Introductory Real Analysis - Lecture notes - Fall 2024

Summary : Fill in later, acknowledgements, etc.

Contents

Lectures

Solutions to Exercises

Sections to Finish :

- unlock solutions as they finish (figure out a better way to do this), remember solutions for any problems you add.
- Intro and closing bit to each chapter.
- Add in way more references
- read through abott and zorich and think
- maybe put this after table of contents and all that, fix lecture numbering as we get to stuff each day.
- Acknowledgements and an intro?
- Ex, Thm numbering per chapter? figure out
- Bring in more history when you can
- Dictionary of symbols, list of definitions and theorems? (Add to this as you go)
- Summary at end of each section?
- Make homework questions for section 6.5, 7.5,7.6, 8.1,8.2,8.3,8.4,8.5,8.6, 9 section 6.4 number 4 fix wording.
- Clean up spacing at the end. newpage and stuff, space after each section before exercises?
- Willard 7C continuity, rudin 46 number 30, and 82 number 21 and 22, Baire?
- fix the its
- decimals from two sides, geometric series via mct to get convergence, or nested interval property for each base. Comment on decimals and dedekind, any base, maybe a appendix. possibly mention continued fractions here?
- generally finding more problems
- go into e a bit more, ln in inverse function theorem, trig in power series, nth roots in general in continuity etc
- Problem on arithrogeometric mean that gauss liked with 1 and root 2, in Gray change and variation book.
- 1.1 - Mention how n to z to q are two step processes, algebraic numbers are finite (maybe not in decimal but in poly) Mention how we know Q and A are small in the sense of cardinality and have gaps more specifically (maybe leave proof to cardinality section), it will be a little awkward to talk about R without R yet, maybe just call it a continuum or infinite precise ruler for now. Fix part about it making Calculus possible. Maybe add in another homework problem. idea about countability of point versus space, how R is made of up of infinite processes and the number is uncountable. Degree of algebraic number.

- 1.2 Maybe mention nested interval property early in terms of cuts or why R is necessarily uncountable due to completeness, write a section on decimals really in an appendix. (and possibly connect this here) Maybe bring up Baire and irrational numbers. (continued fraction?) union of N to the k is not N to the N, mention with decimals, rationals, etc. irrationals, continued fractions, countable from continuum is not nothing oxtoby, how jump from n to z to q is all finite, not countable, In cardinality section, give better argument of $\mathbb{N}^2$ being countable, steal pic from talk you gave in cosmos. Check against 100 notes, try and make potato model example check tikz? check proofs in the appendix and clean them up, maybe give proof of theorem 4 with primes. Check history of what Cantor actually did prove, maybe put oxtoby in here. fix dash in 01 sequences, maybe bring up set exponentiation definitions, maybe add exercises, maybe fix question 3 and clean it up a little. union of N to the k is not N to the N - maybe put this in cardinality appendix.
- 1.3 explain the choice of positive. Maybe put proof of prop in section and not appendix. Maybe explain why we care about field structure.
- 1.4 Rename the section, archimedean property and mention the density versus discreteness
- Make a 1.5 and put more on bounded sets and defs here, and completeness.
- 2.1 Maybe more explanation of what is happening in convergence, bring in neighborhoods, distance, mention topology lingo earlier in sequences (windows, epsilon balls, etc), Add in geogebra links, probably use logarithmic scale to make it easier to see. Maybe use python for picture and use pdf to insert move definition 20 earlier, problem 2.2.8 clean up.
- 2.2 Reread section, In def for boundedness of sequence, put boundedness of sets as well (check sup and inf section), make connection to bounded of sequence and set ASAP, generally connecting sequence and set. Check definition of diverging to infinity, is this right, what Quincy said. Put in acknowledgements. Maybe remark after theorem 11.
- 2.3 Make connection to sequences and tails more explicit, connect to 2.1.
- 2.4 maybe start e example here, reread and change footnotes depending on what real analysis in reverse you do. Bit about Cauchy sequences being cadidates, mention what you mention at start of 3, why it's good bounded does not imply cauchy but convergent does. Problem 2 has typo add show that. How Cauchy does not have bad behavior, osciallation or divergence, it's like a pre-convergent notion, only way a cauchy sequence doesn't converge is because its limit does not exist in the set.
- 3.1 give another example of equivalence relation to start, explain how later we will see x as the limit of its representations, but right now destinations in terms of paths not paths to destinations.
- 3.2 everything has a cost, the cost of equivalence classes. For sake of completeness put remaining field axioms at end.
- 3.3 likely skip this in class, leave for student reading. main order argument point out it is trichotomy law and really point out how we are bootstrapping twice.
- 3.4 explain more of how abs value and triangle inequality still hold here, also injection mapping is a field isomorphism, fix argument of density, where did the Cauchy sequence come from, spend a lot more time with this moving back in forth between representations and sequences

of images under the injection mapping, actually fill in the proofs of things you mention in the notes like triangle inequality and algebraic limit laws. Like explain how triangle inequality really works in R, why the injection mapping is an isometry in this sense, make sure argument of AP or density, or sequences representing x converging to x explains why epsilons is real or rational, just generally more detail when you move to injection mapping. Be more careful about when epsilon is real or not. Maybe give second argument of Archimedean property as proof or homework assignment once we have sup and inf.

- 3.5 maybe mention theorem 31 around injection mapping in previous section also in prior section. explain a bit better at the end why the gaps are gone, why we have gone as far as we can. all norms equivalent in finite dimensions, why always R when completing with archimedean (maybe put this in a section after on equivalent forms of completeness, bring up NIP here after mentioning in 1.2), advantage of this method, why dedekind only really works for R, how Cauchy works in general with metrics, but we need completeness of R first to use this argument. Doing second method of existence and uniqueness of nth roots from sup and inf (maybe in chapter 4)
- chapter 4 is where i should connect completeness to MCT and BW and others to start.
- 4.1 more examples finding the sup, maybe bring in one way of finding nth roots here (leave other as exercise) also do uniqueness on this as well, more on how inf and sup extend max and min for finite sets. how sup and inf generalize least and greatest elements when in linear order.
- 4.2 beginnings of e in section 4.2? leave it as a function, continuity, diff, etc. to later sections 5 on.
- 4.3 bolzano weierstrass completeness of r question.
- 4.4 Section connecting limit sets and limit points? Start the connection here, maybe mention boundary in 4.2 or 4.3, prior we were forcused on sequences going somewhere, other side is which points have sequences going to them.
- Really find a good through line between sequences going somewhere and when we have sequences going to something, kind of tee up 6, 7, etc.
- 5.1 reread and edit, maybe add some stuff
- 5.2 mention how sparseness relation is useful for ln problems.
- 5.3 maybe clean it up a bit more
- 5.4 delete stuff at front if you are done, maybe make theorem 54 better about rearrangement really needs to be infinite. check riemann rearrangement
- 5.5 put off mult until power series maybe or connect to polynomial
- 5.6 in notes
- 5.7 in notes
- 6.1 reread, maybe take out question 4.

