $\varphi(t):=|t|$ from Theorem 11.3.
Finally, letting $c:=\pm1$ be such that

$$\left|\int_a^b f\,\mathrm{d}\alpha\right|=c\int_a^b f\,\mathrm{d}\alpha$$

(the absolute value either flips the sign or not), we have

$$\left|\int_a^b f\,\mathrm{d}\alpha\right|=c\int_a^b f\,\mathrm{d}\alpha=\int_a^b (cf)\,\mathrm{d}\alpha\leqslant\int_a^b |f|\,\mathrm{d}\alpha.$$ □

Now we look at worse examples. We begin by looking at α that is discontinuous: zero all the way until one point where it jumps to some value.

Theorem 11.6

Suppose $s\in(a,b)$ and f is bounded on $[a,b]$ and continuous at s. If

$$\alpha(x):=I(x-s),$$

the **unit step function** defined by $I(z)=0$ for $z\leqslant0$ and $I(z)=1$ for $z>1$, then

$$\int_a^b f\,\mathrm{d}\alpha=f(s).$$

Future reference: Example 11.9

Proof. Let $\epsilon>0$. Since f is continuous at s, there exists $\delta>0$ such that $|f(x)-f(s)|<\epsilon/2$ if $|x-s|<\delta$. (We assume δ is small enough that $(s-\delta,s+\delta)\subset[a,b]$.) Now we pick any partition $P:=\{a,x_1,x_2,b\}$ such that

$$s-\delta<x_1<s \qquad \text{and} \qquad s<x_2<s+\delta.$$

Notice that $I(a)=I(s-\delta)=0$ and $I(b)=I(s+\delta)=1$, so $\Delta\alpha_1=\Delta\alpha_3=0$. Also, $\Delta x_2=I(s+\delta)-I(s-\delta)=1-0=1$. Therefore,

$$U(P,f,\alpha)=\sup_{[x_1,x_2]}f\cdot1<f(s)+\frac{\epsilon}{2}\implies\overline{\int_a^b} f\,\mathrm{d}\alpha<f(s)+\frac{\epsilon}{2} \tag{1}$$

and

$$L(P,f,\alpha)=\inf_{[x_1x_2]}f\cdot1>f(s)-\frac{\epsilon}{2}\implies\underline{\int_a^b} f\,\mathrm{d}\alpha>f(s)-\frac{\epsilon}{2}. \tag{2}$$

Combining (1) and (2), we obtain

$$U(P,f,\alpha) - L(P,f,\alpha) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \implies f \in \mathfrak{R}(\alpha)$$

and in particular $f(s) - \frac{\epsilon}{2} \leqslant \int_a^b f \, \mathrm{d}\alpha \leqslant f(s) + \frac{\epsilon}{2}$. Since ϵ is arbitrary, taking $\epsilon \to 0$ gives

$$\int_a^b f \, \mathrm{d}\alpha = f(s).$$

□

Corollary 11.7

This is a generalization of the previous theorem. Let $\sum c_n$ be a convergent series with $c_n \geqslant$, and let $\{s_n\}$ be a sequence of distinct points in (a,b). Define

$$\alpha(x) := \sum_{i=1}^{n} c_n I(x - s_n).$$

(*For example, if $a < s_1 < s_2 < \ldots < b$ then α equals 0 on (a,s_1), c_1 on $[s_1,s_2)$, $c_1 + c_2$ on $[s_2,s_3)$, and so on.*) If f is continuous on $[a,b]$, then

$$\int_a^b f \, \mathrm{d}\alpha = \sum_{i=1}^{n} c_n f(s_n).$$

Proof: see Rudin, Theorem 6.16.

Theorem 11.8

Suppose α is (strictly) increasing on $[a,b]$ and suppose α is differentiable with a Riemann integrable derivative α'. Let f be a bounded function. Then

$$f \in \mathfrak{R}(\alpha) \iff f\alpha' \text{ is Riemann-integrable,}$$

and if so then

$$\int_a^b f \, \mathrm{d}\alpha = \int_a^b f(x)\alpha'(x) \, \mathrm{d}x.$$

This is basically the u-substitution stated formally.

Proof: see Rudin, Theorem 6.17.

Example 11.9: Why R-S Integral?. Let us consider a wire of length one and a mass function

$$m(x) := \text{ mass of the wire contained in } [0,x].$$

Note that m is nondecreasing and $m(0) = 0$. The total mass is (by Lemma 11.4.6)

$$M := \int_0^1 1 \, \mathrm{d}m = m(1) - m(0) = m(1),$$

and the center of mass is given by

$$\frac{1}{M}\int_0^1 x \, \mathrm{d}m$$

(which is well-defined by Theorem 10.17 since x is continuous).

(1) If the wire has a continuous density, i.e., $m'(x) = \rho(x)$ for some continuous ρ, then

$$M = \int_0^1 m'(x) \, \mathrm{d}x = \int_0^1 \rho(x) \, \mathrm{d}x$$

and so

$$\text{center of mass} = \frac{1}{M}\int_0^1 xm'(x)\,\mathrm{d}x = \int_0^1 \rho(x)\,\mathrm{d}x.$$

(2) If the wire is composed only of point masses (i.e., only has mass at some points) $\{m_1, ..., m_n\}$ at points $\{x_1, ..., x_n\}$ respectively, then we need to use to use Theorem 11.6 and get

$$m(x) = \sum_{i=1}^{n} m_i \overline{I}(x - x_i)$$

(*where* $\overline{I}$ *is defined similarly as* I *but* $I(z) = 1$ *for* $z = 0$ *too, since our mass function* $m(x)$ *increases immediately as* x *touches one of the point masses*). It follows that the center of mass is

$$\frac{1}{M}\sum_{i=1}^{n} x_i m_i = \sum_{i=1}^{n} x_i m_i \Big/ \sum_{i=1}^{n} m_i.$$

The R-S integral covers both cases, and it can also cover more extreme cases where the mass function m is continuous but not differentiable. The formulae for total mass and center of mass remain well-defined.

6.3 Riemann-Integration & Differentiation

From now on we will primarily focus on Riemann integration only and “forget about” Stieltjes.

Theorem 11.10: Fundamental Theorem of Calculus (FTC), Part 1[5]

Let $f \in \mathfrak{R}$, i.e., let f be Riemann-integrable (or R-S integrable with $\alpha(x) := x$). Let

$$F(x) := \int_a^x f(t)\,\mathrm{d}t \qquad \text{for } x \in [a,b].$$

Then $F \in C([a,b])$. Furthermore, if f is continuous at x_0 then F is differentiable at x_0 with $F'(x_0) = f(x_0)$.
Future reference: Definition 12.2, proof of the Picard-Lindelöf Theorem

Proof. Since $f \in \mathfrak{R}$, it is by definition bounded. Therefore there exists $M > 0$ such that $|f(x)| \leqslant M$ for all $x \in [a,b]$. Thus, for all $x, y \in [a,b]$ with $x < y$, we have

$$|F(y) - F(x)| = \left| \int_x^y f(t)\,\mathrm{d}t \right| \leqslant M(y-x) \leqslant M|y-x|$$

where the first $\leqslant$ is by Lemma 11.4.3 and second by Lemma 11.4.4. This shows F is Lipschitz and therefore continuous (Lemma 8.13).
Now we show the second claim. Suppose f is continuous at x_0 and let $\epsilon > 0$ be given. Then there exists $\delta > 0$ (assuming WLOG that this δ is sufficiently small with $(y - x_0, y + x_0) \subset [a,b]$) such that

$$|f(y) - f(x_0)| < \epsilon \qquad \text{whenever} \qquad |y - x_0| < \delta.$$

We claim that the “derivative quotient” for small h is close enough to $f(x_0)$. For $0 < h < \delta$, we have

$$\begin{aligned}\left| \frac{F(x_0+h) - F(x_0)}{h} - f(x_0) \right| &= \left| \frac{1}{h} \int_{x_0}^{x_0+h} f(t)\,\mathrm{d}t - \frac{1}{h} \int_{x_0}^{x_0+h} f(x_0)\,\mathrm{d}t \right| \\ &= \left| \frac{1}{h} \int_{x_0}^{x_0+h} (f(t) - f(x_0))\,\mathrm{d}t \right|\end{aligned}$$

(where we have used the fact that $\frac{1}{h} \int_{x_0}^{x_0+h} 1\,\mathrm{d}t = 1$). By Theorem 11.5, we can move the absolute value inside, which gives

$$\ldots \leqslant \frac{1}{h} \int_{x_0}^{x_0+h} |f(t) - f(x_0)|\,\mathrm{d}t < \frac{1}{h} \int_{x_0}^{x_0+h} \epsilon\,\mathrm{d}t = \epsilon$$

(where the $<$ is by Lemma 11.4.2). Clearly we can also do the same computation for $0 > h > -\delta$, and so $F'(x_0)$ is arbitrarily close to $f(x_0)$, i.e., $F'(x_0) = f(x_0)$, completing our proof. □

Beginning of April 1, 2021

Theorem 11.11: FTC Part 2

If $f \in \mathfrak{R}$ on $[a,b]$ and there exists F, a **primitive function** of f, such that $F' = f$, then

$$\int_a^b f\,\mathrm{d}x = F(b) - F(a).$$

Future reference: Definition 12.2, Taylor’s Theorem of integral form, minimization problems

[5] In the lectures, FTC part 1 didn’t have its name, and FTC part 2 was simply regarded as the FTC. I renamed them following the more popular way on Internet.

Proof. Fix $\epsilon > 0$. Let $P = \{x_0, ..., x_n\}$ be a partition of $[a, b]$ such that $U(P, f) - L(P, f) < \epsilon$ [*note that we are using Riemann sums or equivalently Riemann-Stieltjes sums with* $\alpha(x) := x$; also, such partition exists by Theorem 10.15]. Since $x_0 = a$ and $x_n = b$, "adding and subtracting" multiple terms yields

$$F(b) - F(a) = \sum_{i=1}^{n} [F(x_i) - F(x_{i-1})].$$

Since F is differentiable, by MVT, each $F(x_i) - F(x_{i-1})$ is equal to some $F'(t_i)\Delta x_i$ for some $t_i \in [x_{i-1}, x_i]$. Therefore, using Corollay 10.16.3, since $U - L$ for this partition is already $< \epsilon$, we have

$$\left| [F(b) - F(a)] - \int_a^b f \, \mathrm{d}x \right| = \left| \sum_{i=1}^{n} f(t_i)\Delta x_i - \int_a^b f \, \mathrm{d}x \right| < \epsilon.$$

Since ϵ is arbitrary, we see $F(b) - F(a) = \int_a^b f \, \mathrm{d}x$, proving the claim. □

Corollary 11.12: Integration by Parts

Suppose $F, G \in D([a, b])$ are such that $F' = f$ and $G' = g$. Further suppose $f, g \in \mathfrak{R}$. Then

$$\int_a^b Fg \, \mathrm{d}x = F(b)G(b) - F(a)G(a) - \int_a^b fG \, \mathrm{d}x.$$

This is the "$\int u\mathrm{d}v \, \mathrm{d}x = uv - \int v\mathrm{d}u \, \mathrm{d}x$*" which we are all familiar with.*

Proof. It suffices to notice that this is merely the reverse direction of chain rule. If we define $H(x) := F(x)G(x)$, then $H' = F'G + FG'$ which is Riemann integrable (R-S integrable with respect to α) by Lemma 11.4.1 and Theorem 11.5. Then

$$\int_a^b H' \, \mathrm{d}x = H(b) - H(a),$$

and rearranging the terms gives our desired equation. □

Example 11.13. $\int_1^2 \ln x \, \mathrm{d}x = \int_1^2 \ln x \cdot 1 \, \mathrm{d}x = \ln 2 \cdot 2 - \ln 1 \cdot 1 - \int_1^2 \frac{1}{x} \cdot x \, \mathrm{d}x = 2\ln 2 - 1$, where we have chosen $F(x) := \ln x$, $G(x) := 1$, $F'(x) = 1/x$, and $G'(x) = 0$.

Theorem 11.14: Change of Variables

We will state and prove this theorem using the more general Riemann-Stieltjes integrals. Suppose α is increasing on $[a, b]$, $f \in \mathfrak{R}(\alpha)$, and $\varphi : [A, B] \to [a, b]$ is increasing, <u>continuous, and onto</u> (in particular this implies φ is bijective, so $\varphi(A) = a$ and $\varphi(B) = b$). Then

$$\int_a^b f \, \mathrm{d}\alpha = \int_A^B g \, \mathrm{d}\beta \qquad \text{where} \qquad \begin{cases} g(y) := f(\varphi(y)) \\ \beta(y) := \alpha(\varphi(y)). \end{cases}$$

In the special case where $\alpha(x) := x$ *and* $\beta = \varphi$*, we obtain the "*u*-substition formula":*

$$\int_a^b f(x) \, \mathrm{d}x = \int_A^B f(\varphi(y))\varphi'(y) \, \mathrm{d}y. \qquad \text{(Eq.11.1)}$$

Future reference: Lemma 11.16

Proof. Let $P := \{x_0, \ldots, x_n\}$ be any partition of $[a,b]$. Since φ is an increasing bijection, $\{\varphi^{-1}(x_0), \ldots, \varphi^{-1}(x_n)\}$ is an increasing sequnece with $\varphi^{-1}(x_0) = \varphi^{-1}(a) = A$ and $\varphi^{-1}(x_n) = \varphi^{-1}(b) = B$. Thus, $\{\varphi^{-1}(x_0), \ldots, \varphi^{-1}(x_n)\}$ is a partition of $[A,B]$. For notational convenience we denote this set as Q and define $y_k := \varphi^{-1}(x_k)$. [*In particular, all partitions of* $[A,B]$ *can be obtained this way since* φ *is a bijection.*]
Directly following this definition, we have

$$\mathrm{im}_{[x_{i-1},x_i]} f = \mathrm{im}_{[y_{i-1},y_i]} g,$$

so the suprema of both sides agree and so do the infima. Thus,

$$U(P,f,\alpha) = \sum_{i=1}^{n} \underbrace{M_i}_{\text{of } f} [\alpha(x_i) - \alpha(x_{i-1})] = \sum_{i=1}^{n} \underbrace{M_i}_{\text{of } g} [\underbrace{\alpha(\varphi(y_i)) - \alpha(\varphi(y_{i-1}))}_{\beta(y_i)-\beta(y_{i-1})}] = U(Q,g,\beta) \tag{1}$$

and similarly

$$L(P,f,\alpha) = L(Q,g,\beta). \tag{2}$$

Since $f \in \mathfrak{R}$, we have $\sup_P L(P,f,\alpha) = \int_a^b f \, d\alpha = \inf_P U(P,f,\alpha)$ by definition, so (1) and (2) give

$$\sup_Q L(Q,g,\beta) = \int_a^b f \, d\alpha = \inf_Q U(Q,g,\beta). \tag{3}$$

On the other hand, $\sup L = \inf U$ of g also means $g \in \mathfrak{R}(\beta)$ on $[a,b]$ with Riemann-Stieltjes integral equaling to $\sup L$ and $\inf U$. This, along with (3), implies

$$\int_a^b f \, d\alpha = \int_A^B g \, d\beta.$$

□

Example 11.15. $\displaystyle\int_0^{\pi/2} \underbrace{\sin^3(y)}_{\varphi^3(y)} \underbrace{\cos(y)}_{\varphi'(y)} \, dy = \int_0^1 x^3 \, dx = \frac{1}{4}.$

Lemma 11.16

Equation 11.1 is also true for decreasing φ if we add a "–": $\displaystyle\int_a^b f \, dx = -\int_A^B f(\varphi(y))\varphi'(y) \, dy.$
Future reference: Characterization of odd functions

Proof. First notice that

$$\int_a^b f(x) \, dx = \int_{-b}^{-a} f(-x) \, dx \tag{Eq.11.2}$$

(*the proof of this claim is similar to that of Theorem 11.14 where we need to work with images and preimages of partitions*). Then if we take $\psi(y) := -\varphi(y)$ [*note that* $-\varphi$ *is increasing*], Equation 11.1 gives

$$\int_a^b f \, dx = \int_{-b}^{-a} f(-x) \, dx = \int_A^B f(-(-\varphi(y))) - (\varphi)'(y) \, dy = -\int_A^B f(\varphi(y))\varphi'(y) \, dy.$$

□

Example 11.17.

(1) Similar to the previous example, we consider $\displaystyle\int_0^{\pi/2} \cos^3(y)\sin(y) \, dy$. Note that $\cos(\cdot)$ is decreasing on $[0, \pi/2]$. If we define $\varphi(y) := \cos(y)$ and correspondingly $-\varphi'(y) = \sin(y)$, we get

$$\int_0^{\pi/2} \cos^3(y)\sin(y) \, dy = \int_0^1 x^3 \, dx = \frac{1}{4}.$$

(2) If f is continuous on $[-a, a]$ and is odd (i.e., $f(-x) = -f(x)$), then

$$\int_{-a}^{a} f \, \mathrm{d}x = 0.$$

Proof. $\int_{-a}^{a} f \, \mathrm{d}x = \int_{-a}^{0} f \, \mathrm{d}x + \int_{0}^{a} f \, \mathrm{d}x = \int_{0}^{a} f(-x) \, \mathrm{d}x + \int_{0}^{a} f(x) \, \mathrm{d}x = 0$, *where the second = used Equation 11.2 (right above).*

We will show that the converse holds too, i.e., if $\int_{-a}^{a} f \, \mathrm{d}x = 0$ *for all* $a > 0$ *then* f *is odd.*

6.4 MVT and Taylor's Theorem with Integrals

Lemma 11.18: MVT for Integrals

If $f \in C([a,b])$, $g \geqslant 0$ (i.e., a nonnegative function) integrable, and $g \in \mathfrak{R}$, then there exists a $\xi \in (a,b)$ such that

$$\int_a^b f(x)g(x)\,\mathrm{d}x = f(\xi)\int_a^b g(x)\,\mathrm{d}x.$$

(*Think of this as the "weighted average".*) In particular, if $g \equiv 1$, then

$$\int_a^b f(x)\,\mathrm{d}x = f(\xi)(b-a).$$

Proof. If the integral of g is 0 then we are immediately done, since any ξ works. Now we suppose the integral of g is not 0 (so it's positive). Define

$$m := \inf_{[a,b]} f \qquad \text{and} \qquad M := \sup_{[a,b]} f.$$

It follows that $m \leqslant f(x) \leqslant M$ for all $x \in [a,b]$. Therefore,

$$m\int_a^b g\,\mathrm{d}x \leqslant \int_a^b fg\,\mathrm{d}x \leqslant M\int_a^b g\,\mathrm{d}x$$

by using Lemma 11.4.2 twice. Since $\int_a^b g(x)\,\mathrm{d}x \neq 0$, we can divide the above inequalities by it and obtain

$$m \leqslant \int_a^b fg\,\mathrm{d}x \Big/ \int_a^b g\,\mathrm{d}x \leqslant M.$$

Since f is continuous on a compact domain, it attains its bounds, so there exist x_1, x_2 such that $f(x_1) = m$ and $f(x_2) = M$. The existence of ξ with $f(\xi) = \int_a^b fg\,\mathrm{d}x \big/ \int_a^b g\,\mathrm{d}x$ then follows from the IVP. □

Beginning of April 5, 2021

Lemma 11.19

Suppose that $f \in C([a,b])$ and $\int_\alpha^\beta f\,\mathrm{d}x = 0$ for all $a < \alpha < \beta < b$. Then $f \equiv 0$.