- 6.2 fix definition 41 about isolated point, section 6.2, check definition of isolated and discrete, Number 10 in 6.2, use this an excuse to bring up ordinals at some point, Cantor Bendixson rank? difference between subsequential limit and limit point in more detail.
- 6.3 Reread and edit
- 6.4 Reread, edit, add nested interval property into compactness section, more detail on compact being almost finite (maybe in EVT as well), check proofs in compactness section, limit point used correctly? section at start of relative topology pointing out section 6 was from perspective of r as background and why compactness matters maybe as a subsubsection, Big O in section on uniform cont? wording on 6.4 number 4. make Bolzano weierstrass better here.
- 6.5 in notes,
- 6.6 possibly at end after showing $\mathbb{R}$ is connected, after continuity and connectedness make a problem showing $\mathbb{R}$ is connected this way (steal cosmos problem with half-circle), difference between connected and convex in higher dimensions
- 7.1 continuity on R really, no rn yet, limits, at limit points etc versus discrete points. bring up here. pictures in section 7, write intervals in chapter 7 and 8 as intervals and not I. appendix on functions, when they have inverses.
- 7.2 once again R and not rn, topology connection, any function on discrete is continuous, move to 7.2?, fix devil stair problem in section 7.2
- section on when continuous functions have inverses. lipschitz and holder continuity put more , bring in e somewhere here (use to motivate power series)
- 7.3 compact is almost finite again, Big O notation here, redo homework questions.
- 7.4 move theorem 90 earlier, make inverse claim its own section. baby inverse function theorem.
- 7.5 F sigma and G delta maybe come up? Oscillation will probably, converse to IVT is not true, put in after section on discontinuities, add sinx over x and cosine one into section on discontinuities, missing squeeze theorem in continuous functions, maybe make an exercise
- 7.6 in notes, connect to compactness again, field of rational functions is nonarchimedean?
- 8 miscellany, add in pictures into derivative sections, c1 def in 8.1 makes sense, leave other for different sections, mention higher derivatives in sections that require it, and notation for it as well? finish chapter 8 stuff, Implicit Function Theorem? put Leibniz notation into notes for Chapter 8, clean up theorem 93 on inverse for continuous function existing, maybe ref in inverse function theorem section. reframe local properties of derivative, what do we require, diff at a point? connect this with local properties of first derivative, have the results be mirror, mention strict local extrema and local extrema, be more precise, more pictures in inverse function theorem section, better description of local properties of f given by f prime, two proofs of the chain rule, why give the more complicated one anyways (because it generalizes easier to higher dim)
- 8.1 both notations, reread and edit, define trig later in power series, maybe bring up e here, stuff in notes

- 8.2 local behavior, reread, edit, clean up theorems, stuff in notes
- 8.3 stuff in notes, algebraic versus transcendental, louiville number, same for functions? do liouville number at start in section 1?
- 8.4 in notes,
- 8.5 in notes
- 8.6 in notes
- Solutions chapter 1
- solutions chapter 2
- solutions chapter 3
- solutions chapter 4
- solutions chapter 5
- solutions chapter 6
- solutions chapter 7
- solutions chapter 8
- Try to plan out sections for 9 on.
- 9 miscellany, riemann integration seperate from stieljes, then stieljes and connect to series and preview distributions, bring lebesgue in later, baire sets and borel, regularity, category and measure, connect to dual, riesz-rep, bounded variation and absolute continuity? try to find some history and connective tissue between everything here. stuff in todo from integration section. Maybe get to distributions a little? (after c infinity and power series and linear functional) (Duistermaat quote)
- after this talk a little bit about measure on R and category on R, power series, series of functions, taylor and fourier, then maybe do basic Lebesgue theorem, abs cont and bounded vary.
- Metric and connections to top, sequences, continuity
- Diff and Int on Rn will be it's out things,
- What I wrote awhile back: ok for plan after this, I think I know what I am doing, section at end about smallness with countability, category, and measure. Then make you way in three possible directions from here, one, multivariable stuff and inverse and ift, second measure and getting to lebesgue def, derivatives of monotonic functions, egoroff and lusin, three, stieljes, bounded variation and absolute continuity, beginning of functionals and riesz rep.
- at very end some distributions, Q to R, functions to Dist. all this stuff may be way too much though.
- Maybe no appendices, put into notes in places but as subsubsections.

- Appendix on decimals: poorly named, general decimal bases and completing without archimedean prop, appendix.
- Appendix, Really go into what is equivalent to completeness of R. Check other resources, zorich, spivak, etc. Put books in folder inside of other folder.
- Appendix: Make construction of $\mathbb{N}$ more rigorous. maybe define recursion and peano systems. clean up construction of the other number systems also. Make each subsubsubsection distinct, write 'addition' in color or something
- Appendix: Finish binary representation in appendix and clean it up. Also maybe make some proofs in appendix on countable cleaner as well.
- Appendix: Cardinality - do we need this equivalence concept business, put CSB theorem in notes, clean up theorem 108 TFAE on infinite, example 46 clean, theorem 11 could be simpler probably, maybe cut out lemma 114, clean up theorem 116 or give proof of cartesian product first and deduce it, basically i mean do 117 then 116 (maybe show both), cut out 119 if you do a section on decimals, same with 120, maybe move example 49 into 1.2, maybe don't need a second heading for stuff about uncountable things.
- Maybe move Prop about Fields into section 1.3
- reread induction section and edit if need be, maybe mention open induction as equivalent to completeness (no gaps), maybe bring up transfinite if we have ordinals from cantor bendixson.
- once again, maybe an appendix on functions also
- Appendix on equivalence relations, maybe put in the section we need them in.
- appendix: put in construction via dedekind cut (maybe put at the end of chapter on construction of R)
- Surreal, only do if you have ordinals made either from enderton or cantor bendixson.

1 Why $\mathbb{R}$

1.1 Algebra will only take us so far

Lecture 1 - 9/27/24

Let us start with some of the common number system sets that we will encounter:

$\mathbb{N}$ - the *natural numbers* - the set of positive whole numbers $\{1, 2, 3, \dots\}$. [1]

$\mathbb{Z}$ - the *integers* - the set of all whole numbers $\{\dots, -2, -1, 0, 1, 2, \dots\}$.

$\mathbb{Q}$ - the *rational numbers* - the set of all possible fractions.

$$\mathbb{Q} = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0, \gcd(a, b) = 1 \right\}$$

the last condition involving the greatest common divisor is usually present to make sure we are always looking at a fraction in its *lowest terms* form, i.e. with all common factors from the numerator and denominator cancelled out.

Each of these number systems is larger than the one proceeding it, equivalently, in set notation we have $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q}$. But one can often think that each number system was brought into existence to fix an algebraic malady a prior number system had.

For example, if we try to solve the equation $x + 3 = 0$ in $\mathbb{N}$ we find that there is no solution as $x = -3$ does not exist in $\mathbb{N}$. However, this equation does have a solution in $\mathbb{Z}$ as the integers contain negative numbers. Similarly, the equation $2x + 3 = 0$ has no solution in $\mathbb{Z}$ as $x = -\frac{3}{2}$ is not a whole number, but the solution does exist in $\mathbb{Q}$. [2]

For a last example, we return to a classic proof. The equation $x^2 - 2 = 0$ has no solution in $\mathbb{Q}$ as $\sqrt{2}$ is not rational. We include the proof here for completeness.