Proof. Clearly, it suffices to prove that $f(x_0) = 0$ for all $x_0 \in [a,b]$, so let us first pick an arbitrary $x_0 \in [a,b]$. We also define a sequence $(h_n)_{n=1}^{\infty} \subset (0, b-x_0)$ [i.e., each lying inside $(0, b-x_0)$] such that $h_n \to 0$. It follows that each $x_0 + h_n$ is still in $[a,b]$, so by the previous Lemma (MVT for integrals), to each n corresponds a $\xi_n \in (0,1)$ such that

$$0 = \int_{x_0}^{x_0+h_n} f\,\mathrm{d}x = f(x_0 + \xi_n h_n)h_n.$$

(Here any point in (x_0, x_0+h_n) can obviously be written as $x_0 + kh_n$ for $k \in (0,1)$, and $(x_0+h_n) - x_0 = h_n$.) Since $h_n \neq 0$, we obtain $f(x_0 + \xi_n h_n) = 0$ for all n. Since $|\xi_n|$ is bounded by 1 and $h_n \to 0$, we see $\xi_n h_n \to 0$ as $n \to \infty$, and so $x_0 + \xi_n h_n \to 0$. Using Heine's sequential definition of continuity, $f(x_0) = 0$. This proves the claim. □

Example 12.1: Characterization of Odd Functions. If $f \in C([-a,a])$ is such that $\int_{-\alpha}^{\alpha} f\,\mathrm{d}x = 0$ for all $\alpha \in (0,a)$, then it is an odd function, i.e., $f(-x) = -f(x)$.

Proof. We begin by picking any α and α' satisfying $0 < \alpha' < \alpha < a$. On one hand, using Equation 11.2,

$$0 = \int_{-\alpha}^{\alpha} f \, dx = \int_0^{\alpha} f(x) \, dx + \int_0^{\alpha} f(-x) \, dx = \int_0^{\alpha} [f(x) + f(-x)] \, dx. \tag{1}$$

On the other hand, since $0 < \alpha' < alpha$, we also have

$$0 = \int_{-\alpha'}^{\alpha'} f \, dx = \int_0^{\alpha'} [f(x) + f(-x)] \, dx. \tag{2}$$

Subtracting (2) from (1), we obtain

$$\int_{\alpha'}^{\alpha} [f(x) + f(-x)] \, dx = 0 \qquad \text{for all} \qquad 0 < \alpha' < \alpha < a.$$

The claim then follows from the previous lemma. (*At first we will only obtain the claim for $x \in [0, \alpha]$, but since the function is odd, the claim can be extended to $x \in [-a, a]$.*) □

Definition 12.2

If $a > b$, we define $\displaystyle\int_a^b := -\int_b^a$.

In particular, the theorems we have previously proven still holds for this new definition.

If $x \leqslant b$ and $F(x) := \int_b^x f(t) \, dt$, then $F(x) = -\int_x^b f(t) \, dt$ so F is continuous and $F'(x_0) = f(x_0)$ for all points of continuity of f. This is FTC part 1, and the proof is analogous. Also, FTC part 2 holds as (assuming $a > b$)

$$\int_a^b f(t) \, dt = -\int_b^a f(t) \, dt = -(F(a) - F(b)) = F(b) - F(a).$$

Theorem 12.3: Taylor's Theorem with Remainder of Integral Form

Let $x_0 \in \mathbb{R}$ and suppose that $f : \mathbb{R} \to \mathbb{R}$ is such that $f^{(n)} \in C(I)$ for some open interval containing x_0. Then,

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \int_{x_0}^{x} f^{(n)}(t) \frac{(x-t)^{n-1}}{(n-1)!} \, dt \qquad \text{for all } x \in I.$$

Future reference: Taylor's Theorem (summary)

Proof. By the FTC part 2,

$$\begin{aligned} f(b) &= f(a) + \int_a^b f'(t) \, dt = f(a) + \int_a^b f'(t) \frac{d}{dt}(t - b) \, dt \\ \text{[integrate by parts]} &= f(a) + f(a)(b - b) - f(a)(a - b) - \int_a^b f''(t)(t - b) \, dt \\ &= f(a) + f(a)(b - a) + \int_a^b f''(t)(b - t) \, dt. \end{aligned}$$

Note that we can iteratively integrate by parts:

$$\begin{aligned} \int_a^b \frac{f^{(k)}(t)}{(k-1)!}(b-t)^{k-1} \, dt &= \int_a^b \frac{f^{(k)}(t)}{(k-1)!} \frac{d}{dt}\left[-\frac{(b-t)^k}{k}\right] dt \\ &= \frac{f^{(k)}(a)}{k!}(b-a)^k + \int_a^b \frac{f^{(k+1)}(t)}{k!}(b-t)^k \, dt. \end{aligned}$$

Putting everything together, since $f^{(n)} \in C(I)$, we obtain

$$f(b) = f(a) + \ldots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \int_a^b \frac{f^{(n)}(t)}{(n-1)!}(b-t)^{n-1}\,\mathrm{d}t,$$

and the claim follows by letting $a = x_0, b = x$. □

Remark. We actually recover the Peano form from this integral. By PS11.4 (the *Second Mean Value Theorem*), the integral is equal to

$$\frac{(x-x_0)^{n-1}}{(n-1)!}\int_{x_0}^{\xi} f^{(\xi)}(t)\,\mathrm{d}t \qquad \text{for some } \xi \in (x_0, x).$$

This is $o(x-x_0)^{n-1}$, since continuity of $f^{(n-1)}$ implies $f^{(n-1)}(\xi) - f^{(n-1)}(x_0) \to 0$ as $x \to 0$ as $\xi \in (x_0, x)$.

6.5 Improper Integrals

Definition 12.4: Improper Integrals

If $f \in \mathfrak{R}$ on $[a, b]$ for all $b > a$, then we define

$$\int_0^\infty f \, \mathrm{d}x := \lim_{b \to \infty} \int_a^b f \, \mathrm{d}x, \tag{1}$$

when such limit exists.

If $b > a_0$ and $f \in \mathfrak{R}$ on $[a, b]$ for all $a \in (a_0, b)$, then we define

$$\int_{a_0}^b f \, \mathrm{d}x := \lim_{a \to a_0} \int_a^b f \, \mathrm{d}x. \tag{2}$$

We can also define $\int_{-\infty}^b f \, \mathrm{d}x$. Finally, we define

$$\int_{-\infty}^\infty f \, \mathrm{d}x := \lim_{a \to -\infty} \int_a^0 f \, \mathrm{d}x + \lim_{b \to \infty} \int_0^b f \, \mathrm{d}x, \tag{3}$$

when both limits exist.

If an integral of any form above exists, we say that that **improper integral converges**. If it converges with f replaced by $|f|$ then we say it **converges absolutely**.

Lemma 12.5

(1) If $\int_0^\infty f \, \mathrm{d}x$ converges absolutely then it converges.

(2) If $\int_0^\infty g \, \mathrm{d}x$ converges and $g \geqslant |f|$, i.e., $g(x) \geqslant |f(x)|$ for all x, then $\int_0^\infty f \, \mathrm{d}x$ converges absolutely.

Future reference: Integral test

Proof of (2). Since $g \geqslant |f|$, it is in particular nondecreasing. In particular, this means

$$F \text{ nondecreasing} \implies \lim_{x \to \infty} F(x) = \sup_{x > 0} F(x). \tag{Eq.12.1}$$

Thus,

$$\int_0^x |f| \, \mathrm{d}x \leqslant \int_0^x g \, \mathrm{d}x \leqslant \sup_{y > 0} \int_0^y g \, \mathrm{d}x = \lim_{y \to \infty} \int_0^y g \, \mathrm{d}x = \int_0^\infty g \, \mathrm{d}x.$$

(The first $\leqslant$ is because $|f \leqslant g|$, and the first $=$ is by Equation 12.1.) Therefore, $\int_0^x |f| \, \mathrm{d}x$ is bounded, so

$$\sup_{y > 0} \int_0^y |f| \, \mathrm{d}x \text{ exists and } = \lim_{y \to \infty} \int_0^y |f| \, \mathrm{d}x = \int_0^\infty |f| \, \mathrm{d}x.$$

Therefore $\int_0^\infty f \, \mathrm{d}x$ converges absolutely, proving the claim. □

Example 12.6.

(1) $\displaystyle \int_1^\infty \frac{1}{x^4} \, \mathrm{d}x = \lim_{b \to \infty} \int_1^b \frac{1}{x^4} \, \mathrm{d}x = \lim_{b \to \infty} \frac{1}{3} \left(1 - \frac{1}{b^3} \right) = \frac{1}{3}.$

(2) $\int_{-\infty}^{\infty} \frac{\sin x}{1+x^4}\, dx$ converges absolutely, since

$$\begin{aligned}\int_{-\infty}^{\infty} \left|\frac{\sin x}{1+x^4}\right| dx &= \int_{-\infty}^{-1} + \int_{-1}^{1} + \int_{1}^{\infty} \\ &= 2\int_1^\infty \underbrace{\left|\frac{\sin x}{1+x^4}\right|}_{\leqslant |\sin x/x^4| \leqslant 1/x^4} dx + \int_{-1}^{1} \underbrace{\left|\frac{\sin x}{1+x^4}\right|}_{\leqslant |\sin x| \leqslant 1} dx \\ &\leqslant 2\int_1^\infty x^{-4}\, dx + 2 = \frac{8}{3}.\end{aligned}$$

Beginning of April 9, 2021

Theorem 12.7: Integral Test / Maclurin-Cauchy Test

Suppose that for some $N \in \mathbb{N}$, $f(x)$ is nonincreasing and nonnegative for all $x \geqslant N$. Then

$$\sum_{n \geqslant N} f(n)^6 \text{ converges} \iff \int_N^\infty f\, dx \text{ converges.}$$

Proof. The proof is similar to the Calc II version, only to be more rigorous. For $\implies$,

$$\int_N^\infty f\, dx = \lim_{b\to\infty} \int_N^b f\, dx = \sup_{b>N} \int_N^b f\, dx$$

where the last = is by Equation 12.1. The integral on the RHS can be bounded by

$$\sum_{n=N}^{\lfloor b\rfloor -1} \int_n^{n+1} f\, dx + \int_{\lfloor b\rfloor}^{b} f\, dx \leqslant \sum_{n=N}^{\lfloor b\rfloor -1} \int_n^{n+1} f(n)\, dx + \int_{\lfloor b\rfloor}^{b} f(\lfloor b\rfloor)\, dx \leqslant \sum_{n \geqslant N} f(n) < \infty.$$

Therefore the supremum is finite and the integral converges.

For $\impliedby$, notice that $f(n) \leqslant f(n-1)$. Therefore, the partial sums

$$S_m := \sum_{n=N}^{m} f(n) = f(N) + \sum_{n=N+1}^{m} \int_{n-1}^{n} f(n)\, dx \leqslant f(N) + \int_N^m dx \leqslant f(N) + \int_N^m f\, dx.$$

Taking the supremum over m gives

$$\sup_m S_m \leqslant f(N) + \sup_m \int_N^m f\, dx = f(N) + \lim_{m\to\infty} \int_N^m f\, dx = f(N) + \int_N^\infty f\, dx < \infty.$$

Therefore the series $\sum f(n)$ converges, as claimed. □

[6] The summation $\sum_{n \geqslant N}$ is a shorthand notation for $\sum_{n=N}^{\infty}$.

Chapter 7

Sequences & Series of Functions

7.0 The Moore-Smith Theorem

Recall in Example 0.3 we stated that the interchange of limits can sometimes cause trouble:

$$\text{If } a_{m,n} := \frac{m}{m+n} \text{ then } \lim_{n\to\infty}\lim_{m\to\infty} a_{m,n} = 1 \neq 0 = \lim_{m\to\infty}\lim_{n\to\infty} a_{m,n}.$$

It turns out that the mere convergence of $(a_{m,n})_{n\geqslant 1}$ and $(a_{m,n})_{m\geqslant 1}$ are insufficient for our desired result. Instead, we need what is called *uniform convergence* with respect to the other variable, a stronger condition.

Theorem 12.8: Moore-Smith Theorem

Suppose $(a_{m,n})$ is a sequence with two indices, and suppose that there exist sequences $(y_n)_{n\geqslant 1}, (z_n)_{n\geqslant 1}$ such that

(1) $(a_{m,n})_{m\geqslant 1} \to z_n$ pointwise, i.e., when n is fixed, the sequence $(a_{1,n}, a_{2,n}, \ldots)$ converges to z_n, and

(2) $(a_{m,n})_{n\geqslant 1} \to y_m$ uniformly, i.e.,

Given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $|a_{m,n} - y_m| < \epsilon$ for all $n \geqslant N$ and for all m.

Alternatively, the condition can be written as $\sup_m |a_{m,n} - y_m| \leqslant \epsilon$ if $n \geqslant N$.

(*Compare the difference between two modes of convergence with two modes of continuity; they bear similarities, in particular the choice of $N \in \mathbb{N}$ and the $\delta > 0$ and what their relation to ϵ and the other variable.*)

Then the limit of the doubly indexed sequence is interchangeable, i.e.,

$$\lim_{n\to\infty}\lim_{m\to\infty} a_{m,n} = \lim_{m\to\infty}\lim_{n\to\infty} a_{m,n} = \lim_{n,m\to\infty} a_{m,n}.$$

(*We say $g = \lim_{n,m\to\infty} a_{m,n}$ is the limit of the doubly indexed sequence if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $|a_{m,n} - g| < \epsilon$ when $m, n \geqslant N$. In other words, "the sequence is arbitrarily close to g when both indices are large."*)
Future reference: Uniform convergence and limit points

Proof. (Completed on 4/12.) The prove consists of three steps.

Step 1. *The sequence* (y_m) *converges.* The proof uses a very common and useful "$\epsilon/3$" argument (which we will encounter later too). Since $(a_{m,n})_{n\geqslant 1} \to y_m$ *uniformly*, there exists some $N \in \mathbb{N}$ such that

$$|a_{m,n} - y_m| < \frac{\epsilon}{3} \qquad \text{for all } n \geqslant N \text{ and for all } m. \tag{1}$$

On the other hand, since $(a_{m,N})_{m\geqslant 1}$ converges to z_N pointwise, they form a Cauchy sequence in particular, so there also exists a large $M \in \mathbb{N}$ such that

$$|a_{m_1,N} - a_{m_2,N}| < \frac{\epsilon}{3} \qquad \text{for all } m_1, m_2 \geqslant M. \tag{2}$$

Combining (2) with (1) with $m = m_1$ and $m = m_2$, we obtain

$$|y_{m_1} - y_{m_2}| \leqslant |y_{m_1} - a_{m_1,N}| + |a_{m_1,N} - a_{m_2,N}| + |a_{m_2,N} - y_{m_2}| < \epsilon \qquad \text{for } m_1, m_2 \geqslant M.$$

Therefore $(y_m)_{m\geqslant 1}$ is Cauchy in $\mathbb{R}$ and therefore converges to some $y \in \mathbb{R}$.

Step 2. *We now show that* $y = \lim_{m,n\to\infty} a_{m,n}$*, i.e.,* $a_{m,n}$ *is close to* y *when* m, n *are both large.* Easy. Let $\epsilon > 0$ be given. Since $y_m \to y$, there exists $M \in \mathbb{N}$ such that $|y_m - y| < \epsilon/2$ for $m \geqslant M$. Also, since $(a_{m,n})_{m\geqslant 1} \to y_m$ uniformly, there exists $N \in \mathbb{N}$ such that $|a_{m,n} - y_m| < \epsilon/2$ for all $n \geqslant N$ *and all* m. Thus, for $m, n \geqslant \max(M, N)$, both conditions are simultaneously satisfied, and so

$$|a_{m-n} - y| \leqslant |a_{m,n} - y_m| + |y_m - y| < \epsilon.$$

Step 3. *Finally, we show* $z_n \to y$. From Step 2, we pick $K \in \mathbb{N}$ such that $|a_{m,n} - y| < \epsilon$ for all $m, n \geqslant K$. If we let $m \to \infty$, then $a_{m,n}$ (for this fixed n) converges to z_n. Therefore,

$$|z_n - y| = \lim_{m\to\infty} |a_{m,n} - y| \leqslant \epsilon \text{ for } n \geqslant N$$

(where $<$ becomes $\leqslant$ by Theorem 5.8). This shows $z_n \to y$ and concludes the proof of the theorem. □

Example 12.9. The sequence $a_{m,n} := (m+n)^2/2^{mn}$ satisfies

$$\lim_{n\to\infty} \lim_{m\to\infty} a_{m,n} = \lim_{m\to\infty} \lim_{n\to\infty} a_{m,n} = 0.$$

Proof. It is clear that $a_{m,n} \to 0$ at least pointwise with respect to either variable. By the previous theorem, it suffices to show that one of the convergence is uniform. (In fact, both are.) Using AM-GM,

$$|a_{m,n}| = \frac{(m+n)^2}{2^{mn}} \leqslant \frac{4(mn)^2}{2^{mn}}$$

which can be made arbitrarily small, as

$$\lim_{x\to\infty} \frac{x^2}{2^x} \overset{H}{=} \dots \overset{H}{=} 0.$$

Therefore, we just need to make the product of mn sufficiently large to ensure $|a_{m,n}| < \epsilon$. In particular, since $n \geqslant 1$, we can pick a sufficiently large m such that mn is always sufficiently large regardless of the value of n. This shows uniform convergence with respect to m, and the claim follows by Moore-Smith. □

7.1 Introduction – Why *Uniform* Convergence of Functions?

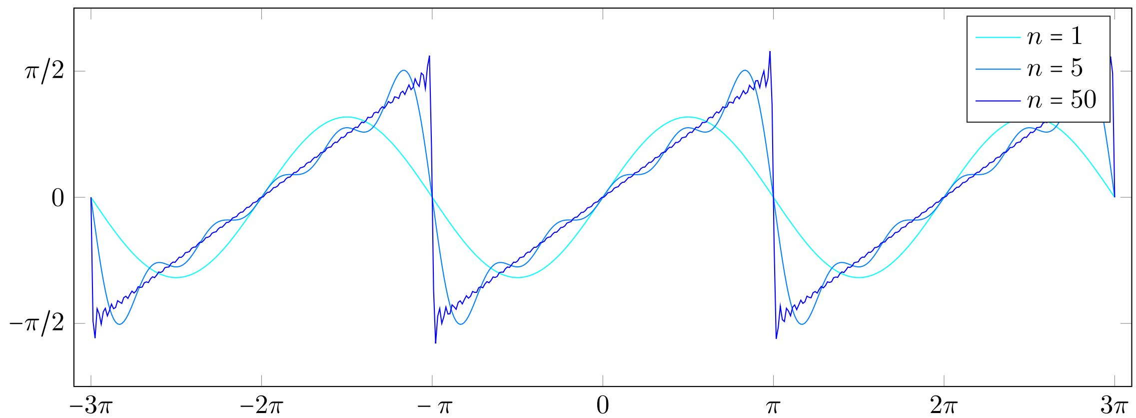
A short history[1] of why we need a stronger mode of convergence of functions called *uniform convergence* and in particular why pointwise convergence of functions isn't sufficient:

(1) In 1821, Cauchy hypothesized that if $f_n : \mathbb{R} \to \mathbb{R}$ are continuous and $\sum_{n=1}^{\infty} f_n$ converges to f pointwise, i.e., if $\sum_{n=1}^{\infty} f_n(x) \to f(x)$ for all $x \in \mathbb{R}$, then f is continuous.