Theorem 1. *$\sqrt{2}$ is not rational.*

Proof. By way of contradiction, we assume that $\sqrt{2} \in \mathbb{Q}$, i.e. $\sqrt{2} = \frac{a}{b}$ for integers a, b with $b \neq 0$ and a and b having no common factors. Squaring both sides and multiplying across gives that

$$a^2 = 2b^2$$

And this implies that a^2 is an even number. It is true that if a^2 is even then it must be that a is even. Thus $a = 2m$ for some integer m, but this gives that

$$4m^2 = 2b^2 \implies 2m^2 = b^2$$

and this says that b^2 is even and thus b is even. But we then have that

$$1 = \gcd(a, b) \geq 2$$

which is a clear contradiction. Thus it must be the case that $\sqrt{2}$ is not rational. □

[1] Sometimes 0 is considered a natural number, in this case $\mathbb{N}$ is the set of nonnegative whole numbers. The notation $\mathbb{N}_0$ is often used to denote natural numbers containing 0.

[2] Another equivalent way of putting this is that while $+$ is a well defined operation on $\mathbb{N}$, subtraction is not ($2 - 5 \notin \mathbb{N}$). Similarly, multiplication is well defined in $\mathbb{Z}$, but division is not. ($\frac{2}{7} \notin \mathbb{Z}$)

Now at this point, some may say this is the value of the real numbers; that it let's us solve even more algebraic equations, but this (in my opinion) is kind of dishonest or misleading about why the real numbers are so important. Let me explain a little more with a definition first.

Definition 1. *A number a is called* **algebraic** *if it is the solution to a polynomial equation with integer coefficients*[3]*, i.e. there exists a polynomial $p(x)$*

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

$a_n, a_{n-1}, \ldots, a_1, a_0 \in \mathbb{Z}$ with $p(a) = 0$. We will label the collection of algebraic numbers by $\mathbb{A}$. Any number that is not algebraic is called **transcendental**.

We have the following:

- Clearly any rational number $\frac{p}{q}$ is also an algebraic number as this solves the equation $qx - p = 0$.
- The algebraic numbers contain any nth root of any prime number p, as $x = \sqrt[n]{p}$ is a solution to $x^n - p = 0$.
- The number $i = \sqrt{-1}$ is an algebraic number as it is a solution to $x^2 + 1 = 0$.
- The numbers π and e are transcendental.

Example 1. *The number $\sqrt{2} + \sqrt{3}$ is algebraic.*

Often the easiest way to see if a number a is algebraic is to begin with the expression $x - a = 0$ and begin manipulating both sides until you have a polynomial equation. So starting with $x - (\sqrt{2} + \sqrt{3}) = 0$ we have

$$\begin{aligned}
x - \sqrt{2} &= \sqrt{3} \\
(x - \sqrt{2})^2 &= 3 \\
x^2 - 2\sqrt{2}x + 2 &= 3 \\
x^2 - 1 &= 2\sqrt{2}x \\
(x^2 - 1)^2 &= (2\sqrt{2}x)^2 \\
x^4 - 2x^2 + 1 &= 8x^2 \\
x^4 - 10x^2 + 1 &= 0
\end{aligned}$$

and this shows that $\sqrt{2} + \sqrt{3}$ is a solution to $x^4 - 10x^2 + 1 = 0$ and thus is an algebraic number.

My point is that from the view of algebra if we were only interested in solving equations then we would only need to go as far as the algebraic numbers $\mathbb{A}$, as this solves the issue $\mathbb{Q}$ has of being unable to solve certain polynomial equations. But this is why I say algebra can only take us so far, the real numbers $\mathbb{R}$, once we construct them are much larger than $\mathbb{A}$.[4]

The value of the real numbers is that it makes Calculus possible. The rational numbers and the algebraic numbers are not **complete**.[5] Both $\mathbb{Q}$ and $\mathbb{A}$ have 'gaps' in them in the sense that neither of them contain any solid interval $(a, b) = \{x | a < x < b\}$. While we will later see that $\mathbb{Q}$ and $\mathbb{A}$ are

[3] equivalently, rational coefficients

[4] with the notable major exception that $i \notin \mathbb{R}$

[5] a topic we will see shortly

dense inside of the real numbers, their **totally disconnected**[6] nature keeps Calculus from being possible on them. The study of Calculus involves limiting processes and analysis of local behaviors of functions, both are things that will require letting a variable in a domain traverse a connected continuous interval and this is what $\mathbb{R}$ will allow.

Exercises for section 1.1:

1. Show that the number $\sqrt[3]{2} + \sqrt[3]{3}$ is algebraic.

[6] another definition we will see in the future

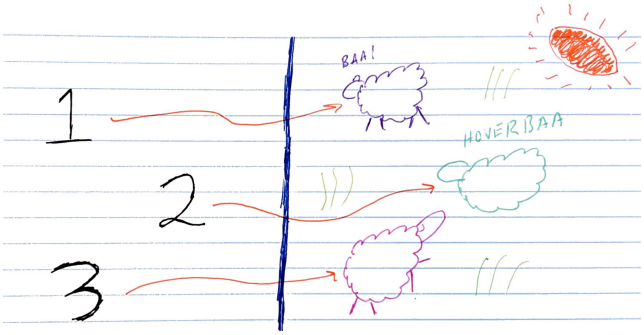
1.2 Cardinality of $\mathbb{R}$

In general, cardinality, is the study of 'size' for a set in terms of amount of things. It is a fascinating topic that is typically enccountered in the study of set theory. The basics are covered at the end of Math 100 typically, and we will not dip our toes in any deeper than we have to but for those that are curious I will put far more details on this topic in an appendix.

Definition 2. *For a set X, its* **cardinality**, *written $|X|$, is the size of X in terms of number of elements. For finite sets, $|X|$ will be an element of $\mathbb{N}$.*

In particular, it is often very simple to compare the sizes of finite sets, i.e. it is often simple to tell if two finite sets have the same number of objects or if one has more than the other etc. This can be done by simply counting the elements in the sets and then comparing, but the question is how we compare cardinalities of infinite sets. For an infinite set, counting its elements would literally take forever, so "How does one count when you can no longer count on counting?"

The answer lies in how we learned to count as children [E]



We counted by pairing up numbers with objects. We would say there are 4 objects present if there was a way to pair each object with one number from $\{1, 2, 3, 4\}$. This process of pairing is actually defining a function that maps each object to a specific number. Because of this we can phrase 'X has the same number of elements as Y' or A has less elements than B' in terms of functions with specific properties. It turns out that this generalization is the appropriate manner to handle 'counting' infinite sets.

So let us first recall the following definitions:

Definition 3. *A function $f : A \to B$ is said to be* **one–to–one** *or* **injective** *if at least one of the following hold*

- *For a given $b \in B$, there is at most one $a \in A$ for which $f(a) = b$.*
- *For any $a, b \in Dom(f)$, if $a \neq b$, then $f(a) \neq f(b)$. In other words distinct points in the domain map to distinct points in the range.*
- *For $a, b \in Dom(f)$, if $f(a) = f(b)$, then $a = b$.*

Definition 4. *A function $f : A \to B$ is called* **onto** *or* **surjective** *if at least one of the following hold*

- *$\forall b \in B$ there exists an $a \in A$ such that $f(a) = b$.*

- $Codom(f) = Ran(f)$.
- $f^{-1}(\{b\}) \neq \varnothing$ *for all* $b \in B$.

Definition 5. *A function is called* **bijective** *if it is both injective and surjective.*

These definitions of functions with these specific properties will help us define the notions of 'smaller', 'bigger', and 'equal' even in the context of infinite sets.

Definition 6. *For two sets A and B we have the following:*

- *Two sets A and B have the same* cardinality*, denoted $|A| = |B|$, if there exists a bijective function $f : A \to B$.*
- $|A| \leq |B|$ *if there exists an injective map $f : A \to B$.*
- $|A| \geq |B|$ *is there exists a surjective map $f : A \to B$.*
- $|A| < |B|$ *if there exists an injective map $f : A \to B$ and no map $g : A \to B$ is surjective.*

What follows will be what we need from the topic of Cardinality for our course. Proofs of the following theorems will be relegated to the appendix.

Definition 7. *The cardinality of the natural numbers is called $\aleph_0$, and it is read 'aleph null'. The natural numbers, in the sense defined above, are the smallest infinite set.*[7]

Definition 8. *A set X is called* **countable** *if it is finite or if it has the same cardinality as $\mathbb{N}$. Equivalently, by the definition above, a set X is called countable if there exists a bijective map $f : X \to \mathbb{N}$ or $f : \mathbb{N} \to X$.*

Lecture 2 - 9/30/24

Example 2. *The integers, $\mathbb{Z}$, are countable.*

We will construct a map that sends the odd numbers to the naturals and the even numbers to the negative naturals and zero. Thus, define a map $h : \mathbb{N} \to \mathbb{Z}$ by

$$h(x) = \begin{cases} \frac{x+1}{2} & \text{if } x \in \mathbb{O} \\ -\left(\frac{x}{2} - 1\right) & \text{if } x \in \mathbb{E} \end{cases}$$

Given any $m \in \mathbb{Z}$, by the trichotomy law m is either positive, negative, or zero.

- If m is positive. Then $2m - 1 \in \mathbb{N}$ and $2m - 1$ is odd, thus
$$h(2m-1) = \frac{2m-1+1}{2} = \frac{2m}{2} = m.$$
- If $m = 0$, then $2 \in \mathbb{N}$ and 2 is even, thus
$$h(2) = -(\frac{2}{2} - 1) = -(1 - 1) = 0.$$

[7] This will be justified in a proof in the appendix

- If m is negative, then $2(-m+1) \in \mathbb{N}$ and $2(-m+1)$ is even, thus,

$$h(2(-m+1)) = -\left(\frac{2(-m+1)}{2} - 1\right) = -(-m+1-1) = m.$$

Thus, h is an onto map.

Take $m, n \in \mathbb{N}$. If m and n have different parity, without loss of generality, assume that m is even and n is odd, then h maps m to nonpositive integers and h maps n to positive integers. Thus, if we assume that $h(m) = h(n)$, then m and n must have the same parity, i.e. both m and n are odd, or both m and n are even.

- If both m and n are odd, then

$$\begin{aligned} h(m) &= h(n) \\ \frac{m+1}{2} &= \frac{n+1}{2} \\ m+1 &= n+1 \\ m &= n. \end{aligned}$$

- If both m and n are even, then

$$\begin{aligned} h(m) &= h(n) \\ -(\frac{m}{2} - 1) &= -(\frac{n}{2} - 1) \\ \frac{m}{2} - 1 &= \frac{n}{2} - 1 \\ \frac{m}{2} &= \frac{n}{2} \\ m &= n. \end{aligned}$$

Thus in either case, $h(m) = h(n)$ implies that $m = n$. Thus h is injective. Thus h is a bijection from $\mathbb{N}$ to $\mathbb{Z}$. Thus $\mathbb{Z}$ is countable.

Theorem 2. *A countable union of countable sets is countable, i.e. if $A_1, A_2, \ldots$ is a listing of countable sets (i.e. A_j countable for all $j \in \mathbb{N}$) then*

$$\bigcup_{n=1} A_n$$

is countable.

Theorem 3. *For a countable set A, any subset $B \subseteq A$ is countable.*

Theorem 4. *For any fixed $k \in \mathbb{N}$, the k-Cartesian product of $\mathbb{N}$, written $\mathbb{N}^k$*

$$\mathbb{N}^k = \underbrace{\mathbb{N} \times \mathbb{N} \times \cdots \times \mathbb{N}}_{k-times} = \{(n_1, n_2, \ldots, n_k) \mid n_j \in \mathbb{N}\ 1 \leq j \leq k\}$$

i.e. the collection of all k-tuples with natural number inputs, is countable.

One proof of the above theorem by induction will be given in the appendix, but there is a nice proof of this that can be done using the unique factorization of a number into primes and the Cantor-Schroder-Bernstein Theorem.

Example 3. *The rational numbers, $\mathbb{Q}$, are countable*

We first break the rationals into three pieces: the negative rationals, zero, and the positive rationals.

$$\mathbb{Q} = \mathbb{Q}_- \cup \{0\} \cup \mathbb{Q}_+$$

And we really only need to show that $\mathbb{Q}_+$ is countable, as 0 is clearly countable and the negative rationals will be if the positive rational numbers are.

Each element $\frac{a}{b} \in \mathbb{Q}_+$ with $\gcd(a, b) = 1$ can be paired with the coordinate $(a, b) \in \mathbb{N}^2$. In this sense, $\mathbb{Q}_+ \subset \mathbb{N}^2$, and so by the above theorems $\mathbb{Q}_+$ is countable.

It is interesting to the note that the algebraic numbers, $\mathbb{A}$, are also countable. However, with all of this said, we now come to the real numbers, $\mathbb{R}$, and what Cantor proved in 1874 is that the real numbers are uncountable. In other words, the real numbers have a cardinality larger than that of the natural numbers, and are an infinite beyond the smallest infinity (in terms of size/counting) In effect, a finite quantity, no matter how large, is seen as small from the perspective of a countable infinite, and in the same manner a countable infinite is seen as small from the perspective of an uncountable infinite. And so, we can see that the cost of studying Calculus, of requiring a continuum is that we must work with such a large collection of numbers.

Definition 9. *A set that is not countable is called* **uncountable**.

Sets that are uncountable can not be listed by the natural numbers, i.e. it is impossible to label every element of an uncountable set with natural numbers. Another way of thinking of this is as follows: If you can imagine an uncountable set and $\mathbb{N}$ side by side, and you began to take one element from each set and pair them together (i.e. you are trying to build a bijection directly), then you would exhaust the natural numbers before the uncountable set becomes empty. To be clear, uncountable sets have strictly more elements than $\mathbb{N}$.