(2) In 1826, Abel provided a counterexample to above: consider the *Fourier series* of the *discontinuous* function $f(x) = x/2$ on $(-\pi, \pi)$ but extended 2π-periodically to $\mathbb{R}$ with $f(2\pi) = 0$. (The graph consists of parallel line segments. See graph below.)

It was known that

$$f(x) = \sin(x) = \frac{1}{2}\sin(2x) + \frac{1}{3}\sin(3x) - \ldots = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\sin(nx).$$



From this we see that the pointwise limit of a sequence of continuous functions may be discontinuous. We would also run into problems if we try to differentiate the series term by term, obtaining

$$\cos(x) - \cos(2x) + \cos(3x) - \ldots$$

which diverges for most x, whereas the derivative of f exists *almost everywhere* and equals to $1/2$.
We will now formally introduce the notion of uniform convergence and derive many of its nice properties — for example, how uniform convergence interacts with limits, series, integrals, and derivatives in ways that (standard) convergence may fail to.

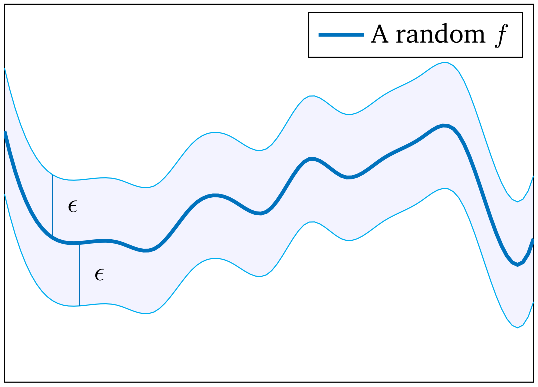
7.2 Uniform Convergence of Functions

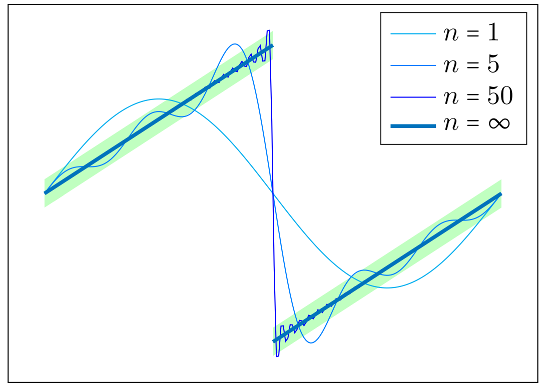
Definition 12.10: Pointwise and Uniform Convergence of Functions

Let $(f_n)_{n\geqslant 1}$ be a sequence of functions. We say that $f_n : X \to Y$ **converges** to $f : X \to Y$ **pointwise** if $f_n(x) \to f(x)$ for each $x \in X$. We say f_n **converges** to f **uniformly** (on X)[2] if $\sup_{x\in X} |f_n(x) - f(x)| \to 0$.
Equivalently, given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \epsilon$ for all $x \in X$ and all $n \geqslant N$. (*In particular, uniform convergence implies pointwise convergence.*)

[1] From my 425b, taught by Prof. Andrew Manion. Lecture notes, sample HWs, and exams can be found on my website.

To illustrate the difference between convergence and uniform convergence: if $f_n \to f$ uniformly, then for all large enough n's, the corresponding f_n's need to be contained by the "ϵ-tube[3]" of f, as shown in the left figure. Letting $\epsilon \to 0$, it becomes clear that all $x \in X$ need to synchronously approach their corresponding limits on f, hence the word "uniform". The counterexample provided by Abel clearly fails to satisfy this criterion: near π, every single f_n jumps out of the green ϵ-tube, so the convergence is not uniform, which is (we'll show soon) precisely why Cauchy's hypothesis is false.



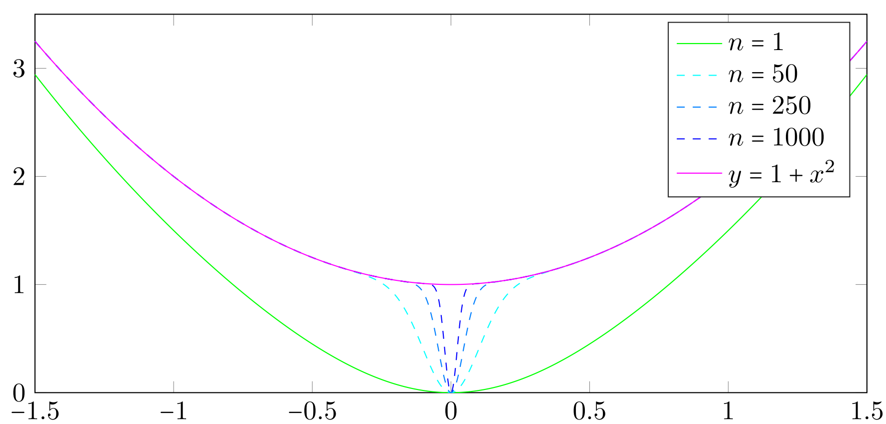


Example 12.11: Examples Where Convergence isn't Good Enough.

(1) (Another example that) continuity may fail if only assuming pointwise convergence: define

$$f_n(x) := x^2 \sum_{k=0}^{n} (1+x^2)^{-k} \qquad \text{and} \qquad f(x) := \begin{cases} \dfrac{x^2}{1 - 1/(1+x^2)} = 1 + x^2 & x \neq 0 \\ 0 & x = 0 \end{cases}$$

(Note that $(1 + x^2) > 1$ as $x \neq 0$, and since $-k < 0$, the sum forms a geometric series, so indeed $f_n(x) \to f(x)$ pointwise.) However, it is clear that although each f_n is continuous (they are in fact C^∞, i.e., smooth!), f is not at 0 (not even C^1).



(2) Limit of integral need not equal integral of limit if only assuming pointwise convergence: if

$$f_n(x) := n^2 x(1 - x^2)^n \qquad \text{on } [0, 1],$$

[2]Notations include $f_n \rightrightarrows f$, but I will instead say "$f_n \to f$ uniformly" every time to strengthen memory.

[3]I first saw this in Pugh's book, and I loved this description. Vivid, intuitive, and self-explanatory.

then $f_n(x) \to 0$ for all $x \in [0,1]$ (see Rudin, Theorem 3.20(d)). However,

$$\int_0^1 f_n(x)\,\mathrm{d}x = n^2 \int_0^1 (\underbrace{1-x^2}_{=:y})^n\,\mathrm{d}x = \frac{n^2}{2}\int_0^1 y^n\,\mathrm{d}y = \frac{n^2}{2n+1} \to \infty,$$

so this example demonstrates

$$\int_0^1 \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = 0 \neq \lim_{n\to\infty}\int_0^1 f_n(x)\,\mathrm{d}x.$$

Theorem 12.12: Uniformly Convergent & Cauchy w.r.t. Sup Metric

For the space of real-valued functions on X, we define a metric, called the **sup metric**, written $\|\cdot\|_{\sup}$, by

$$\|f\|_{\sup} := \sup_{x\in X}|f(x)|.$$

Then (*think of completeness, i.e., convergence* $\Leftrightarrow$ *Cauchy-ness but w.r.t. function norms; we will talk about this in detail in Theorem 13.3.*)

$$f_n \to f \text{ uniformly} \iff (f_n)_{n\geqslant 1} \text{ forms a Cauchy sequence w.r.t. } \|\cdot\|_{\sup}.$$

In other words, the RHS states that, for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $\|f_n - f_m\| = \sup_{x\in X}|f_n(x)-f_m(x)| < \epsilon$ for all $m,n \geqslant N$.

Future reference: Uniform convergence & derivatives, Weierstraß M-Test, Arzelá-Ascoli Theorem and its proof

Proof. We first show $\Longleftarrow$ using an $\epsilon/2$ argument. Given $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $\|f_n - f\| < \epsilon/2$ for all $n \geqslant N$. Therefore, for $m,n \geqslant N$, and any x, we have

$$|f_n(x) - f_m(x)| \leqslant \underbrace{|f_n(x) - f(x)|}_{<\epsilon/2} + \underbrace{|f(x) - f_m(x)|}_{<\epsilon/2} < \epsilon.$$

Therefore, taking supremum over all x, we obtain

$$\|f_n - f_m\|_{\sup} = \sup_{x\in X}|f_n(x) - f_m(x)| \leqslant \epsilon.$$

For $\Longleftarrow$, also let $\epsilon > 0$ be given. By "Cauchy-ness"[4], there exists $N \in \mathbb{N}$ such that

$$|f_n(x) - f_m(x)| < \epsilon \qquad \text{for all } x \in X \text{ and all } m,n \geqslant N. \tag{Δ}$$

In particular, this means that, for all x, the sequence of *real numbers* $(f_n(x))_{n\geqslant 1}$ is Cauchy. Since $\mathbb{R}$ is complete, there exists some real number, which we call $f(x)$, such that $f_n(x) \to f(x)$. *Therefore we can define a function f by setting its value at x to be $f(x)$, the limit of $(f_n(x))_{n\geqslant 1}$.*

Taking $\lim_{m\to\infty}$ in (Δ), we see that $f_m(x) \to f$, and since $|\cdot|$ is continuous, $|f_n(x) - f_m(x)| \to |f_n(x) - f(x)|$. By Theorem 5.8, we obtain

$$|f_n(x) - f(x)| \leqslant \epsilon \qquad \text{for all } x \in X \text{ and } n \geqslant N.$$

which proves the uniform convergence. ☐

[4] I quoted this "Cauchy-ness" because we haven't taken a rigorous approach to prove the important properties of $\|\cdot\|_{\sup}$.

7.3 Uniform Convergence & Continuity

Beginning of April 12, 2021

Theorem 13.1: Uniform Convergence & Limit Points

Suppose that $f_n \to f$ uniformly (both functions are from X to $\mathbb{R}$). Also suppose that there exists a limit point $x \in X$ such that $\lim_{y\to x} f_n(y) = \alpha_n$ for all n. Then $\alpha_n \to \alpha := \lim_{y\to x} f(y)$.

Notice that we did not make any assumption on continuity; the α's only needed to be limit points.

Future reference: Uniform convergence & derivatives, uniform convergence & series

Proof. Let $(y_m)_{m\geqslant 1} \subset X$ be any sequence that converges to x. We need to show that $f(y_m) \to \alpha$. We define $a_{m,n} := f_n(y_m)$. By assumption, as $m \to \infty$, $f_n(y_m) \to \alpha_n$ for all n, so this gives the first condition in Moore-Smith. The uniform convergence of f_n gives the second condition. Therefore by Moore-Smith, $\lim_{n\to\infty} \alpha_n$ exists and limits are interchangeable, i.e.,

$$\alpha = \lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \lim_{m\to\infty} f_n(y_m) = \lim_{m\to\infty} \lim_{n\to\infty} f_n(y_m) = \lim_{m\to\infty} f(y_m). \qquad \square$$

Corollary 13.2: Uniform Convergence Preserves Continuity

Following the previous theorem: if in particular f_n converges uniformly to f and if each f_n is continuous at x, then f is continuous at x. If each f_n is continuous on all of X, then f is continuous on all of f. Continuity is preserved by uniform convergence!

Future reference: Uniform convergence & series, radii of convergence & derivatives

Corollary 13.3

Let $K \subset X$ be a compact set. Then $(C(K), \|\cdot\|_{\sup})$ [i.e., the space of continuous functions on K with sup norm as metric] is a complete metric space.

Proof. It is quite easy to show that this is indeed a metric space (directly by definition). Note that since functions on a compact domain are bounded, the corresponding $\|\cdot\|_{\sup}$ is always finite.

Now for completeness, suppose that $(f_n) \subset C(K)$ is Cauchy. By Theorem 12.12, $f_n \to f$ uniformly for some f, and by the previous corollary, this f is continuous! This concludes the proof. $\square$

I would like to include an interlude here, inspired by Pugh's book, my 425a, and also an exercise from Rudin's Real and Complex Analysis (RCA). Since we have shown the superiority of uniform convergence, it is natural to attempt to "upgrade" convergence to uniform convergence. However, this process turns out to be not easy. Let $f_n \to f$ pointwise.

(1) *Can we upgrade it if f_n and f are on a compact domain (recall compactness is nice)? The answer is no. This should be obvious from the examples plotted before. Compact domain upgrades continuity to uniform continuity but not convergence to uniform convergence!*

(2) *What if in addition we require f to continuous? Still no. For example, Pugh mentiond the growing steeple,*

$$f_n(x) := \begin{cases} n^2 x & 0 \leqslant x \leqslant 1/n \\ 2n - n^2 x & 1/n \leqslant x < 2/n \\ 0 & 2/n \leqslant x \leqslant 1. \end{cases}$$

In other words, the graph of f_n connects $(0,0), (1/n, n), (2/n, 0)$, and $(1,0)$. See the first figure below. As spikes keep coming up, there is no way to make f_n uniformly close to the zero function, the pointwise limit of f_n.

(3) *What if in addition we require $\lim_{n\to\infty} \int_X f_n \, \mathrm{d}x = \int_X f \, \mathrm{d}x$? Still no. I have come up with a "modified growing steeple" while solving a problem from Rudin's RCA:*

$$f_n(x) := \begin{cases} n(n+2)(n+3) & 1/(n+3) \leqslant x < 1/(n+2) \\ n & 1/(n+2) \leqslant x < 1/(n+1) \\ n(n+1) - n^2(n+1)x & 1/(n+1) \leqslant x < 1/n \\ 0 & 1/n \leqslant x \leqslant 1. \end{cases}$$

The graph of f_n connects $(0,0), (1/(n+3), 0)$, $(1/(n+2), n)$, $(1/(n+1), n)$, $(1/n, 0)$, and $(1,0)$. In addition to not converging uniformly, the pointwise supremum of f_n, $\sup f_n$ defined by

$$(\sup f_n)(x) := \sup_{n\in\mathbb{N}} f_n(x)$$

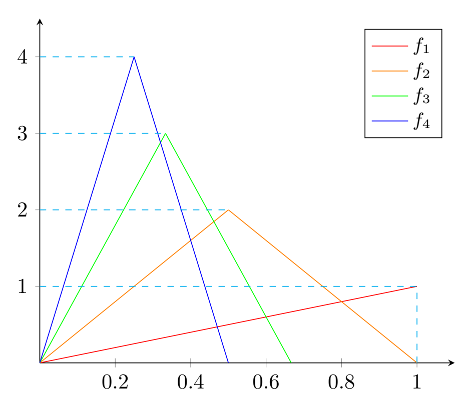
is "not L^1", meaning that the integral $\int_0^1 \sup f_n \, \mathrm{d}x$ is in fact ∞, even though $\lim_{n\to\infty} \int_0^1 f_n \, \mathrm{d}x \to 0$[!]

(4) *What if in addition we require f_n to be uniformly bounded, i.e., there exists $M \in \mathbb{R}^+$ such that $|f_n(x)| \leqslant M$ for all $x \in [0,1]$ and all $n \in \mathbb{N}$? You might have guessed... still no. Consider*

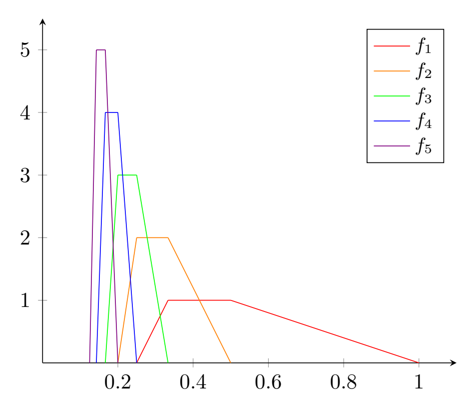
$$f_n(x) := \begin{cases} nx & 0 \leqslant x \leqslant x < 1/n \\ 2 - nx & 1/n \leqslant x < 2/n \\ 0 & 2/n \leqslant x \leqslant 1. \end{cases}$$

The graph of f_n connects $(0,0)$, $(1/n, 1)$, $(2/n, 0)$, and $(1,0)$. It is similar to the growing steeples except it is no longer growing – all of them peak at 1.

Growing Steeples



Growing Steeples, Modified



However, persistence does pay off — what we are missing is monotonicity, and this leads to the following theorem.

Theorem 13.4: Dini's Theorem

Let K be compact. Suppose that $f_n : K \to \mathbb{R}$ are continuous and nonincreasing (or monotone decreasing, i.e., $f_n \geqslant f_{n+1}$). If $f_n \to f$ for some *continuous* $f : K \to \mathbb{R}$, then $f_n \to f$ uniformly.
To sum up: compact domain + monotone pointwise convergence + continuous limit = uniform convergence.

Proof. WLOG assume $f \equiv 0$ (otherwise we can consider $g_n := f_n - f$ whose limit is still the zero function). Let $\epsilon > 0$ and define $K_n := \{x \in K : f(x) \geqslant \epsilon\}$. Since $[\epsilon, \infty)$ is closed in $\mathbb{R}$ and f continuous, the closed set condition implies that K_n is closed. Since K is compact, we see K_n is also compact (Example 4.1). Also, it is clear that $K_{n+1} \subset K_n$, since f_n is monotone decreasing, so if $f_{n+1}(x) \geqslant \epsilon \Rightarrow f_n(x) \geqslant f_{n+1}(x) \geqslant \epsilon$.
Since for all x, $f_n(x) < \epsilon$ for sufficiently large n (it needs to converge to 0 and $\epsilon > 0$), so x cannot lie in the *infinite* intersection of K_n's. Since this holds for all $x \in X$,

$$\bigcap_{n=1}^{\infty} K_n = \varnothing.$$

By the contrapositive of Corollary 4.4, this implies that some K_N (and $K_{N+1}, \ldots$) must be empty. Therefore, $f_n(x) < \epsilon$ for all $x \in K$ and $n \geqslant N$, which is precisely the condition for uniform convergence of $f_n \to f \equiv 0$. ☐

Beginning of April 14, 2021

Example 13.5: Dini's Theorem Fails Without Compact Domain. For $x \in (0,1)$, consider

$$f_n(x) := \frac{1}{nx+1}.$$

It follows that $f_n \to 0$ pointwise, $f_{n+1} \leqslant f_n$, but f_n does not converge uniformly to 0. *Think of the graph of* $1/x$. *Near zero, the function is close to* 1, *not* 0. To justify rigorously, given $\epsilon > 0$, we have

$$\frac{1}{nx+1} > \epsilon \qquad \text{if } x < \frac{1-\epsilon}{n\epsilon}.$$

7.4 Uniform Convergence & Derivatives

To connect uniform convergence with derivatives, we need stronger conditions than what we used previously. First, a non-example showing that insufficient conditions can lead to undesired results:

Example 13.6. Define $f_n(x) := n^{-1/2}\sin(nx)$ for $x \in \mathbb{R}, n \in \mathbb{N}$. It follows that $|f_n(x)| \leqslant n^{-1/2} \to 0$ as $n \to \infty$, so indeed $f_n \to 0$ uniformly as $n \to \infty$. However,

$$f_n'(x) = \sqrt{n}\cos(nx)$$

which approaches ∞ as $n \to \infty$. In particular, the derivatives do not even converge pointwise, not to mention uniformly.
This is because we cannot control derivatives using only f (or lower-order derivatives).

Theorem 13.7: Uniform Convergence & Derivatives

Let $f_n : [a,b] \to \mathbb{R}$ be such that

(1) Each f_n is differentiable, with $f_n' \to g$ uniformly for some $g : [a,b] \to \mathbb{R}$, and

(2) f_n converges (pointwise) at least at one point x_0.

Then, $f_n \to f$ uniformly for some f with $f' = g$. *In particular, the real number $f_n(x_0)$ converges to is $f(x_0)$.* (Note that convergence at at least one point is necessary here – otherwise we can define $f_n(x) \equiv n$, the constant functions, which clearly don't converge to any function, even though their derivatives are uniformly 0.)
To sum up, convergence at one point + uniform convergence of derivatives = preservation of derivatives.
Future reference: Uniform convergence & series

Proof. This proof also uses the "$\epsilon/3$" trick.

Step 1. *We first show that f_n converges uniformly.* Let $\epsilon > 0$ be given. By pointwise convergence at x_0 *and* uniform convergence of derivatives, there exists $N \in \mathbb{N}$ such that

(1) (convergence at x_0) $|f_n(x_0) - f_m(x_0)| < \epsilon/2$ for all $m,n \geqslant N$, and

(2) (uniform convergence of derivatives) $|f_n'(x) - f_m'(x)| < \epsilon/(2(b-a))$ for all $m,n \geqslant N$ and all x.

Then, for all $x \in [a,b]$ and $m,n \geqslant N$, we have

$$\begin{aligned}|f_n(x) - f_m(x)| &\leqslant \underbrace{|(f_n - f_m)(x) - (f_n - f_m)(x_0)|}_{|(f_n - f_m)'(\xi)|\cdot|x - x_0| \text{ by MVT}} + |(f_n - f_m)(x_0)| \\ &< \frac{\epsilon}{2(b-a)}(b-a) + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$

(The existence of ξ is guaranteed since $f_n - f_m$ is differentiable.) Therefore $f_n \to f$ uniformly.

Step 2. *We now show that $f'(x) = g(x)$ for all $x \in [a,b]$.* We pick and fix $x \in [a,b]$ and define two functions

$$\varphi_n(t) := \frac{f_n(t) - f_n(x)}{t - x} \qquad \text{and} \qquad \varphi(t) := \frac{f(t) - f(x)}{t - x} \qquad \text{for } t \in [a,b] - \{x\}.$$

Note that as $t \to x$, $\varphi_n(t) \to f_n'(x)$ and $\varphi(t) \to f'(x)$ (if it exists).

First observe that $\varphi_n \to \varphi$ uniformly on $[a,b] - \{x\}$: for all $t \in [a,b] - \{x\}$,

$$\begin{aligned} |\varphi_n(t) - \varphi_m(t)| &= \frac{1}{|t-x|}|(f_n - f_m)(t) - (f_n - f_m)(x)| \\ [\text{MVT}] &= \frac{1}{|t-x|}|(f_n - f_m)'(\xi)| \cdot |t-x| \end{aligned}$$

which can be made arbitrarily small, as shown in Step 1. Hence φ_n forms a Cauchy sequence with respect to $\|\cdot\|_{\sup}$, and Theorem 12.12 asserts that it converges to some function. Since the numerator converges to $f(t) - f(x)$, φ_n converges to φ.

On the other hand, for all n, $\varphi_n(t) \to f_n'(x)$ as $t \to x$.

Note that now we have a sequence of functions φ_n such that

(1) they converge uniformly to φ, and

(2) there exists a limit point x such that $\varphi_n(t)$ converges (to $f_n'(x)$) as $t \to x$, for all n.

This means that we can invoke Theorem 13.1 and interchange the limits!

$$g(x) = \lim_{n\to\infty} f_n'(x) = \lim_{n\to\infty}\lim_{t\to x} \varphi_n(t) = \lim_{t\to x}\lim_{n\to\infty} \varphi_n(t) = \lim_{t\to x}\varphi(t) = f'(x) \text{ [so it exists!]}$$ □

Remark. With an extra assumption that each f_n' is continuous, we can prove the theorem faster.

7.5 Uniform Convergence & Series

Since series are closely related to sequences, many claims we made on sequences can be translated into series versions.

Theorem 13.8: Uniform Convergence & Series

(1) If $f_n \in C(X)$ and $\sum_{n\geqslant 1} f_n$ converges uniformly (i.e., the partial sums converge uniformly) to some function S, then

$$f_n(x) := \sum_{n\geqslant 1} f_n(x)$$

is continuous (by Corollary 13.2), and Theorem 13.1 applied to $S_m := \sum_{n=1}^{m} f_n$ gives

$$\lim_{y\to x} f(y) = \sum_{n\geqslant 1} \lim_{y\to x} f_n(y).$$

(2) If $f_n \in C^1([0,1])$ [i.e., continuous differentiable], $\sum_{n\geqslant 1} f_n'$ [partial sum of derivatives] converges uniformly, and $\sum_{n\geqslant 1} f_n$ converges *at some* $x_0 \in [a,b]$, then

$$f := \sum_{n\geqslant 1} f_n \text{ converges uniformly,} \qquad f \in C^1([a,b]), \qquad \text{and } f' = \sum_{n\geqslant 1} f_n'.$$

This is basically Theorem 13.7 but applied to the partial sums.

Future reference: Example 13.11, radii of convergence & derivatives, infinite series for e, Weierstraß' Monster

Theorem 13.9: Weierstraß M-Test / Weierstraß Criterion

Consider $f_n : X \to \mathbb{R}$, a sequence of functions. If there exists a convergent series of *real numbers* $\sum_{n\geqslant 1} M_k$ such that $\|f_n(x)\|_{\sup} \leqslant M_n$ for each n then $\sum_{n\geqslant 1} f_n$ converges uniformly (on X, and absolutely as well).

(*Heuristically, if $\sum M_k$ converges and dominates $\sum f_n$, then $\sum f_n$ needs to converge.*)
Future reference: Example 13.11, Theorem 13.16, Weierstraß' Monster

Proof. By Theorem 12.12, $\sum f_n$ converges if and only if the series is Cauchy with respect to $\|\cdot\|_{\sup}$, i.e., for all $\epsilon > 0$, there exists $N > 0$ such that

$$\sup_{x\in X}\left|\sum_{k=n+1}^{m} f_k(x)\right| < \epsilon \qquad \text{for all } m,n \geqslant N.$$

Since

$$\sup_{x\in X}\left|\sum_{k=n+1}^{m} f_k(x)\right| \leqslant \sup_{x\in X} \sum_{k=n+1}^{m} |f_k(x)| \leqslant \sum_{k=n+1}^{m} M_k$$

where the last term can be made $< \epsilon$ for large enough m,n according to the CCC, we are done. ☐

Beginning of April 16, 2021

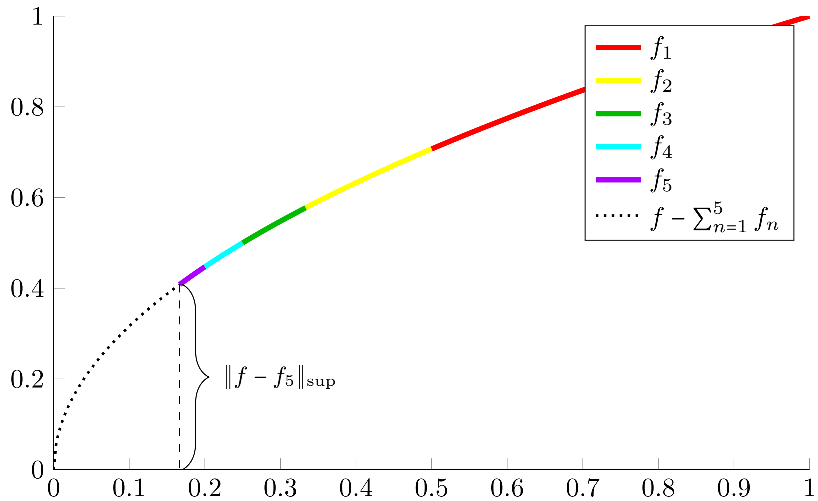
Example 13.10: Converse of Weiwestraß M-Test is False. We will consider an example where $\sum f_n$ converges uniformly to some function f, but there does not exist a convergent $\sum M_k$ dominating the functions. We define the **characteristic function** (also called the **indicator function**, with notation $\mathbf{1}_A$)

$$\chi_A(x) := \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

and consider

$$f_n(x) := \sum_{k=1}^{n} \chi_{[1/(k+1),1/k]}(x)\sqrt{x}.$$

(*The graph roughly looks like the diagram above, but I only plotted the nonzero parts of each function and ignored the zero parts, so it only serves as a visual representation.*)



Clearly $\|f - f_n\|_{\sup} \to 0$, so $f_n \to f \equiv \sqrt{x}$ uniformly. However,

$$\sum_{n=1}^{\infty} \sup_{x\in[0,1]} f_n(x) = \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \text{ diverges by Example 6.8.}$$

Example 13.11. Compute $\displaystyle\sum_{n=1}^{\infty} \frac{n}{2^n}$.

Solution. Notice that this looks almost like a geometric series. In fact, if we define $f_n(x) := x^n/2$, then

$$f_n'(1/2) = \frac{n}{2^{n-2}} \cdot \frac{1}{2} = \frac{n}{2^n}.$$

To connect derivatives with convergence, we want to use Theorem 13.8.2, so we need to verify (1) convergence of f_n at at least one point and (2) uniform convergence of the derivatives f_n'.

(1) Each f_n is clearly differentiable, and in particular $f_n \in C^1([0,3/4])$. (*We chose* $3/4$ *because (i) the geometric series correpsonding to* $x^n/2$ *converges if and only if* $|x| < 1$ *and (ii) our special point* $1/2$ *is contained in* $[0,3/4]$.) The convergence at $x_0 := 0$ is clear, as $\sum 0 = 0$.

(2) To show the uniform convergence of f_n', we invoke the Weierstraß M-test. Since (on $[0,3/4]$)

$$|f_n'(x)| = \frac{nx^{n-1}}{2} \leqslant 10\left(\frac{8}{7}\right)^n \cdot \frac{1}{2} \cdot \left(\frac{3}{4}\right)^{n-1} \leqslant \frac{20}{3}\left(\frac{6}{7}\right)^n := M_n,$$

we see that $\|f_n'\|$ are bounded by a convergent geometric series. (*Note the smart choice of numbers here! The key is that* $(3/4)\cdot(8/7)$ *is still* < 1*, thereby creating a geometric series. The constant* 10 *is to ensure that* $n \leqslant 10(8/7)^n$ *on* $[0,3/4]$.) Therefore $\sum f_n'$ converges uniformly on $[0,3/4]$.

Thus, by Theorem 13.8.2,

$$\sum_{n=1}^{\infty} f_n'(x) = \Big(\sum_{n=1}^{\infty} f_n(x)\Big)' = \left(\frac{1}{2}\cdot\frac{x}{1-x}\right)' = \frac{1}{2(1-x)^2} \qquad \text{for all } x \in [0,3/4].$$

Taking $n := 1/2$, we obtain

$$\sum_{n=1}^{\infty} \frac{n}{2^n} = \sum_{n=1}^{\infty} f_n'(1/2) = 2.$$

7.6 Uniform Convergence & Integration

Theorem 13.12: Uniform Convergence & Riemann-Stieltjes Integrals

Suppose $f_n \to f$ uniformly on $[a, b]$ and each $f_n \in \mathfrak{R}(\alpha)$ on $[a, b]$. Then

$$\int_a^b f \, \mathrm{d}\alpha = \lim_{n\to\infty} \int_a^b f_n \, \mathrm{d}\alpha.$$

(*This tells us that (1) f is R-S integrable* and *(2) the limit of the RHS exists!*)

To sum up, uniform convergence + R-S integrability = preservation of R-S integrability.

Proof. The main idea is that as f_n gets uniformly close to f, the integral also gets close to that of f. Think of Riemann integrals for example, in which case the difference is bounded by $2\|f_n - f\|_{\sup}(b-a)$.

Let $\epsilon > 0$ be given. By uniform convergence, there exists $N > 0$ such that

$$\sup_{x\in[a,b]} |f_n(x) - f(x)| < \epsilon \qquad \text{for all } n \geqslant N.$$

Since

$$0 \leqslant \overline{\int_a^b} f \, \mathrm{d}\alpha - \underline{\int_a^b} f \, \mathrm{d}\alpha \leqslant \int_a^b 2\epsilon \, \mathrm{d}\alpha = 2\epsilon\left(\alpha(b) - \alpha(a)\right),$$

Letting $\epsilon \to 0$, we see that $f \in \mathfrak{R}(\alpha)$.

Also, we claim that the RHS integral converges:

$$\left| \int_a^b f \, \mathrm{d}\alpha - \int_a^b f_n \, \mathrm{d}\alpha \right| \leqslant \int_a^b |f - f_n| \, \mathrm{d}\alpha \leqslant \int_a^b |f - f_n| \, \mathrm{d}\alpha \leqslant \epsilon\left(\alpha(b) - \alpha(a)\right)$$

where the first $\leqslant$ is by Theorem 11.5. Letting $\epsilon \to 0$, we obtain $f_n \to f$ as well, which concludes the proof. □

Corollary 13.13

A direct consequence of the previous theorem: if $f := \sum_{n=1}^{\infty} f_n$ converges uniformly on $[a, b]$ and each f_n is in $\mathfrak{R}(\alpha)$, then

$$\int_a^b f \, \mathrm{d}\alpha = \int_a^b \sum_{n=1}^{\infty} f_n \, \mathrm{d}\alpha = \sum_{n=1}^{\infty} \int_a^b f_n \, \mathrm{d}\alpha,$$

i.e., uniform convergences allows us to interchange integration and summation operators.

Proof: consider the sequence $S_m := \int_a^b \sum_{n=1}^{m} f_n \, \mathrm{d}\alpha$.

7.7 Infinite Taylor Series & Power Series

Previously we have always been analyzing finite Taylor expansions involving a finite sum and a remainder. Now, having discussed the notion of uniform convergence, we are finally ready to study infinite Taylor series, as done (unrigorously) in Calc II!

First, let us restate Example 6.17:

Definition 13.14: Power Series

If $f_n(x) = a_n(x - x_0)^n$ then $\sum_{n=0}^{\infty} f_n = \sum_{n=0}^{\infty} a_n(x - x_0)^n$ is a **power series** centered at x_0.

Theorem 13.15: Cauchy-Hadamard Theorem

A power series converges (pointwise) inside its **interval of convergence** $(x_0 - R, x_0 + R)$, where

$$R := \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|a_n|}}.$$

If the denominator is 0, we set $R = \infty$, and if the denominator is ∞, we set $R = 0$. The power series diverges if $x \in (-\infty, x_0 - R) \cup (x_0 + R, \infty)$. If $x = x_0 - R$ or $x_0 + R$, the theorem is indeterminate.

For simplicity we will let $x_0 := 0$. (The general case follows right away.)

Theorem 13.16

Suppose $\sum a_n x^n$ converges for $x = z \neq 0$. Then:

(1) it converges absolutely for $x \in (-|z|, |z|)$, and

(2) it converges uniformly on $[-a, a]$ for any $a \in (0, |z|)$.