We are about to show that the real numbers are uncountable, however we will not do so directly. Granted, we have not rigorously constructed the real numbers yet, but I assume most students have a passing familiarity of $\mathbb{R}$ from prior classes. Instead we will show why the collection of $0 - 1$ sequences is uncountable, and in the appendix, argue that this is equivalent to $\mathbb{R}$ from using a base 2 (binary) number system.

The set $\{0, 1\}^{\mathbb{N}}$ notationally stands for the set of all sequences of zeros and ones, i.e.

$$\begin{aligned}\{0, 1\}^{\mathbb{N}} &= \{f : \mathbb{N} \to \{0, 1\} \mid f \text{ a function}\}. \\ &= \{a_1 a_2 a_3 \ldots \mid a_k \in \{0, 1\} \ \forall k \in \mathbb{N}\}.\end{aligned}$$

We now present a classical argument of Cantor, the so–called Cantor Diagonal argument, to show why $\{0, 1\}^{\mathbb{N}}$ is uncountable.

Theorem 5. *The set* $\{0,1\}^{\mathbb{N}}$ *is uncountable.*

Proof. By way of contradiction, we suppose that $\{0,1\}^{\mathbb{N}}$ is countable. Then there exists a bijection $f : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$. In other words, we can label every element of $\{0,1\}^{\mathbb{N}}$ with a natural number. In general call $a_k = f(k)$ where $a_k \in \{0,1\}^{\mathbb{N}}$. For shorthand, as a_k is a sequence of zeros and ones, we call

$$a_k = a_{k1}a_{k2}a_{k3}a_{k4}\dots$$

where a_{kj} is shorthand for the jth term of sequence k. Let us list all elements of $\{0,1\}^{\mathbb{N}}$.

$$\begin{array}{cccccccc} & - & - & - & - & - & - & \\ a_1 : & a_{11} & a_{12} & a_{13} & a_{14} & \dots \\ a_2 : & a_{21} & a_{22} & a_{23} & a_{24} & \dots \\ a_3 : & a_{31} & a_{32} & a_{33} & a_{34} & \dots \\ a_4 : & a_{41} & a_{42} & a_{43} & a_{44} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Now, what we do is we look at the diagonal terms in the array above.

$$\begin{array}{cccccccc} & - & - & - & - & - & - & \\ a_1 : & a_{11} & a_{12} & a_{13} & a_{14} & \dots \\ a_2 : & a_{21} & a_{22} & a_{23} & a_{24} & \dots \\ a_3 : & a_{31} & a_{32} & a_{33} & a_{34} & \dots \\ a_4 : & a_{41} & a_{42} & a_{43} & a_{44} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

and we will define a sequence $b \in \{0,1\}^{\mathbb{N}}$ in the following manner.

$$b = b_1 b_2 b_3 \dots$$

where each term b_k is defined as

$$b_k = \begin{cases} 1 & \text{if } a_{kk} = 0 \\ 0 & \text{if } a_{kk} = 1 \end{cases}$$

In other words, b is a sequence formed by taking opposite values on the diagonal terms. Now, here is the beautiful thing, we have that $b \neq a_k$ for any $k \in \mathbb{N}$ as b and a_k differ at the kth term as sequences, i.e. $b_k \neq a_{kk}$. However, the assumption was that $\{0,1\}^{\mathbb{N}}$ is countable, i.e. that $f : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$ is bijective. Thus, there must exist some $j \in \mathbb{N}$ such that $b = f(j) = a_j$. This is a contradiction. Thus, $\{0,1\}^{\mathbb{N}}$ is uncountable. [8] □

[8] An alternate way this is often presented is: if $f : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$ is an injective map, then the diagonal argument gives a method of creating a sequence outside the image of f. Because of this $f(\mathbb{N})$ is always a proper subset of $\{0,1\}^{\mathbb{N}}$ for any injection, and hence, must be 'smaller'.

Exercises for section 1.2:

1. Show that $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined by $f(m, n) = 2^{m-1}(2n - 1)$ is a bijective function.
2. Create an injective map $f : \mathbb{N}^k \to \mathbb{N}$ and show that your creation is injective. *Hint*: For $k = 3$, $f(x, y, z) = 2^x 3^y 5^z$.
3. Show that the algebraic numbers are countable by making use of the following:
   (a) For a polynomial of degree n with integer coefficients,
   $$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$
   explain without proof the cardinality of the set $\{a \mid p(a) = 0\}$.
   (b) Create a surjection from $\mathbb{Z}^{n+1} \times \{1, 2, \ldots, n\}$ to zeroes of any nth degree integer coefficient polynomial. Explain why $\mathbb{Z}^{n+1} \times \{1, 2, \ldots, n\}$ is countable by quoting theorems in this section.
   (c) Explain why the set of algebraic numbers is countable using the prior part and quoting a theorem in this section.
4. Show that if a countable subset is removed from an uncountable set, the remainder is still uncountable.

1.3 Field Axioms and Totally Ordered Fields

As we saw in section 1.1, the number systems $\mathbb{N}, \mathbb{Z}$, and $\mathbb{Q}$ carry with them algebraic operations that help in the solving of equations. In this section, we want to formally describe and give names to these properties. This is a topic we touch on quickly as one of our goals in the construction of the real numbers is to show it inherits the algebraic properties that $\mathbb{Q}$ has, and these are topics that can be seen in more detail in a future class on abstract algebra. With that said, some proofs in this section will be put into the appendix. Most of this section follows the discussion about fields in [R].

Definition 10. *A* **binary operation** *on a set X is a function from $X \times X$ to X, i.e. it is a function that takes in a pair of elements from X and outputs a single element in X.*

In particular for us, we will focus on the binary operations of $+$ and $\cdot$ on sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$. It is implicit in the definition of a binary operation on X that the operation is **closed** on X, i.e. that the operation will take pairs from X to a result that stays within X. For some enlightening non-examples:

- Subtraction $-$ is **not** a binary operation on $\mathbb{N}$ as it is not closed. The numbers 2 and 7 are natural numbers, but $2 - 7$ is not.
- Division $\div$ is **not** a binary operation on $\mathbb{Z}$ as it is not closed. The numbers 2 and 5 are whole numbers, but $\frac{2}{5}$ is not.

We now come to the **field axioms**, which is a collection of properties we require of an abstract number system X (with operations $+, \cdot$) to guarantee we will be able to solve most abstract equations.