Proof. (1) We want to show that $\sum_{n=0}^{\infty} |a_n x^n|$ converges. Indeed, notice that

$$\sum_{n=0}^{\infty} |a_n x^n| = \sum_{n=0}^{\infty} |a_n z^n| \cdot |x/z|^n \leqslant M \sum_{n=0}^{\infty} \underbrace{|x/z|}_{<1}^n \qquad \text{for some } M > 0.$$

(Since $\sum a_n x^n$ converges, the tail must converge to zero, i.e., $a_n z^n \to 0$, so the entire sequence $(a_n x^n)_{n\geqslant 1}$ is bounded by Lemma 4.12 by M, say.) Since the RHS is a geometric series with $|x/z| < 1$, we are done.

(2) For this one, we use the Weierstraß M-test. For all $x \in [-a, a]$, we have

$$|a_n x^n| \leqslant M|x/z|^n \leqslant M \frac{a^n}{|z|^n} =: A_n$$

where $\sum A_n$ is a convergent geometric series. The claim then follows.

□

Beginning of April 19, 2021

Corollary 14.1: Radii of Convergence & Derivatives

Suppose $f(x) := \sum_{n=0}^{\infty} a_n x^n$ has a radius of convergence $R \in (0, \infty]$. Then:

(1) $f \in C((-R, R))$.

(2) $\sum_{n=1}^{\infty} (a_n x^n)'$ has the same radius of convergence R, and it converges uniformly to f' on every compact interval $[-a, a] \subset (-R, R)$. Furthermore, $f' \in C((-R, R))$.

(3) Repeating (1) and (2), for all $k \in \mathbb{N}$, $\sum_{n=k}^{\infty} (a_n x^n)^{(k)}$ has the same radius of convergence and converges uniformly to $f^{(k)}$ on any $[-a, a] \subset (-R, R)$. Also, $f^{(k)} \in C((-R, R))$.

Put informally, power series are infinitely differentiable, i.e., smooth, inside its interval of convergence.

Future reference: Infinite series for e, Taylor's Theorem (summary)

Proof. (1) *The key is to show that f is continuous on $[-R + \epsilon, R - \epsilon]$ for any $\epsilon > 0$.* If $a \in (0, R)$, then we can pick any $z \in (a, R)$. Since z is still in the radius of convergence, $\sum a_n x^n$ converges, so by the previous theorem (part 2), $[-a, a] \subset [z, z]$ implies uniform convergence on $[-a, a]$. Since a can be arbitrarily close to R, we conclude that the series converges uniformly to f on $(-R, R)$. Then since uniform convergence preserves continuity of series, we conclude that $f \in C((-R, R))$.

(2) By definition

$$R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|a_n|}}.$$

We now compute R', the radius of convergence of the series of derivatives. Since $(a_n x^n)' = n a_n x^{n-1}$,

$$R' = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|n a_n|}} = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{n} \sqrt[n]{|a_n|}}.$$

This is because $\sqrt[n]{n}$ as $n^{1/n} = \exp((1/n)\log(n))$[5], L'Hôpital's rule suggests that the exponent $(1/n)\log(n)$ satisfies

$$\lim_{n\to\infty} \frac{\log n}{n} \overset{H}{=} \lim_{n\to\infty} \frac{1}{n} = 0,$$

so $\lim_{n\to\infty} n^{1/n} = \exp(0) = 1$. This limit is finite, so we have

$$\limsup_{n\to\infty} \sqrt[n]{n} \sqrt[n]{|a_n|} = \limsup_{n\to\infty} \sqrt[n]{n} \limsup_{n\to\infty} \sqrt[n]{|a_n|} = \limsup_{n\to\infty} \sqrt[n]{|a_n|}.$$

(For proof, see this excellent answer on Math SE.) Therefore R' is really just R. Using (1) again, we see that f' is also continuous on $(-R, R)$. (f' is the limit since (1) the derivatives converge uniformly and (2) the series of the (original) power series converge at least at the origin. Then use Theorem 13.8.)

(3) Induction! □

[5] Since many steps in the lecture simply said "cf. problem sheet", I decided to write them for completeness. However, I am assuming some basic properties of the exponential and logarithmic functions, e.g., $a^b = \exp(b \log a)$ which are not entirely trivial.

Definition 14.2: Analytic Functions

Let I be an open interval. We say $f : I \to \mathbb{R}$ is (real) **analytic** at $x_0 \in I$ if

There exists $R > 0$ and $(a_n)_{n \geqslant 0} \subset \mathbb{R}$ such that $f(x) = \sum_{n=0}^{\infty} a_n x^n$ for $x \in (x_0 - R, x_0 + R) \subset I$.

Note that $R > 0$*, since if* $R = 0$, $f(x_0) = 1 \cdot f(x_0) + 0 + \ldots$ *which holds trivially.* We say f is **analytic in** $\boldsymbol{I}$ if f is analytic at every $x_0 \in I$.

Corollary 14.3: Analytic ⇒ Smooth

If f is analytic at x_0, then $f \in C^\infty(I)$ for some open interval I containing x_0. In particular, this means that if f is analytic in I then f is smooth on I — being analytic is stronger than being smooth!

Proof. Within the radii of convergence of each point, we can take derivatives, and since the radii does not change, we can do this as many times as we want. □

Example 14.4: Euler Number as an Infinite Taylor Series. Finally, we can show that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ [!]

Proof. Notice that $n!$ easily outgrows x^n, so the radius of convergence $R = \infty$. It follows that the series

$$f(x) := \sum_{n=0}^{\infty} \frac{x^n}{n!} \in C^\infty(\mathbb{R})$$

by Corollary 14.1. We will now show that $f(x)$ is precisely e^x. Notice that the series converge to f at $x = 0$. Also, for any x, since it is in the radius of convergence (of the series and therefore of its derivatives, as they share the same R), the derivatives converge uniformly. Therefore by Theorem 13.8.2 f' is equal to the infinite sum of derivatives. This gives

$$f'(x) = \sum_{n=1}^{\infty} \frac{nx^{n-1}}{n!} = \sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} = \sum_{n=0}^{\infty} \frac{x^n}{n!} = f(x).$$

To show that $f(x) = e^x$, it suffices to show that $f(x)/e^x = 1$ for all x. Notice that we have

$$(f(x)e^{-x})' = f'(x)e^{-x} - f(x)e^{-x} = 0$$

so $f(x)/e^x$ is a constant C. Taking $x := 0$, we get $f(0)e^0 = C$; since $f(0) = e^0 = 1$, $C = 1$, and we are done. □

Summary of Infinite Taylor Expansions at x_0

We now briefly summarize our discussions on Taylor expansions so far.

(1) Consider $f \in C^n(I)$ and let $x_0 \in I$. For $x \in I$, the finite Taylor expansion gives

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + \begin{cases} \dfrac{f^{(n)}(\xi)}{n!}(x-x_0)^n & \text{Lagrange form, Theorem 10.4} \\ o\big((x-x_0)^{n-1}\big) & \text{Peano form, Theorem 10.7} \\ \displaystyle\int_{x_0}^{x} f^{(n)}(t)\frac{(x-t)^{n-1}}{(n-1)!}\,\mathrm{d}t & \text{integral form, Theorem 12.3.} \end{cases}$$

(2) If, in addition to (1), we also have $f \in C^\infty(I)$, then we define

$$R := \frac{1}{\limsup_{n\to\infty} \sqrt[n]{f^{(n)}(x_0)/n!}}.$$

If $R > 0$, then the infinite Taylor series $\tilde{f}(x) := \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$ converges on $(x_0 - R, x_0 + R)$. More importantly, it converges to a C^∞ function $\tilde{f}(x)$. (In fact $\tilde{f}$ is analytic by Corollary 14.1.)

(3) Directly from construction we see $\tilde{f}(x_0) = f(x_0)$ and $\tilde{f}^{(k)}(x_0) = f^{(k)}(x_0)$ for all $k \in \mathbb{N}$. However, *there is nothing more we can assert*. If we want these equations to hold for $x \neq x_0$, we need to repeat (1) and (2) at x, not x_0. This is because not all smooth functions are analytic. See the example below.

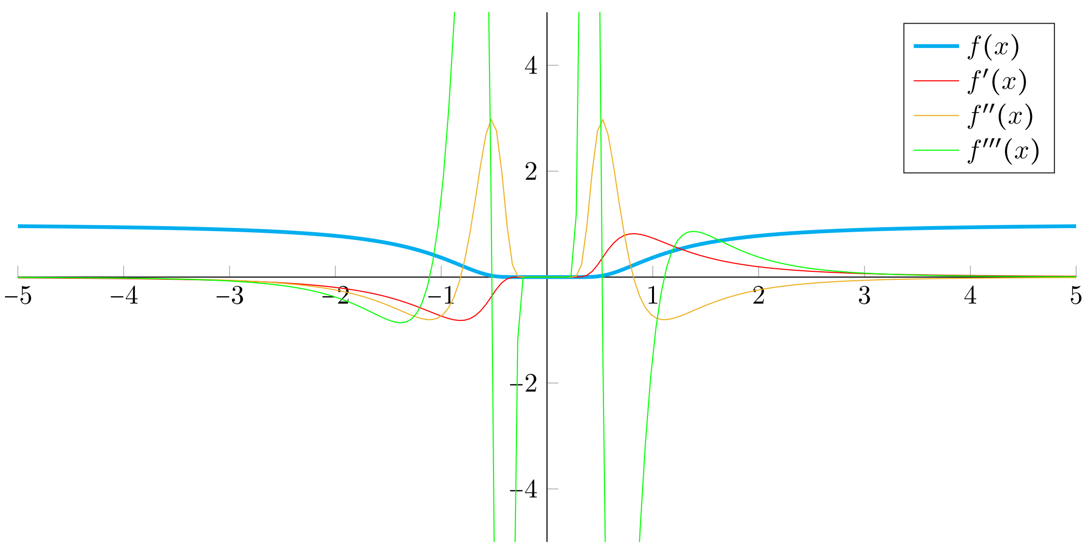
Example 14.5: A Smooth but Non-analytic Function. Consider the **bump function**

$$f(x) := \begin{cases} \exp(-1/x^2) & x \neq 0 \\ 0 & x = 0. \end{cases}$$

Using the quotient-limit definition of derivative and L'Hôpital's rule, all derivatives of f are 0, i.e.,

$$f(0) = f'(0) = \ldots = f^{(n)}(0) = \ldots = 0[!]$$

Therefore the infinite Taylor expansion gives the zero function, but clearly f is not the zero function! This is because f is not analytic at 0 (but it is everywhere else).



Chapter 8

Some Selected Topics

8.1 ODEs & Banach Contraction Principle

We are now equipped with the knowledge for a sneak peak of **ordinary differential equations**, ODEs. We are given a function f of *time* t, and we are asked to solve for another function x of t, subject to the condition that the *rate of change* of x is precisely described by $f(t)$. Also, we are given an initial condition.

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}x(t) = f(x(t)) \\ x(0) = x_0 \end{cases} \tag{Eq.14.1}$$

A most basic example is $\frac{\mathrm{d}}{\mathrm{d}t}x(t) = 2x$ and $x(0) = 1$. Just by inspection we see that $x(t)$ is $x^2 + C$, the antiderivative of x. The initial condition implies $0^2 + C = 1$, so $C = 1$ and our solution is $x(t) = x^2 + 1$.

ODEs are ubiquitous in applications such as chemistry, economics, physics, and so on. They tell us another aspect of these functions – the *rate of change* – which, in some situations, are more useful than the functions themselves. For example, consider a simplified biological model of a forest that only consists of one species of predator and one of producer, say wolves and deers, for example. If there are too many wolves, they will overhunt the deers, resulting in a *decline* of deer population. But then since there are not enough preys, the wolf population starves, resulting in a *decrease* in their population too. The deer population would then *increase* as a result of decreasing number of predators, and subsequently the wolf population would *increase* as food sources once become ample. The cycle goes on, and it is much more convenient to look at the *differential* equations modelling these phases than the equations of populations if we want to know what phase the forest is at.

Questions in the theory of ODEs. What kind of functions f and x do we require? Do we want C^1? Do we want D^1? When do the solutions exist? When are they locally or globally unique[1]? The list goes on.

Definition 14.6: Integral Solutions to ODEs

Let $f \in C(\mathbb{R})$ [a sufficient but not necessary condition]. We say that $x \in C([0,T])$ satisfies Equation 14.1 if

$$x(t) = x_0 + \int_0^x f(x(s))\,\mathrm{d}x \qquad \text{for all } t \in [0,T]. \tag{Eq.14.2}$$

[*Note that f is continuous so the integral is well-defined. Furthermore, $x \in C^1([0,T])$.*]

Future reference: Proof of the Picard-Lindelöf Theorem

[1]The Picard-Lindelöf Theorem analyzes existence and uniqueness under some conditions. This is covered in 425b. Pure magic!

Beginning of April 21, 2021

Theorem 14.7: Picard-Lindelöf Theorem

If f is Lipschitz with $|f(x) - f(y)| \leqslant L|x - y|$ for $L > 0$ and if $T < 1/L$, then there exists a unique $x \in C([0,T])$ satisfying the above Equation 14.2. *In fact this holds for all T.* To be prove later.

Theorem 14.8: Banach Contraction Principle / Banach Fixed-Point Theorem (Banach FPT)

Let (X, d) be a complete metric space, let c be constant satisfying $c < 1$, and let $S : X \to X$ be such that

$$d(S(x), S(y)) \leqslant c \cdot d(x, y) \qquad \text{for all } x, y \in X.$$

(Such S is called a **(strong) contraction**; if we only require $d(S(x), S(y)) < d(x, y)$, such S is called a **weak contraction**. *I will briefly mention an example at the end of this section.*) Then there exists a *unique* fixed point $z \in X$ of S, i.e., $S(z) = z$ for some unique z.
Future reference: Devil's Staircase

Proof. The main idea is to construct a Cauchy sequence using this contractive property of S and then use completeness of X to show that the sequence converges. Then we will (magically) show that the limit is the fixed point!
For existence, we start by picking any $x_0 \in X$, and for $n \geqslant 1$ we define iteratively $x_n := S(x_{n-1})$ [i.e., the sequence starts off by $(x_0, S(x_0), S(S(x_0)), \ldots)$]. We will show that this sequence is Cauchy.
Let $\epsilon > 0$ be given. For $n \geqslant m$, we can repeatedly use $d(S(x), S(y)) \leqslant c \cdot d(x, y)$ and reduce every single term consisting of $d(x_k, x_{k+1})$ to some power of c times $d(x_0, x_1)$:

$$\begin{aligned} d(x_n, x_m) &\leqslant d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \ldots + d(x_{n-1}, x_n) \\ &= d(S(x_{m-1}), S(x_m)) + d(S(x_m) + S(x_{m+1})) + \ldots + d(S(x_{n-2}), S(x_{n-1})) \\ &\leqslant c \cdot d(x_{m-1}, x_m) + c \cdot d(x_m, x_{m+1}) + \ldots + c \cdot d(x_{n-2}, x_{n-1}) \\ &\ \ \vdots \\ &\leqslant c^m \cdot d(x_0, x_1) + c^m \cdot d(x_1, x_2) + \ldots + c^m \cdot d(x_{n-m-1}, x_{n-m}) \\ &\ \ \vdots \\ &\leqslant c^m \cdot d(x_0, x_1) + c^{m+1} \cdot d(x_0, x_1) + \ldots + c^{n-1} \cdot d(x_0, x_1). \end{aligned}$$

This is a geometric series, and the sum can can be bounded:

$$d(x_n, x_m) \leqslant \sum_{k=m}^{n-1} c^k \cdot d(x_0, x_1) \leqslant \sum_{k=m}^{\infty} c^k \cdot d(x_0, x_1) = d(x_0, x_1) \cdot \frac{c^m}{1 - c}.$$

Note that the RHS $\to 0$ as $n \to \infty$, so by picking sufficiently large N (in particular, large enough N satisfying $c^N \cdot d(x_0, x_1)/(1 - c) < \epsilon$), we obtain $d(x_n, x_m) < \epsilon$ whenever $m > n \geqslant N$. Cauchy-ness!
Thus, since (X, d) is assumed to be complete, $(x_n)_{n \geqslant 1}$ converges to some $z \in X$. Since

$$\begin{aligned} d(S(z), z) &\leqslant d(S(z), S(x_n)) + d(S(x_n), x_{n+1}) + d(x_{n+1}, z) \\ &\leqslant c \cdot \underbrace{d(z, x_n)}_{\to 0} + \underbrace{d(x_{n+1}, x_{n+1})}_{\to 0} + \underbrace{d(x_{n+1}, z)}_{\to 0} \end{aligned}$$

we see that $d(S(z), z)$ can be bounded above by any positive number [*letting $n \to \infty$, the RHS becomes arbitrarily small, whereas the $\leqslant$ still needs to hold*], so $d(S(z), z) = 0$, as claimed. [2]

[2] Another magical proof that I feel obliged to introduce: $S(z) = S(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} S(x_n) = \lim_{n\to\infty} S(x_{n+1}) = \lim_{\tilde{n}-1\to\infty} S(x_{\tilde{n}}) = z$, where the

Uniqueness is obvious: if $S(z_1) = z_1$ and $S(z_2 = z_2)$, then

$$d(z_1, z_2) = d(S(z_1), S(z_2)) \leqslant c \cdot d(z_1, z_2),$$

so $d(z_1, z_2)(1-c) \leqslant 0$. Since $1 - c \neq 0$ we must have $d(z_1, z_2) = 0$, i.e., $z_1 = z_2$. *After all, if they are apart, S sends them closer to each other, but how can fixed points move under S?* □

Back to proving the previous theorem:

Proof of Picard-Lindelöf Theorem. Given a *function* x of t, we define a new *function* $S(x)$, often times written as S_x [because it is a function of time and we need to save the parentheses for $S_x(t)$] by

$$S_x(t) := x_0 + \int_0^t f(x(s))\, \mathrm{d}s.$$

(Un)surprisingly, this theorem can be proven by the Banach FCT. Assuming S is a strong contraction, if x is a fixed point of S, i.e., $x = S_x$ or equivalently $x(t) = S_x(t)$ for all $t \in [0, T]$, then it is a solution to Equation 14.2, and since Banach FPT asserts the uniqueness of such fixed point, the unique existence of x follows directly.

We shall now show that the conditions for Banach FPT indeed apply. Right away we know that S is a mapping from $C([0, T])$ into itself, as $x \in C([0, T])$ implies that S_x is also continuous, by Theorem 11.10. Therefore, our work reduces to verifying that S is a strong contraction.