Definition 11. *A set X with binary operations $+$ and $\cdot$, often labeled by $(X, +, \cdot)$ is called a* **field** *if it has the following properties.*

(A1) *The operation of addition $+$ is* **associative**, *i.e. for $x, y, z \in X$ we have*

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

(A2) *There exists an* **additive identity**, *0, such that for all $x \in X$,*

$$x + 0 = 0 + x = x$$

(A3) *Every element $x \in X$ has an* **additive inverse** *labeled by $-x$ with*

$$x + (-x) = (-x) + x = 0$$

(A4) *Addition is* **commutative**, *i.e. for $x, y \in X$,*

$$x + y = y + x$$

(M1) *The operation of multipllication $\cdot$ is* **associative**, *i.e. for $x, y, z \in X$ we have*

$$x(yz) = (xy)z = xyz$$

(M2) *There exists a* **multiplicative identity**, *1, such that for all* $x \in X$

$$x \cdot 1 = 1 \cdot x = x$$

(M3) *Multiplication is* **commutative**, *i.e. for* $x, y \in X$,

$$xy = yx$$

(M4) *For all* $x \in X$ *with* $x \neq 0$ *there is a* **multiplicative inverse** *labeled by* $\frac{1}{x}$ *with*

$$x \cdot \frac{1}{x} = \frac{1}{x} \cdot x = 1$$

(D) *Multiplication distributes over addition, i.e. for* $x, y, z \in X$ *we have*

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

Many of these properties are things we were likely taught in grade school and have taken for granted for most of our mathematical careers, and while we will stay in the safe waters of fields, be careful, in the mathematical world there are objects that do not satisfy even the simplest of these properties. [9]

The rational numbers $\mathbb{Q}$ form a field, and our goal in the construction of the real numbers will be to show that it is a field as well. [10]. As you will see in future algebra classes there are many examples of fields beyond the rational and real numbers, but these will suffice for our study of analysis.

Remark 1. *It should be noted that we often simplify the formal notation above in the following manner*

$$x \cdot x = x^2, \quad x \cdot \frac{1}{y} = \frac{x}{y}, \quad x + (-y) = x - y$$

But, the rationals actually have even more structure besides being a field (as it will be with $\mathbb{R}$ as well) and that is an **order** structure.

Definition 12. *A set* X *is called* **totally ordered** *(alternatively called* **linearly ordered***) if the order* $<$ *on* X *satisfies*

- *the* **trichotomy law***: for* $a, b \in X$ *only one of the following holds*

$$a < b, \quad a = b, \quad a > b$$

- *And the order is* **transitive***: for* $a, b, c \in X$

$$a < b \quad \textit{and} \quad b < c \implies a < c$$

[9] For example, the Octonion numbers are not associative. You can read more about them here [JB]. *Hic sunt dracones*

[10] A set satisfying A1 and A2 is called a *monoid*, for example $\mathbb{N}$ is a monoid. A set satisfying A1-3 is called a *group* and A1-4 an *Abelian group*. A set satisfying all of these properties except M4 is a *commutative ring*, for example $\mathbb{Z}$ is a commutative ring

And of course a natural question to ask for a set X with an order structure $<$ and two binary operations $+, \cdot$ is how these structures interact. With that said we have

Definition 13. *A field $(X, +, \cdot)$ is called a* **totally ordered field** *(sometimes just an ordered field) if the following properties hold for $a, b, c \in X$*

- *If $a < b$, then $a + c < b + c$.*
- *If $a > 0$ and $b > 0$ then $ab > 0$.*

Elements $a \in X$ with $a > 0$ are called positive.

In particular the rational numbers are a totally ordered field, and we will show that the real numbers are a totally ordered field. We close out this section with a proposition containing properties any totally ordered field will satisfy. The proof of this proposition will be left to the appendix. 131

Proposition 6. *Let $a, b, c, d \in X$ with X a totally ordered field.*

i. If $a \neq 0$, then $a^2 > 0$. In particular, $1 > 0$.

ii. If $a > 0$, then $a^{-1} > 0$.

iii. If $a > b$ and $c > 0$, then $ac > bc$. If $a > b$ and $c < 0$, then $ac < bc$.

iv. If $0 < a < b$, then $0 < b^{-1} < a^{-1}$.

v. If $a < b$ and $c < d$, then $a + c < b + d$.

vi. If $0 < a < b$ and $0 < c < d$, then $a \cdot c < b \cdot d$.

vii. If $a > 0$ and $b > 0$ then $a + b > 0$.

Exercises for section 1.3:

1. Prove part vii. of Proposition 6.
2. Prove the following proposition 129 in the appendix.

1.4 Some last important properties

Lecture 3 - 10/2/24

We have seen that the rational numbers $\mathbb{Q}$ are a totally ordered field. In this section we will mention some of the remaining important properties that the rational numbers have before we move on to the formal construction of the real numbers.

Definition 14. *For x an element of the rational numbers, we define it's* **absolute value** *by*

$$|x| = \max(x, -x) = \begin{cases} x, & \text{if } x \geq 0 \\ -x, & \text{if } x < 0 \end{cases}$$

This definition agrees with how we typically think of absolute value in terms of measuring the distance of x from 0. A result that we will use time and time again in this course is the triangle inequality. This result tells us the interplay between absolute and addition depending on the order of the operations.

Theorem 7. The Triangle Inequality*: For rational numbers x and y we have*

a). $|x + y| \leq |x| + |y|$.

b) $||x| - |y|| \leq |x - y|$.

We will provide two proofs of part a), part b) is left as an exercise.

Proof. For our first proof of part a) of the triangle inequality, we exploit the definition of maximum, i.e. the definition of $|x|$ as a maximum means that $x \leq |x|$ and $-x \leq |x|$. The latter of these is equivalent to $-|x| \leq x$ from the proposition following the totally ordered field axioms, and as such we can combine both inequalities into one

$$-|x| \leq x \leq |x|$$

Doing the same thing for y: $-|y| \leq y \leq |y|$ and adding these inequalities gives

$$-|x| - |y| \leq x + y \leq |x| + |y|$$

and unpacking this inequality means

$$x + y \leq |x| + |y| \quad \text{and} \quad -(|x| + |y|) \leq x + y \iff -(x + y) \leq |x| + |y|$$

But then we have $|x + y| = \max(x + y, -(x + y)) \leq |x| + |y|$. □

Proof. For our second proof of part a), we proceed by cases using the piecewise definition of absolute value.

Case 1: If x, y are both positive, then $|x| = x$ and $|y| = y$ and as the sum of two positive numbers is positive, we have $x + y > 0$ and $|x + y| = x + y$, thus

$$|x + y| = x + y = |x| + |y|$$

Case 2: If x, y are both negative, then $|x| = -x$ and $|y| = -y$ and as the sum of two negative numbers is negative, we have $x + y < 0$ and $|x + y| = -(x + y)$, thus

$$|x + y| = -(x + y) = -x - y = |x| + |y|$$

Case 3: If x is positive and y is negative, then $-y$ is positive, which means we have the following by trichotomy $-y > 0 > y$. From here we have two subcases: *Subcase 1*: If $x \geq |y|$, then $x + y$ is nonnegative and thus $|x + y| = x + y$ and we have

$$|x + y| = x + y < x - y = x + (-y) = |x| + |y|$$

Subcase 2: If $x < |y|$, then $x + y$ is negtaive and thus $|x + y| = -x - y$ and we have

$$|x + y| = -x - y < x - y = x + (-y) = |x| + |y|$$

where the inequality is justified from x being positive.

Thus we see the triangle inequality holds in all cases. □

There is an extension of the triangle inequality to an arbitrary number of elements, and we will find need for it in this class so we provide a proof here.

Theorem 8. General Triangle Inequality for Finite Sums*: If $x_1, \ldots, x_n$ is a collection of n rational numbers, then*

$$\left|\sum_{k=1}^{n} x_k\right| \leq \sum_{k=1}^{n} |x_k|$$

Proof. The statement of the theorem is a collection of logical statements dependent on n

$$P(n): \qquad \left|\sum_{k=1}^{n} x_k\right| \leq \sum_{k=1}^{n} |x_k|$$

as such, it will be best to proceed by induction.

We first check the base case of $n = 1$ which is trivially true as it says $|x_1| \leq |x_1|$. We now assume that $P(n)$ holds and prove the result for $P(n+1)$. Beginning with the left hand side of our claim we have

$$\left|\sum_{k=1}^{n+1} x_k\right| = \left|\sum_{k=1}^{n} x_k + x_{n+1}\right| = |A + x_{n+1}|$$

if we let A denote the sum $\sum_{k=1}^{n} x_k$. By part a) of the prior theorem we have

$$\left|\sum_{k=1}^{n+1} x_k\right| = |A + x_{n+1}| \leq |A| + |x_{n+1}| = \left|\sum_{k=1}^{n} x_k\right| + |x_{n+1}|$$

If we now make use of our inductive hypothesis (assumption that $P(n)$ is true) then

$$\left|\sum_{k=1}^{n+1} x_k\right| \leq \left|\sum_{k=1}^{n} x_k\right| + |x_{n+1}| \leq \sum_{k=1}^{n} |x_k| + |x_{n+1}| = \sum_{k=1}^{n+1} |x_k|$$

where the inductive hypothesis was the second inequality above. This proves the statement $P(n+1)$, thus by the principle of mathematical induction, the result holds for all $n \in \mathbb{N}$. □

Another tool or property that will be used often in our future studies is the Archimedean Property. In the rational numbers, the first version of this result states that there is no smallest positive rational number in terms of the ordering $<$. The second equivalent form states that any positive rational number, no matter how large, is dominated by some natural number.

The Archimedean Property of $\mathbb{Q}$: The Archimedean property states

- For every positive rational number a there exists a natural number n with $0 < \frac{1}{n} < a$.
- Equivalently, for every positive rational number b, there exists a natural number n with $b < n$.

Proof. If $a > 0$ is a positive rational number, then we can assume that a is of the form $a = \frac{p}{q}$ with both p, q positive whole numbers. With this define $n = q + 1 \in \mathbb{N}$, then we have

$$na = (q+1)\frac{p}{q} = \frac{q+1}{q}p > p \geq 1$$

Thus as $na > 1$ we have $a > \frac{1}{n}$ and we know $\frac{1}{n} > 0$ as reciprocals of positive numbers are positive.

For the equivalent statement, if we let $a = \frac{1}{b}$, then the prior result gives $0 < \frac{1}{n} < \frac{1}{b}$, which is equivalent to $b < n$. □

Typically the Archimedean property is stated in another equivalent form: For x, y two positive rational numbers, there exists a natural number n such that $nx > y$. This is clearly equivalent to the second version above, but is closer to the original historical definition.

Another way to view the two equivalent forms of the Archimedean property is that a number system satisfying the Archimedean property has no **infinitely large** or **infinitesimal** (infinitely small) elements. The second version of the AP says that there is no positive rational larger than all natural numbers, and the first says there is no nonzero positive rational smaller than $\frac{1}{n}$ for all natural numbers n.[11]

Theorem 9. *For $x, y \in \mathbb{Q}$ with $x < y$, there exists a rational number z such that $x < z < y$.*

Proof. As $y - x > 0$, the Archimedean Property immediately gives the existence of n with $0 < \frac{1}{n} < y - x$. By adding x to all sides we have $x < x + \frac{1}{n} < y$. Taking $z = x + \frac{1}{n}$ proves the result. □

This theorem is often thought of as $\mathbb{Q}$'s density within itself, and we will come back to this when we speak about the density of $\mathbb{Q}$ in $\mathbb{R}$. It is interesting to note that this is where $\mathbb{Q}$ radically departs from $\mathbb{N}$ and $\mathbb{Z}$ in terms of number line structure.[12] Viewed on a number line, $\mathbb{N}$ and $\mathbb{Z}$ are what we would call **discrete** sets in that there is space between elements. And in terms of ordering on the number line, both $\mathbb{N}$ and $\mathbb{Z}$ have immediate predecessors and successors for elements (i.e. 4 is the immediate successor of 3, 2 is the immediate predecessor of 3).[13] But the rationals do not have either property: while the rational numbers are riddled with gaps, there is no space or room to breathe in the rational numbers in the sense that no matter how much you zoom in or get close to a rational number, there will always be other rationals close by. Put another way, there is no amount of zooming in on a rational number that will make it look discrete. Similarly for the order structure, the prior theorem says for a rational x, there is no 'next' rational number that comes after it.

With all of this said, we have found so far that

The rational numbers are a totally ordered field with the Archimedean property.

[11] This is typically the definition of an *infinitesimal number*. If you are curious about the *hyperreals* and their study of the collection of all infestimals about a point (*monad*) and the collection of all infinities (*galaxy*) then please check out the first chapter of [LW]. Note that the hyperreals $\mathbb{R}^*$ do not have the AP and are called *non-archimedean*.

[12] we saw the differences in the algebraic properties $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ have in the previous section.

[13] Well, not for 0 or 1 in $\mathbb{N}$ depending on how its defined.

As a last piece of motivation for why, even with all of the structure $\mathbb{Q}$ has, we will still need to move onto the real numbers, we give a few remaining definitions that will hellp describe these 'gaps' the rationals have.

Definition 15. *For a totally ordered set X with ordering $<$ and a subset $A \subseteq X$, we define the following:*

i). an element a is called an **upper bound** *if $x \leq a$ for all $x \in A$.*

ii). and element a is called a **lower bound** *if $x \geq a$ for all $x \in A$.*

We say that a set A is **bounded above/below** if the set has an upper bound/lower bound respectively.

Definition 16. *For a totally ordered set X with ordering $<$ and a subset A that is bounded above, then an element α satisfying the properties*

a). α is an upper bound of A.

b). Any $\beta < \alpha$ is not an upper bound of A

is called the **least upper bound** *or* **supremum** *of A, and is labeled* $\sup A$.

Definition 17. *For a totally ordered set X with ordering $<$ and a subset A that is bounded below, then an element α satisfying the properties*

a). α is a lower bound of A.

b). Any $\beta > \alpha$ is not a lower bound of A

is called the **greatest lower bound** *or* **infimum** *of A, and is labeled* $\inf A$.

Example 4. *If we look at the set $A = \{q \in \mathbb{Q} \mid q^2 < 2\}$, then this set is bounded above. In fact there are an infinite number of upper bounds, but this set has no least upper bound within $\mathbb{Q}$.*

Put another way, the least upper bound is $\sqrt{2}$ but this is not a rational number.