For all $t \in [0, T]$ and all $x, y \in C([0, T])$ [continuous functions, inputs of S], we have

$$\begin{aligned}
|S_x(t) - S_y(t)| &= \left| \left(x_0 + \int_0^t f(x(s))\, \mathrm{d}s\right) - \left(x_0 + \int_0^t f(y(s))\, \mathrm{d}s\right) \right| \\
&= \left| \int_0^t [f(x(s)) - f(y(s))]\, \mathrm{d}s \right| && \text{(by Lemma 11.4.1)} \\
&\leqslant \int_0^t \left| f(x(s)) - f(y(s)) \right| \mathrm{d}s && \text{(by Theorem 11.5)} \\
&\leqslant \int_0^t L|x(s) - y(s)|\, \mathrm{d}s && \text{(by Lipschitz assumption)} \\
&\leqslant \int_0^t L\|x - y\|_{\sup}\, \mathrm{d}s = \|x - y\|_{\sup} \cdot Lt \leqslant \|x - y\|_{\sup} \cdot LT. && \text{(by Lemma 11.4.2 \& .6)}
\end{aligned}$$

Since we have assumed int he first place that $T < 1/L$, $CL < 1$, and thus S is a strong contraction, as claimed. This completes the proof. □

Weak Contraction ⇏ Banach FPT

I would like to end this section by supplementing the discussion of contraction with an example showing that <u>*Banach FPT may not work for weak contractions*</u>*. To come up with a weak contraction, we consider a mapping a real function f such that*

(1) $|f(x) - f(y)| < |x - y|$ *for all $x \neq y$, but*

(2) *the quotient $|f(x) - f(y)|/|x - y|$ can be made arbitrarily close to 1 for certain x, y.*

Since the quotient resembles the derivative, one natural construction is to design a function $f(x)$ whose derivative never reaches 1 but nevertheless approaches 1 as $x \to \infty$. The easiest example is of course setting the derivative to $1 - 1/x$. The antiderivative is $x - \log(x) + C$, and if we define

$$f : (1, \infty) \to (1, \infty) \qquad \textit{by} \qquad f(x) := x - \log(x),$$

second = is because (S Lipschitz ⇒ S continuous) and the fourth by a "change of dummy variable" $\tilde{n} := n - 1$.

we obtain a weak contraction, but since $\log(x) \neq 0$ *on* $(1, \infty)$*, it is impossible that* $x = x - \log(x)$*, so* f *does not admit a fixed point. Another example is an antiderivative of* 1 *minus the* ***Signoid function*** *(a frequent visitor to my 425a),*

$$f'(x) := 1 - \frac{1}{e + e^{-x}} \implies f(x) = x - \ln(e^x + 1) \qquad \textit{on } (-\infty, \infty).$$

8.2 Equicontinuity & the Arzelá-Ascoli Theorem

Recall that we demonstrated just how nice and useful compactness can be, and we have analyzed compact sets in Euclidean spaces (intervals, boxes, etc). Now, moving to function spaces (e.g., $C([a,b])$), can we still identify compact sets?

Definition 14.9: Equicontinuity

Let $\mathcal{F}$ be a collection of functions $f : X \to \mathbb{R}$. (It is customary to say that $\mathcal{F}$ is a **family** of functions.) We say $\mathcal{F}$ is **equicontinuous** if they are "equally continuous":

> Given $\epsilon > 0$, there exists $\delta > 0$ such that the ϵ-δ continuity condition holds for all $f \in \mathcal{F}$, i.e., for all $x, y \in X$ and all $f \in \mathcal{F}$, if $d(x,y) < \delta$ then $|f(x) - f(y)| < \epsilon$.

Note that each f is already uniformly continuous given the condition above. The extra assumption that δ works for all $f \in \mathcal{F}$ makes the family, in some sense, uniformly uniformly continuous.

Example 14.10. If $\mathcal{F}$ consists of Lipschitz functions with Lipschitz constant L, then they are equicontinuous. (For $\epsilon > 0$ simply take $\delta := \epsilon/L$).

Theorem 14.11: Arzelá-Ascoli Theorem

If K is compact and if $\mathcal{F} \subset C(K)$ (a collection of continuous functions on K) is **precompact** (i.e., the closure $\overline{\mathcal{F}}$ of $\mathcal{F}$ is compact; the closure is taken with respect to $\|\cdot\|_{\sup}$, i.e., $f_n \to f$ if $\|f_n - f\|_{\sup} \to 0$) if

(1) $\mathcal{F}$ is equicontinuous, and

(2) $\mathcal{F}$ is bounded, i.e., there exists $M > 0$ such that $\sup_{f \in \mathcal{F}} \|f\|_{\sup} \leqslant M$, one bound for all f.

For example, if we let $\mathcal{F}$ to be the collection of constant functions $1/n$, $n \in \mathbb{N}$, then it safiesfies (1) and (2) but is not compact: the zero function is in $\overline{\mathcal{F}} - \mathcal{F}$.

We will prove this remarkable theorem later. For now, we will first see an application.

Remark. The Arzelá-Ascoli Theorem characterizes compact sets of $C(K)$. In particular, it says when we can extract a uniformly convergent sequence in $C([a,b])$ (recall that convergence in $\|\cdot\|_{\sup}$ is the same as uniform convergence of functions by Theorem 12.12). This is similar to how the Heine-Borel Theorem characterizes compact sets in $\mathbb{R}^n$.

Beginning of April 23, 2021

Example 14.12: Application: Calculus of Variations. *The **calculus of variations** is a field of mathematical analysis [...] that finds maxima and minima of **functionals** (real-valued functions).*[3]
Given a continuous function $f \in C^1([0,1])$, we define

$$I[f] := \frac{1}{2}\int_0^1 \left(\underbrace{(f'(x))^2}_{\substack{\text{energy of} \\ \text{the system}}} + \underbrace{V(f(x))}_{\text{potential}} \right) \mathrm{d}x$$

where $V(y) \geqslant 0$ for all $y \in \mathbb{R}$.
(*For example, think of a heavy, elastic rope hanging on two endpoints. The further it drops down, the higher the rope's elastic energy is. In the mean time, as it drops down, the rope has a lower gravitational potential energy. The system (rope) would always want to minimize its energy, so it needs to find a minimizer of* $I[f]$.)
Given a subset $\Gamma \subset C^1([0,1])$, can we minimize $I[f]$ where $f \in \Gamma$? In other words, is there a $u \in \Gamma$ that attains the infimum, i.e.,

$$I[u] = \inf_{f \in \Gamma} I[f]?$$

A **minimization** problem like this is often (but not always) solved by using compactness.

A baby example of minimization: suppose we are given $g \in C([0,1])$*. How can we find a minimizer* $u \in [0,1]$ *satisfying* $g(u) = \inf_{x \in [0,1]} g(x)$*?*

Solution 1. Use compactness and invoke Theorem 8.5 Solution 2. Note that the infimum is a limit point of $g([0,1])$*.Therefore we can pick a sequence* $(x_n)_{n \geqslant 1} \subset [0,1]$ *such that* $g(x_n)$ *converges to* $\inf g$*. By the Bolzano-Weierstraß Theorem, we can extract a subsequence* $x_{n_k} \to u$ *for some* $u \in [0,1]$*. Since* g *is continuous,* $g(u)$ *is the limit of a subsequence of* $(g(x_n))_{n \geqslant 1}$*, so the mother sequence (i.e.,* $(g(x_n))_{n \geqslant 1}$ *itself) must also converge to the same limit (because limits are unique and the mother sequence converges). Therefore,*

$$g(u) = \lim_{k \to \infty} g(x_{n_k}) = \inf_{x \in [0,1]} g(x).$$

(We have again used compactness to find a minimizer.) Clearly, for some g*, the minimizer needs not to be unique.*

Solution. We consider a special case where $K > 0$ and $\Gamma := \{f \in C^1([0,1]) : |f(x)| \leqslant K \text{ for all } x\}$. Let $(f_n) \subset \Gamma$ be such that

$$I[f_n] \to \inf_{f \in \Gamma} I[f].$$

(In particular $I[f_n]$ is a sequence of real numbers.) Since $V(y) \geqslant 0$, rewriting the original equation of $I[f]$ gives

$$\int_0^1 (f_n')^2 = 2I[f_n] - \int_0^1 V \leqslant 2I[f_n] \leqslant C \qquad \text{for some } C > 0,$$

where the last $\leqslant$ is because $I[f_n]$ is assumed to be convergent and convergent sequences are bounded.
Also, for all n and all $1 \geqslant t > s \geqslant 0$, by FTC part 2 and Cauchy-Schwarz inequality of integral form (*this is because* $\left(\int_a^b |g(x)|^2 \, \mathrm{d}x\right)^{1/2}$ *define a so-called 2-norm of a function and* $\int_a^b g(x)h(x) \, \mathrm{d}x$ *defines an inner product between functions* g *and* h), we have

$$\begin{aligned} |f_n(t) - f_n(s)| &= \left| \int_s^t f_n'(x) \, \mathrm{d}x \right| \leqslant \int_s^t |f_n'(x)| \, \mathrm{d}x \\ &\leqslant \left(\int_s^t |f_n'(x)|^2 \, \mathrm{d}x \right)^{1/2} \left(\int_s^t 1 \, \mathrm{d}s \right)^{1/2} \leqslant \sqrt{C}\sqrt{t-s}. \end{aligned}$$

This shows that $(f_n)_{n \geqslant 1}$ is equicontinuous (for a given ϵ, pick $\delta := \epsilon^2/C$). Also, $(f_n)_{n \geqslant 1}$ is bounded (because Γ consists only of bounded functions – continuous functions on compact domain). By Arzelá-Ascoli (which we are about to show right away), there exists a subsequence $(f_{n_k})_{k \geqslant 1}$ that converges to f in $\|\cdot\|_{\sup}$. It is guaranteed that

[3] From Wikipedia.

$f \in C([0,1])$, but there is no guarantee that $f \in C^1([0,1])$ (*Arzelá-Ascoli gives precompactness, not compactness, if we do not have further assumption*). Therefore we have obtained a *candidate* f for minimizer. If in addition $f \in C^1$ then it is the minimizer.

Proof of the Arzelá-Ascoli Theorem. Since part of the proof is covered in discussion sessions and omitted by lectures, I will try to reconstruct them based on the version I learned. They should be mostly the same. For a remarkable theorem like this, the proof is long and is therefore broken into smaller steps.

Step 1. *We first show that if K is compact then it admits a countable dense subset.*

Let $\delta_n := 1/n$ and so $(\delta_n)_{n\geqslant 1}$ forms a sequence. For each δ_n, we claim that there exist *finitely many* balls of radii $1/n$ that covers K. To see this, first notice that

$$K = \{x : x \in K\} \subset \bigcup_{x\in K} B_{1/n}(x),$$

so the RHS forms an open cover of K. By covering compactness it admits a finite cover, so

$$K \subset \bigcup_{i=1}^{m} B_{1/n}(x_i)$$

for some finite m. It follows that K can be covered by finitely many $1/n$-balls. We define $E_n := \{x_1, ..., x_m\}$, the corresponding centers of these $1/n$-balls.

Letting $n \in \mathbb{N}$ vary, we obtain different E_n's, but all of them are finite. Taking the countable union of these finite sets, we obtain a countable collection of points

$$E := \bigcup_{i=1}^{\infty} E_n$$

and we claim that E is dense in K. Indeed, if $x \in K$ then x is contained in some arbitrarily small open balls (since the balls of this radius cover K), so x is arbitrarily close to the ball's center, which is a point in E. This completes Step 1.

Step 2. *If $(f_n)_{n\geqslant 1}$ is a bounded sequence with respect to ($\|\cdot\|_{\sup}$), and $E \subset K$ is dense, then there exists a subsequence $(f_{n_k})_{k\geqslant 1}$ that converges pointwise at <u>every</u> $x \in E$.* The proof is done by using the Bolzano-Weiwestraß Theorem along with *Cantor's diagonalization argument*. We first let $\{e_i\}_{i\geqslant 1}$ be an enumeration of the countable dense subset E, and we look at the *real-valued* sequence $(f_n(e_1))_{n\geqslant 1}$. Since $\|f_n\|_{\sup}$ are (uniformly) bounded [i.e., one bound for all f_n], so is the sequence evaluated at e_1. Thus, by Bolzano-Weierstraß there exists a subsequence $(f_{1,n}(e_1))$ that converges to some $y_1 \in \mathbb{R}$.

Then, we look at the subsequence $f_{1,n}$ and evaluate them at e_2, which forms another bounded real-valued sequence. By Bolzano-Weierstraß again, there exists a subsequence of $f_{1,n}$, or a sub-subsequence of f_n, which we call $f_{2,n}$, such that $(f_{2,n}(e_2))$ converges to some $y_2 \in \mathbb{R}$. Notice that $f_{2,n}$ not only converges at e_2 but also e_1[!]

Doing this iteratively, for each sub-sub-...-subsequence $f_{k-1,n}$, we can further extract $f_{k,n}$ that converges at $e_1, ..., e_k$.

Finally, we define a *diagonal* sequence (of functions) $(g_n)_{n\geqslant 1}$ such that g_1 is the first element(function) in $f_{1,n}$, g_2 is the second element in $f_{2,n}$, and g_k is the k^{th} element in $f_{k,n}$. It follows that (g_n) is a subsequence of all $(f_{i,n})$'s, so it converges at e_1 as $f_{1,n}$ (and all other sequences) does, at e_2 as $f_{2,n}$ does, and at e_k as $f_{k,n}$ does. This means that $(g_n)_{n\geqslant 1}$ converges at all points of E, completing Step 2.

Step 3. *This step known as the **Arzelá-Ascoli Propagation Theorem**. (I will fully state it.)*

> If (g_n)[4] is a sequence of equicontinuous functions on X, if $E \subset X$ is dense, and if g_n converges *pointwise* on E, then the pointwise limit on E can be *propagated* to some function on all of X, to which (g_n) converges *uniformly*!

Proof of the A-A Propagation Theorem (Step 3). By Theorem 12.12 it suffices to show that (g_n) forms a Cauchy sequence with respect to $\|\cdot\|_{\sup}$. Let $\epsilon > 0$. By equicontinuity, there exists $\delta > 0$ such that

$$|g_n(x) - g_n(y)| < \frac{\epsilon}{3} \qquad \text{for all } n \in \mathbb{N} \text{ if } d(x,y) < \delta \text{ (and } x, y \in K). \tag{1}$$

Also, by Step 1, we are able to cover K by *finitely many* δ-balls, so let $x_1, \ldots, x_k \in E$ be such that

$$K \subset \bigcup_{i=1}^{m} B_\delta(x_i).$$

Now, since $(g_n(x_1))_{n\geqslant 1}$ converges, it in particular forms a Cauchy sequence, so there exists $N_1 \in \mathbb{N}$ such that $|g_n(x_1) - g_m(x_1)| < \epsilon/3$ whenever $m, n \geqslant N_1$. Likewise, there exists $N_2, N_3, \ldots, N_k$ satisfying the Cauchy condition respectively. Since there are only finitely many N_i's, $N := \max\{N_1, \ldots, N_k\}$ is well-defined and *finite*. More importantly, we have

$$|g_n(x_i) - g_m(x_i)| < \frac{\epsilon}{3} \qquad \text{for all } m, n \geqslant N \text{ and } x_i \in \{x_1, \ldots, x_k\}. \tag{2}$$

Finally, using (1) twice and (2) once, for $m, n \geqslant N$ and any $x \in X$, we obtain

$$|f_m(x) - f_n(x)| \leqslant |f_m(x) - f_m(x_i)| + |f_m(x_i) - f_n(x_i)| + |f_n(x_i) - f_n(x)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,$$

where x_i is chosen from $\{x_1, \ldots, x_k\}$ such that $d(x, x_i) < \delta$. Therefore, for large m, n, $\|f_m - f_n\|_{\sup} \leqslant \epsilon$, completing both the proof of A-A Propagation Theorem and the A-A Theorem itself! □

[4] I chose to write (g_n) instead of (f_n) to stress the fact that these g_n's come from the previous step and to avoid potential confusion.

8.3 The Weierstraß Approximation Theorem

Previously, we have shown that $\mathbb{Q}$ is dense in $\mathbb{R}$; by the same token we can easily show that $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$. Now that we have shifted our focus to the space of continuous functions, a natural question arises –

Is there a dense and countable subset of $C([a,b])$, in particular $C([0,1])$?

Answer: *yes!*

Theorem 14.13: Weierstraß Approximation Theorem

The set of all *polynomials* is dense in $C([0,1])$. In particular, given $f \in C([0,1])$, the **Bernstein polynomials**

$$P_n(x) := \sum_{k=0}^{n} f(k/n)\binom{n}{k}x^k(1-x)^{n-k}$$

approximate f uniformly on $[0,1]$.[5]

Remark 1. This is a *global* approximation on the entire interval $[a,b]$, stronger than the Taylor expansion, which only gives *local* approximation near some x_0.

Remark 2: Probabilistic Interpretation of the Bernstein Polynomials. *Before moving to the proof, I would like to give a probabilistic interpretation of why and how these seemingly arbitrary p_n's approximate f. Let $f \in C([0,1])$ be given.*

(1) *Suppose we have a loaded coin with probability $p \in [0,1]$ of showing up heads.*

(2) *Suppose we have a game in which we toss this coin n times. Obviously, the head will show up k times, where k can be any integer between 0 and n. Whatever k is, we will be awarded $f(k/n)$ amount of money. [For example, if all were tails, we earn $f(0/k) = f(0)$ amount of money; if all were heads, we earn $f(k/k) = f(1)$.]*

(3) *In a game of tossing n coins, what would the **expected value** $E_n(p)$ be? In other words, how much money, on average, are we going to earn? A basic formula in probability says we need to sum up the weighted average,*

$$E_n(p) = \sum_{k=0}^{n}(\textit{money earned with } k \textit{ heads}) \cdot (\textit{probability of getting exactly } k \textit{ heads}).$$

(4) *Translating the above expression into mathematical equation, the probability of getting precisely k heads is*

$$\binom{n}{k}p^k(1-p)^{n-k}.$$

Thus,

$$E_n(p) = \sum_{k=0}^{n} f(k/n)\binom{n}{k}p^k(1-p)^{n-k}. \tag{1}$$

(5) *What happens if n is really large? We digress a bit to think of a very basic example. If we have a fair coin and we only toss it once, we either get head or tail (assuming it doesn't land on its side!), so the "head rate" is either 100% or 0%, very far from the 50% as suggested by the name "fair". However, if we toss this fair coin 100 times or a million times, it is completely understandable that the "head rate" is going to be much, much closer to 50%. This is a heuristic example of the **Law of Large Numbers**, LLN.*

[5] The notation $\binom{n}{k}$ refers to the binomial coefficient $n!/(k!(n-k)!)$.

Back to our loaded coin – when n is large, it becomes increasingly likely that the head will come up around np times (number of tosses · probability). Thus, it becomes increasingly likely that we will be awarded

$$f(np/n) = f(p) \text{ amount of money.}$$

This means that $E_n(p)$ approaches $f(p)$ as $n \to \infty$[!]

(6) *Now we can pick another $p \in [0,1]$ and do exactly the same thing. Thus, $E_n \to f$ (at least) pointwise on $[0,1]$. Also notice that each E_n is indeed a polynomial of degree $(\leqslant)n$:*

$$f(k/n)\binom{n}{k}$$

is really just some fixed constant once n and k have been prescribed, and $p^k(1-p)^{n-k}$ is a polynomial of degree $(\leqslant)n$. Then $\sum_{k=0}^{n}$ simply adds up these polynomials.

To facilitate probabilistic intuition, I used the letter p and E to denote probability and expectation. Now we shall go back to proving our Weierstraß Approximation Theorem, adopting the notation $P_n(x)$ rather than $E_n(p)$.

Proof of the Weierstraß Approximation Theorem. For notational convenience, we will write $r_k(x)$ to denote "the probability of exact k heads showing up among n tosses of a loaded coin with head probability x", i.e.,

$$r_k(x) := \binom{n}{k} x^k (1-x)^{n-k}.$$

Step 1. We shall first derive some equations involving $r_k(x)$, which will be useful later on.

Using the binomial identity, for x, y we have

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}. \tag{2}$$

Taking $\dfrac{\mathrm{d}}{\mathrm{d}x}$ on both sides and multiplying by x, we obtain

$$nx(x+y)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k x^k y^{n-k}. \tag{3}$$

Taking $\dfrac{\mathrm{d}^2}{\mathrm{d}x^2}$ on both sides of (2) and multiplying by x^2, we obtain

$$n(n-1)x^2(x+y)^{n-2} = \sum_{k=0}^{n} \binom{n}{k} k(k-1) x^k y^{n-k}. \tag{4}$$

Setting $y := 1-x$, (2), (3), and (4) give us

$$\sum_{k=0}^{n} r_k(x) = 1 \qquad \sum_{k=0}^{n} k r_k(x) = nx \qquad \text{and} \qquad \sum_{k=0}^{n} k(k-1) r_k(x) = n(n-1)x^2. \tag{5}$$

Therefore, by (5),

$$\begin{aligned}
\sum_{k=0}^{n} (k-nx)^2 r_k(x) &= \sum_{k=0}^{n} k^2 r_k(x) - \sum_{k=0}^{n} 2knx \cdot r_k(x) + \sum_{k=0}^{n} n^2x^2 r_k(x) \\
&= \sum_{k=0}^{n} [k(k-1)+k]\, r_k(x) - 2nx \sum_{k=0}^{n} k r_k(x) + n^2x^2 \sum_{k=0}^{n} r_k(x) \\
&= n(n-1)x^2 + nx - 2nx^2x^2 + n^2x^2 = nx(1-x).
\end{aligned} \tag{6}$$

(That $\sum_{k=0}^{n} r_k(x) = 1$ is a direct consequence of the fact that the probability of *all* events is 1; $\sum_{k=0}^{n}(k-nx)^2 r_k(x)$ is called the 2nd **central moment**, or the **variance**, and in this case it's the variance of the binomial distribution $B(n,x)$.)

Step 2. Now we proceed to the main step of the proof. Let $\epsilon > 0$ be given. Since f is continuous on the compact domain $[0,1]$, it is uniformly continuous by Theorem 8.11 and bounded by Lemma 8.4. Therefore, uniform continuity says there exists some $\delta > 0$ such that $|f(x) - f(x')| < \epsilon/2$ for all $x, x' \in [0,1]$ with $|x - x'| < \delta$, and boundedness says there exists there exists M such that $|f(x)| \leqslant M$ for all $x \in [0,1]$.

Let $N \geqslant M/(\epsilon\delta^2)$ be a sufficiently a large integer. We claim that if $n \geqslant N$ then $|P_n(x) - f(x)| < \epsilon$ for all $x \in [0,1]$.

Notice that we can re-write $f(x)$ as $f(x)\sum_{k=0}^{n} r_k(x)$ by (5). Then,

$$\begin{aligned}
|P_n(x) - f(x)| &= \left|\sum_{k=0}^{n} f(k/n)\binom{n}{k} r_k(x) - \sum_{k=0}^{n} f(x) r_k(x)\right| \\
&= \left|\sum_{k=0}^{n} (f(k/n) - f(x)) r_k(x)\right| \\
&\leqslant \sum_{\substack{k=0 \\ |x-\frac{k}{n}<\delta|}}^{n} \left|(f(k/n) - f(x)) r_k(x)\right| + \sum_{\substack{k=0 \\ |x-\frac{k}{n}\geqslant\delta|}}^{n} \left|(f(k/n) - f(x)) r_k(x)\right| && \text{for convenience, denote as } \sum_1 \& \sum_2{}^{6} \\
&< \sum_1 (\epsilon/2) r_k(x) + \sum_2 2M r_k(x) && \epsilon/2 \text{ by uni. cont; } 2M \text{ b/c bounded} \\
&\leqslant \frac{\epsilon}{2} + \sum_2 2M \cdot \frac{|k-nx|^2}{(n\delta)^2} && \text{since } 1 \leqslant \left(\frac{|k/n - x|}{\delta}\right)^2 = \frac{|k-nx|^2}{(n\delta)^2} \\
&\leqslant \frac{\epsilon}{2} + \frac{2M}{(n\delta)^2} \sum_2 |k-nx|^2 r_k(x) \\
&\leqslant \frac{\epsilon}{2} + \frac{2M}{n\delta^2} x(1-x) && \text{since } \sum_2 \leqslant \sum_{1+2} \leqslant nx(1-x) \text{ by (6)} \\
&\leqslant \frac{\epsilon}{2} + \frac{2M}{n\delta^2}\cdot\frac{1}{4} = \frac{\epsilon}{2} + \frac{M}{2n\delta^2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \text{ for sufficiently large } n.
\end{aligned}$$

□

Corollary 15.1

Even better, $C([a,b])$ admits a *countable* dense subset

$$\mathcal{P} := \{\text{polynomial with rational coefficients}\}.$$

Proof. By the previous theorem, we know at least that there exists a polynomial (of real coefficients) $P_n(x) = a_n x^n + \ldots + a_0$ such that

$$\|P_n - f\|_{\sup} < \frac{\epsilon}{2}.$$

Now we need to find a polynomial with rational coefficients that is sufficiently close to P_n. We pick rationals $q_0, \ldots, q_n \in \mathbb{Q}$ such that

$$|q_k - a_k| < \frac{\epsilon}{2(n+1)\sup\{x^k : x \in [a,b]\}}.$$

[6] Recall how we used the "splitting the sum" trick in Theorem 11.3?

We claim that the polynomial defined by $Q_n(x) := q_n x^n + \ldots + q_0$ is the one we are looking for:

$$\begin{aligned}|Q_n(x) - P_n(x)| &= |(q_n - a_n)x^n + \ldots + (q_0 - a_0)| \\ &\leqslant |(q_n - a_n)x^n| + \ldots + |q_0 - a_0| \\ &< \frac{\epsilon}{2(n+1)} \underbrace{\frac{x^n}{\sup\{x^n : x \in [a,b]\}}}_{\leqslant 1} + \ldots + \frac{\epsilon}{2(n+1)} < \frac{\epsilon}{2}.\end{aligned}$$

Thus $\|Q_n - P_n\|_{\sup} \leqslant \epsilon/2$ and triangle inequality gives $\|Q_n - f\|_{\sup} < \epsilon$. □

8.4 Function Algebra & the Stone-Weierstraß Theorem

Having shown that the family of polynomials are dense in $(C([a,b]), \|\cdot\|_{\sup})$, our next question is, are there other families of functions that are also dense in this space? Maybe trig functions, for example?

Definition 15.2: Function Algebra

We call a family $\mathcal{A}$ of functions a **function algebra** on K ($[a,b]$ or any compact domain) if:

$$f + g \in \mathcal{A} \quad af \in \mathcal{A}, \quad \text{and} \quad fg \in \mathcal{A} \qquad \text{for all } f, g \in \mathcal{A} \text{ and } a \in \mathbb{R}.$$

(The product fg is defined by $(fg)(x) := f(x)g(x)$, not in other ways.)

Example 15.3.

(1) The set of all polynomials is a function algebra on any $[a,b]$.

(2) $C(K)$, where K is compact, is a function algebra.

(3) The set of all **affine functions** (of form $ax + b$, "translation of linear functions") is *not* a function algebra: $x \cdot x = x^2$ which is not affine, for example.

Definition 15.4

Let $\mathcal{A}$ be a function algebra on K and fix some point $p \in K$.

(1) We say $\mathcal{A}$ **vanishes** at p if $f(p) = 0$ for all $f \in \mathcal{A}$.

(2) We say $\mathcal{A}$ **separates points** if for each distinct pair of points $p_1, p_2 \in K$, there exists $f \in \mathcal{A}$ such that $f(p_1) \neq f(p_2)$.

Theorem 15.5: Stone-Weierstraß Theorem

If $\mathcal{A}$ is a function algebra on K that vanishes nowhere and separates points, then $\mathcal{A}$ is dense in $C(K)$.
This is an equally remarkable theorem! Unfortunately we are running out of time, so the proof is omitted. See Rudin's Theorem 7.32 or Pugh's Theorem 4.20, for example.

Remark. There is an analogous version for complex functions, in which we need to replace "function algebra" by a $\mathbb{C}$-function algebra: closed under sums, $\mathbb{C}$-scalar multiples, function multiplication, *and* in addition complex conjugation, i.e., if $f \in \mathcal{A}$ then $\overline{f} \in \mathcal{A}$ where $\overline{f}(x) := \overline{f(x)}$.

Example 15.6. The trigonometric polynomials are dense in $C([a,b])$. The trig polynomials are of form

$$T_n(x) = \sum_{k=0}^{n} a_k \sin(kx) + b_k \cos(kx).$$

8.5 Devil's Staircase

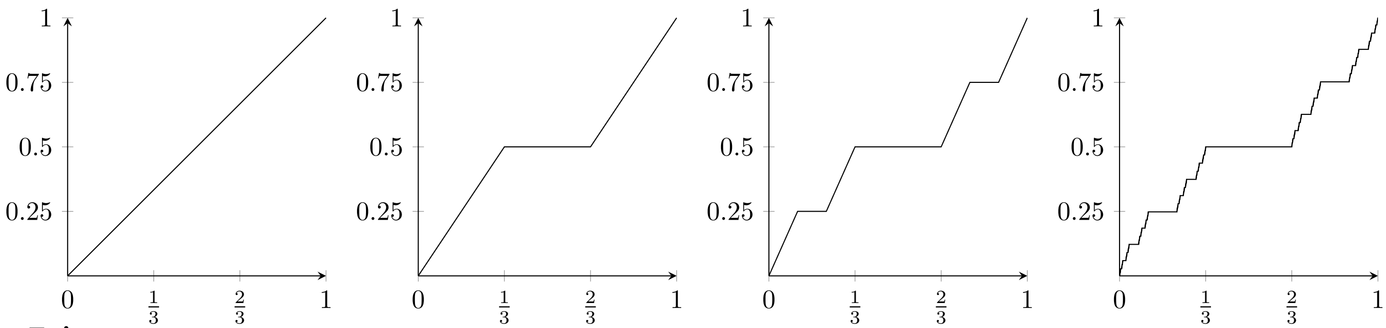
Beginning of April 28, 2021

Now we finally have enough knowledge to analyze Example 0.2 rigorously.

Example 15.7: Devil's Staircase / Cantor Function. The construction of Devil's Staircase is done iteratively. We let $f_0(x) := x$ on $[0,1]$. Then we let

$$f_{n+1}(x) := \begin{cases} f_n(3x)/2 & x \in [0,1/3] \\ 1/2 & x \in (1/3,2/3) \\ 1/2 + f_n(3x-2)/2 & x \in [2/3,1]. \end{cases}$$

We claim that these f_n's converge to some limit function f, and this limit is the **Devil's Staircase**.



Existence

By construction, $f_0 \in C([0,1])$ and so is each f_n. Notice that $\|f_{n+1} - f_n\|_{\sup} \leqslant \|f_n - f_{n-1}\|_{\sup}/2$:

(1) If $x \in [0,1/3]$ then

$$|f_{n+1}(x) - f_n(x)| = \left| \frac{f_n(\overbrace{3x}^{\in[0,1]})}{2} - \frac{f_{n-1}(\overbrace{3x}^{\in[0,1]})}{2} \right| \leqslant \frac{\|f_n - f_{n-1}\|_{\sup}}{2}.$$

(2) If $x \in (1/3,2/3)$ then $|f_{n+1}(x) - f_n(x)| = 0$.

(3) If $x \in [2/3,1]$ then again

$$|f_{n+1}(x) - f_n(x)| = \left| \frac{f_n(\overbrace{3x-2}^{\in[0,1]})}{2} - \frac{f_{n-1}(\overbrace{3n-2}^{\in[0,1]})}{2} \right| \leqslant \frac{\|f_n - f_{n-1}\|_{\sup}}{2}.$$

This should remind us of the Banach FPT. In fact, if we define a mapping $S : C([0,1]) \to C([0,1])$ by

$$S(f_n) := f_{n+1}$$

then (f_n) is Cauchy in $C([0,1])$. Therefore there exists a $f \in C([0,1])$ to which f_n converges uniformly! The limit is what we define to be the Devil's Staircase (and it is a fixed point of S, which shows that f has a *fractal*, self-similar property: if we only look at f on $[0,1/3]$ and scale it, we can recover f).

Uniform Continuity and Monotonicity

Since each f_n is continuous and they converge uniformly, the limit f is also continuous, and since the domain $[0,1]$ is compact, we see that the Devil's Staircase, despite its *fractal* appearance, is **uniformly continuous**!

Since f is the pointwise limit of a sequence of nondecreasing functions, it is also nondecreasing.

Also, f is **constant outside the Cantor Set $\mathcal{C}$**. Notice that the complement $[0,1]-\mathcal{C}$ is the union of the open "middle-third" intervals:

$$[0,1]-\mathcal{C}=\bigcup_{k=1}^{\infty}\bigcup_{i=1}^{2^{k-1}} I_{k,i}=I_{1,1}\cup I_{2,1}\cup I_{2,2}\cup\ldots:=\underbrace{\left(\frac{1}{3},\frac{2}{3}\right)}_{\text{first iteration}}\cup\overbrace{\left(\frac{1}{9},\frac{2}{9}\right)\cup\left(\frac{7}{9},\frac{8}{9}\right)}^{\text{second iteration}}\cup\ldots$$

By construction, all intervals are pairwise disjoint! The first interval has length $1/3$, the second and third together have $2\cdot 1/3^2$, the next 4 together have $2^2/3^3$, so the total length is given by the geometric series

$$\text{Total length }=\sum_{k=1}^{\infty}\frac{2^{k-1}}{3^k}=\frac{1}{3}\sum_{k=0}^{\infty}\left(\frac{2}{3}\right)^k=1. \tag{1}$$

This shows that f **is constant "almost everywhere" yet somehow $f(0)=0$ and $f(1)=1$[!]**

Differentiability

We will now show that f **is not differentiable**. In particular, it is not differentiable at $2/3$, for example.

Proof. Suppose $f'(2/3)$ exists. Then in particular the right limit

$$\lim_{x\to 2/3^+}\frac{f(x)-f(2/3)}{x-2/3}$$

exists. Recall that f is a fixed "point" of S, i.e., $S(f)=f$, so we can rewrite $f(x)$ as $Sf(x)$ [or $[S(f)](x)$, but we drop the parentheses for convenience]. Using the definition for $x\in[2/3,1]$, this gives

$$f(x)=\frac{1}{2}+\frac{f(3x-2)}{2}.$$

Also, $f(2/3)=1/2$. Therefore,

$$\lim_{x\to 2/3^+}\frac{f(x)-f(2/3)}{x-2/3}=\lim_{x\to 2/3^+}\frac{f(3x-2)/2}{x-2/3}=\lim_{x\to 2/3^+}\frac{3}{2}\frac{f(3x-2)}{3x-2}. \tag{2}$$

Defining $y:=3x-y$, (2) is equivalent to $\frac{3}{2}\cdot\lim_{y\to 0^+}f(y)/y$. We claim that such limit does not exist. Consider $y_n:=3^{-n}$, which gives

$$\frac{f(y_n)}{y_n}=\frac{Sf(\overbrace{y_n}^{\in[0,1/3]})}{y_n}=\frac{f(\overbrace{3y_n}^{=y_{n-1}})/2}{\underbrace{y_n}_{=3y_{n-1}}}=\frac{3}{2}\frac{f(y_{n-1})}{y_{n-1}}=\ldots=\left(\frac{3}{2}\right)^n\frac{f(1)}{1}=\left(\frac{3}{2}\right)^n,$$

so as $n\to\infty$, $f(y_n)/y_n$ also $\to\infty$. Therefore the limit does not exist! ☐

It turns out that, for a weird example like the Devil's Staircase, uniform continuity is not good enough. It is missing what is called the —

Absolute Continuity

Definition 15.8: Absolute Continuity

We say $f : I \to \mathbb{R}$ is **absolutely continuous** if

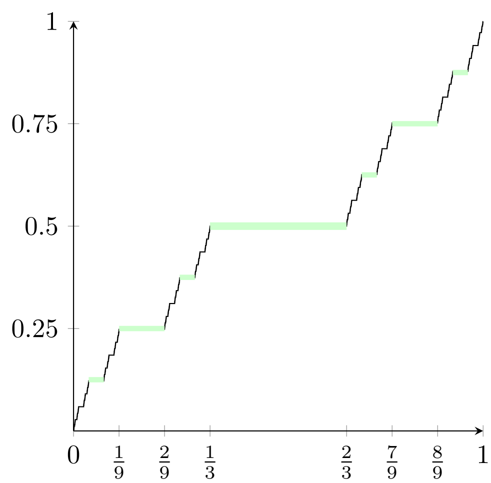
> Given $\epsilon > 0$, there exists $\delta > 0$ such that *any finite collection of pairwise disjoint intervals* $(x_k, y_k) \subset I$ satisfies the following ϵ-δ condition:
>
> $$\sum_{k=1}^{n} |y_k - x_k| < \delta \implies \sum_{k=1}^{n} |f(y_k) - f(x_k)| < \epsilon.$$

(In other words, we generalize the notion of ϵ-δ condition to beyond just one open interval.)
*It is clear that **absolute continuity ⇒ uniform continuity**: if the absolute continuity condition holds, then of course it holds for a single open interval, which is precisely the condition for uniform continuity. The converse, however, is false —*

Theorem 15.9: Devil's Staircase is Not Absolutely Continuous

The Devil's Staircase is uniformly but not absolute continuous, thereby serving as a counterexmple to the $\not\Leftarrow$ direction above.

Proof. The following diagram basically illustrates the proof: with the green segments removed, the remaining intervals has shorted length, but $\sum |f(y_k) - f(x_k)|$ is still 1, as f is constant on every single green segment.



To put formally, let us take $\epsilon := 1/2$. We claim that no δ works! By (1), for any $\delta > 0$, however small, we can find a sufficiently large N such that

$$\frac{1}{3} \sum_{k=0}^{N} \left(\frac{2}{3}\right)^k,$$

the total length of all "middle-third" intervals is sufficiently sloe to 1 ($> 1 - \delta$, in particular). If we take away these finitely many "middle-third" intervals (the green ones in the figure), the remaining ones have total length $< \delta$. However, these removed intervals have no effect on $\sum |f(y_k) - f(x_k)|$ as f is constant on all of them. Hence $\sum |f(y_k) - f(x_k)| = 1 > \epsilon$, even though δ is small. Thus no δ works, i.e., f is not absolutely continuous. $\square$

8.6 Weierstraß' Monster

Having just studied the *Devil*, we now conclude the course with a *monster*[!]

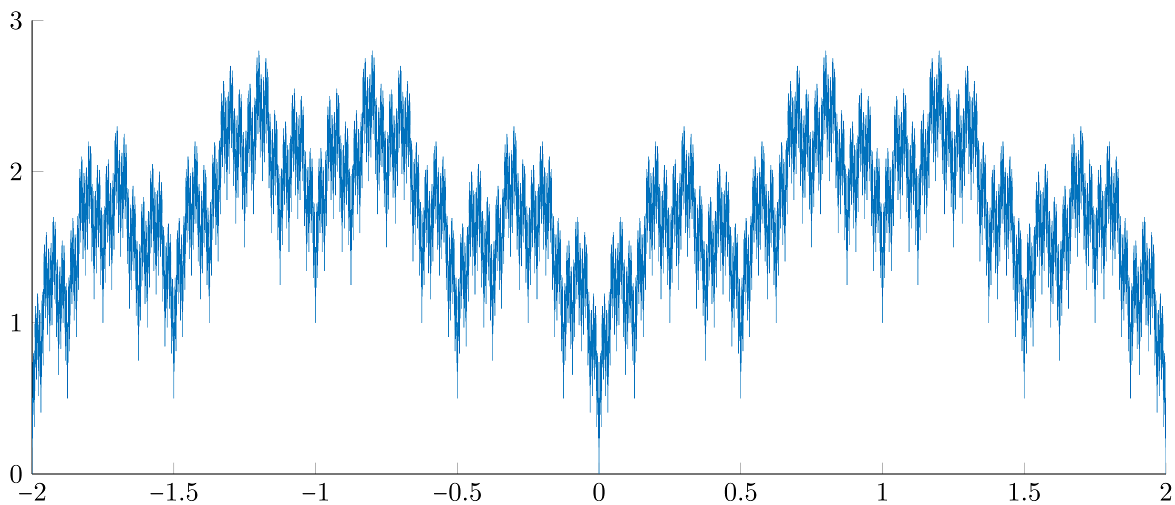
Theorem 15.10: Weierstraß' Monster

There exists $f \in C(\mathbb{R})$ that is *nowhere*[!] differentiable.

When Weierstraß first published this result in 1872, his contemporaries denounced such construction and considered it a **monster**. Henri Poincaré griped that it was "an outrage against common sense," and Charles Hermite famously wrote:

> *I turn with terror and horror from this lamentable scourage of continuous functions with no derivatives.*

While the original *Weierstraß' monster* involves an infinite series of transcendental functions (cosines), we will investigate a purely algebraic one, as seen in both Rudin's Theorem 7.18 and Pugh's Theorem 4.31.
The main idea is simple — we will add up some **sawtooth functions** to make the partial sum non-differentiable at more and more points, eventually resulting in all of $\mathbb{R}$.



Zoom in to see its craziness! This graph consists of $4^7 + 1 = 16385$ data points. This (and almost all other diagrams in this notes) is a vector graph and zooming in will not blur the image, only to make it clearer.

Construction of the *Monster*

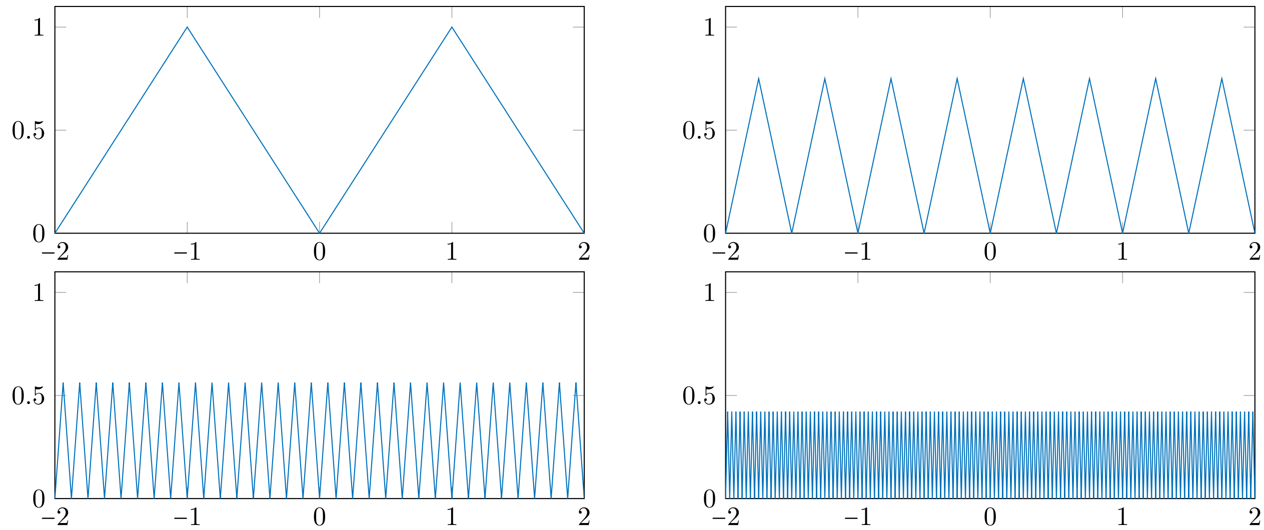
We begin with a **sawtooth function** $\varphi_0(x)$ defined by

$$\varphi_0(x) := |x - 2n| \text{ for } x \in [2n-1, 2n+1) = \begin{cases} x - 2n & x \in [2n, 2n+1) \\ 2n + 2 - x & x \in [2n+1, 2n+2) \end{cases}$$

The graphs clearly illustrates why it is called the "sawtooth function". Next, we define $\varphi_n(x)$ iteratively by

$$\varphi_n(x) = \left(\frac{3}{4}\right)^n \varphi_0(4^n x).$$

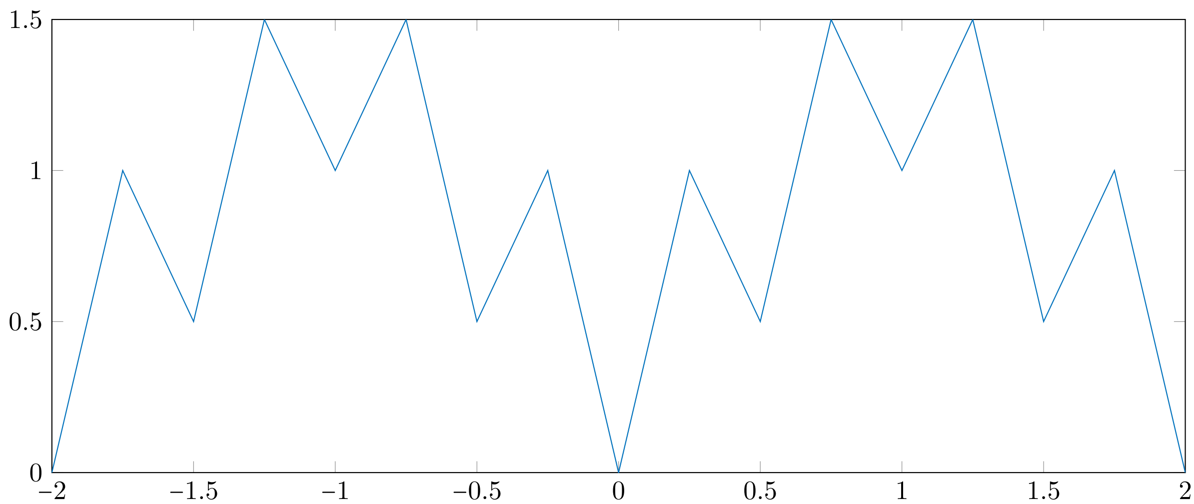
Below are the graphs for $\varphi_0, \varphi_1, \varphi_2$, and φ_3 on $[-2, 2]$. The exponential growth is just scary...

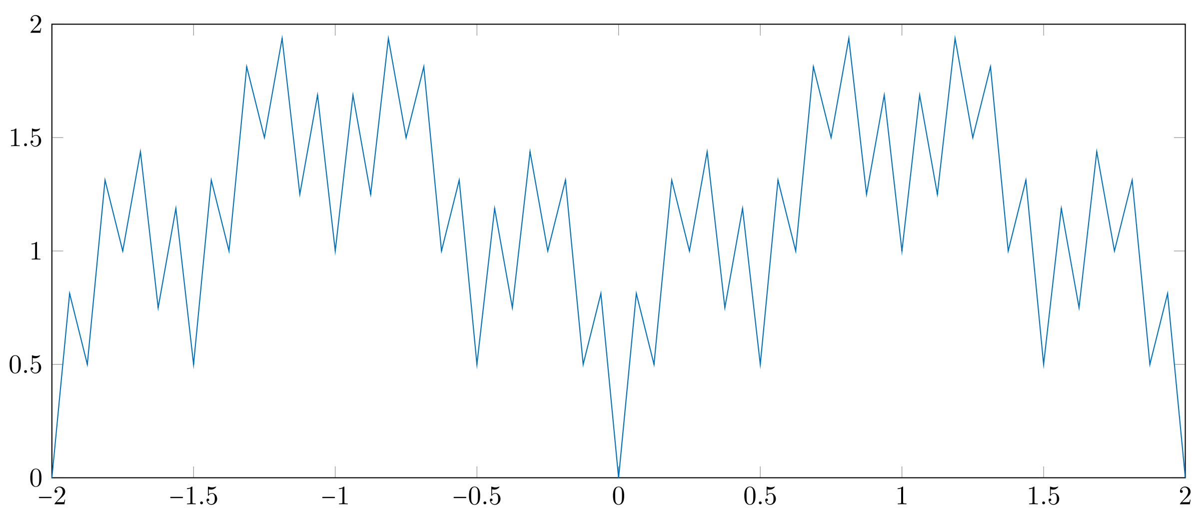


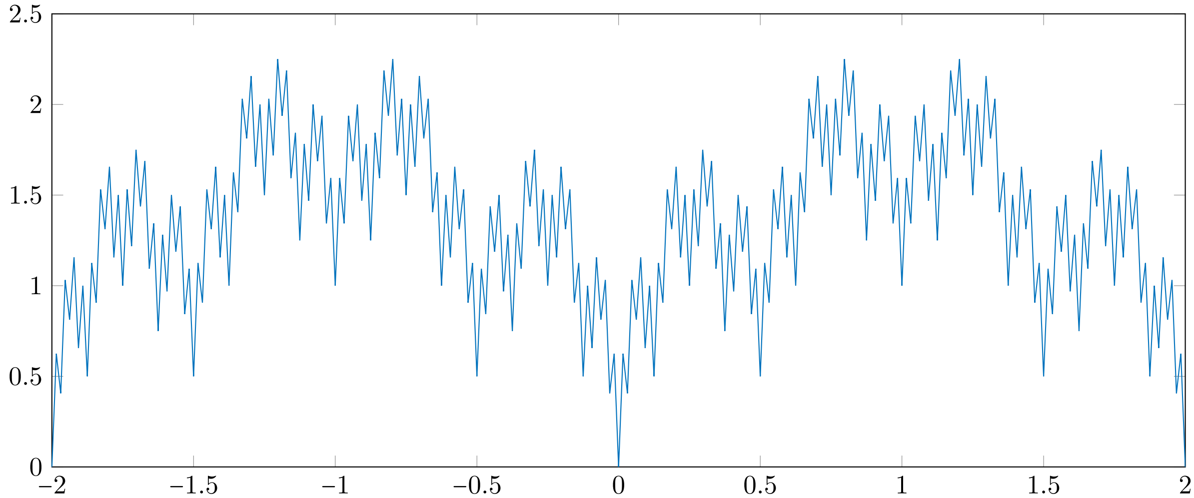
We define the *monster* function to be

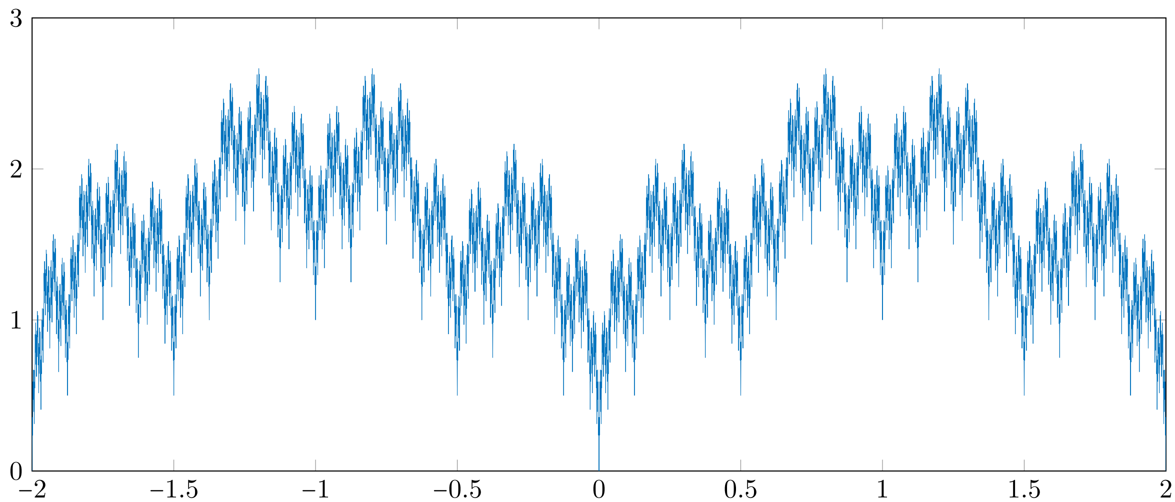
$$f(x) := \sum_{n=0}^{\infty} \varphi_n(x) = \sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n \varphi_0(4^n x).$$

I have plotted $n = 1, 2, 3$, and $n = 5$; the graphs below are put in this order. (The one above is $n = 6$.)









Continuity and Differentiability

By the Weierstraß M-Test, each φ_n is bounded by $M_n := (3/4)^n$, so $\sum \varphi_n$ converges uniformly to f. Then, since each φ_n is continuous, Theorem 13.8.1 asserts that the limit f is **continuous**.

Now we show that f is **nowhere differentiable**. We pick $x \in \mathbb{R}$ and show that, by setting $\delta_m := \pm 4^{-m}/2$, the quotient

$$\frac{f(x+\delta_m) - f(x)}{\delta_m}$$

blows up (just like the 3^{-n} sequence in Devil's Staircase). The sign of δ_m is determined by requiring that no integer lies in the interval between $4^m x$ and $4^m(x+\delta_m) = 4^m x \pm 1/2$. (This is always possible!)
Next, we define

$$\gamma_n := \frac{\varphi_0(4^n(x+\delta_m)) - \varphi_0(4^n x)}{\delta_m}$$

(so the quotient $(f(x+\delta_m) - f(x))/\delta_m$ is just $\sum \gamma_n$). Now we bound γ accordingly:

(1) If $n > m$, then $4^n \delta_m = 4^{n-m}/2$ is an even integer. Since φ_0 is 2-periodic (i.e., periodic with period 2; check the first sawtooth graph), $\varphi_0(4^n x + 4^n \delta_m) = \varphi_0(4^n x)$, so the numerator = 0 and $\gamma_n = 0$.

(2) If $n = m$, $4^n \delta_m$ is precisely $1/2$, so in this case

$$|\gamma_n| = \left|\frac{1/2}{4^{-m}/2}\right| = 4^m.$$

(3) Finally, if $n < m$, notice that φ_0 is Lipschitz with Lipschitz constant = 1. Then,

$$|\varphi_0(4^n(x+\delta_m)) - \varphi_0(4^n x)| \leqslant |4^n(x+\delta_m) - 4^n x| = 4^n \delta_m.$$

Therefore $|\gamma_n| \leqslant 4^n \delta_m / \delta_m = 4^n$.

Finally, using (reverse) triangle inequality $|a + \sum b_i| \geqslant |a| - \sum |b_i|$, we have

$$\begin{aligned} \left| \frac{f(x+\delta_m) - f(x)}{\delta_m} \right| &= \left| \sum_{n=0}^{m} \left(\frac{3}{4}\right)^n \gamma_n \right| \\ [\Delta] &\geqslant \left(\frac{3}{4}\right)^m \underbrace{|\gamma_m|}_{=4^m} - \sum_{n=0}^{m-1} \left(\frac{3}{4}\right)^n \underbrace{|\gamma_n|}_{\leqslant 4^n} \\ &\geqslant 3^m - \sum_{n=0}^{m-1} 3^n = 3^m - \frac{1-3^m}{2} = \frac{3^m - 1}{2} \end{aligned}$$

which blows up as $n \to \infty$. Therefore $f'(x)$ exists nowhere, and we are done!! □