Following the example here, there are many subsets of $\mathbb{Q}$ that are bounded above that have no least upper bound within $\mathbb{Q}$.[14] This makes more precise this notion of 'gaps' in $\mathbb{Q}$. With the example above we can find rational numbers that get closer and closer to the ceiling of $\sqrt{2}$ but we can never say that we reach it, as $\sqrt{2}$ does not exist in $\mathbb{Q}$, it's just a hole.[15] So, with all of this said we come to our last axiom.

Least Upper Bound Property : A totally ordered set X has the least upper bound property if every nonempty subset A of X that is bounded above has a supremum that exists within X.[16]

A set is often called **complete** if it has the least upper bound property. The rationals $\mathbb{Q}$ are not complete.

[14] similar for subsets bounded below with no greatest lower bound.

[15] once again, this is one of the pitfalls of all of us having some familiarity with $\mathbb{R}$ before constructing it. We know that the ceiling is $\sqrt{2}$ here, but formally $\sqrt{2}$ does not exist yet to us. The process of building $\mathbb{R}$ will be a process of creating these numbers in terms of rationals and then filling in these holes.

[16] similarly one could say A bounded below and the infimum exists in X.

Completeness axiom/property: : The real numbers, $\mathbb{R}$, are complete.

And this is the one property that $\mathbb{Q}$ fails to have that will be paramount in the study of analysis, thus we will spend the next few sections making our way through a few tools we will need first before beginning the construction of $\mathbb{R}$ in earnest. In the end, we will show that

The real numbers $\mathbb{R}$ are a complete totally ordered field with the Archimedean property that contains $\mathbb{Q}$ as a dense subfield.[17].

And so $\mathbb{R}$ has the algebraic structure, the order structure, and the completeness we require to solve equations, define limits, and create the proper foundations for Calculus. [18]

Exercises for section 1.4:

1. Prove part b) of the Triangle inequality (Theorem 7). *Hint*: Start with

$$|x| = |x - y + y| \leq |x - y| + |y|$$

[17] In fact, a later result you will learn is that $\mathbb{R}$ is the only complete totally ordered field up to isomorphism

[18] It should be said that a second definition of the completeness of $\mathbb{R}$, is that every Cauchy sequence is convergent. However this requires that $\mathbb{R}$ has the Archimedean Property as well to be equivalent to the least upper bound property.

1.5 Summary

As an end to this section, before we move on to our study of sequences, let us look at a quick recovery of what we have found so far.

- Starting with $\mathbb{N}$, the jump to $\mathbb{Z}$ and then $\mathbb{Q}$ (and even $\mathbb{A}$) can be justified by wanting to solve more and more types of equations, and hence the study of algebra. However, the 'gaps' in these number systems keeps Calculus[19] from being possible, so this is why we desire the real numbers.
- The cost of filling in these gaps and having $\mathbb{R}$ is that the reals form an *uncountable* infinite. In the same way that any finite value takes up 0% of a countable infinite, similarly a countable infinite takes up 0% of an uncountable infinite.
- However, we don't want to throw everything away. The rationals $\mathbb{Q}$ form a *field* and in fact a *totally ordered field* with its order structure $<$. As this makes solving many types of equations possible, we want to maintain this when building the reals.
- Similarly, we want the real numbers to retain the Archimedean Property as we do not want infinite or infinitesimal numbers.
- Lastly, we formalized this notion of the 'gaps' in $\mathbb{Q}$ by defining supremums and infimums and stating the completeness axiom. We will show that $\mathbb{R}$ satisfies this axiom.

The construction of $\mathbb{R}$ is typically done either with *Dedekind cuts* or *Equivalence classes of Cauchy sequences*. The method of cuts to construct the reals was published by Dedekind in 1872, however notions of these 'cuts' existed in papers by Hamilton in 1833 and 1835. Cantor's construction of the reals using equivalence classes of Cauchy sequences was also put forward in 1972[20].

It is the author's opinion that Cantor's construction is more intutive for a first viewing of the construction of $\mathbb{R}$ and so that is what will follow. As such, our next section will be devoted to the study of sequences more generally before beginning the construction process. I will also try to present Dedekind's method at a later time, but this will be supplementary material and not 'required' reading for the course.

[19] i.e. our study of limiting and local behaviors of functions. I am 'technically' lying to you though.
[20] some references say 1871

2 Sequences of Rational Numbers

Lecture 4 - 10/4/24

Before we begin, as we have not formally crafted the real numbers as of yet, in this section when we talk about terms of sequences and values that they converge to, all values should be thought of as rational numbers. These definitions and theorems will still hold true in $\mathbb{R}$ once we build it, but for right now everything is in the rational numbers.

2.1 Definition of Sequences and Convergence

Definition 18. *A* **sequence** *of rational numbers is a function $f : \mathbb{N} \to \mathbb{Q}$. However, typically the function notation is suppressed and instead of writing $f(1) = a_1$, $f(2) = a_2$, ... we simply write the terms of the sequence*

$$a_1,\ a_2,\ a_3,\ \ldots \qquad a_k \in \mathbb{Q},\ \forall k \in \mathbb{N}$$

and will use the notation $\{a_k\}_{k\in\mathbb{N}}$ or $\{a_k\}_{k=1}^{\infty}$ to define a sequence (sometimes just $\{a_k\}$ if it is clear it is a sequence from context)

Example 5. *We have the following examples of sequences*

a). $\{a_n\} = \{(-1)^n\}$.

b). $\{b_n\} = \left\{\frac{1}{n}\right\}$.

c). $\{c_n\} = \left\{\frac{(-1)^n}{n}\right\}$.

The first sequence simply bounces back and forth between -1 and 1, but our prior knowledge from Calculus courses tells us that sequences b) and c) both converge to 0, even though b) is made up of entirely positive terms and c) bounces between positive and negative.

Some sequences have the nature of clustering/converging around a single point while other sequences can cluster around many points[21] or cluster nowhere at all. As it will turn out, the method of building $\mathbb{R}$ that we will pursue will close those 'gaps' in $\mathbb{Q}$ by defining numbers that do not exist in the rationals by sequences of rational numbers that approach that value. Because of this, it will be very important that we have a solid definition of this clustering/converging process.

Definition 19. *A sequence $\{a_n\}$ of rational numbers is said to* **converge** *to a rational number $a \in \mathbb{Q}$ if for all $\epsilon \in \mathbb{Q}_+$ there exists $N \in \mathbb{N}$ such that for all $n > N$ we have*

$$|a_n - a| < \epsilon$$

Written more compactly in quantifiers, this will be

$$\forall \epsilon \in \mathbb{Q}_+, \exists N \in \mathbb{N},\ \forall n > N, |a_n - a| < \epsilon$$

In this case we often write $a_n \to a$ as $n \to \infty$ or $\lim_{n\to\infty} a_n = a$. If a sequence does not converge it is said to **diverge**.

Remark 2. *Later once we have constructed the reals, we will replace the condition $\forall \epsilon \in \mathbb{Q}_+$ with $\forall \epsilon > 0$.*

[21] example a) above

In the definition of convergence, it is good to think of ϵ as a choice of 'error' made by a user of this convergence algorithm. The value N is dependent upon ϵ[22] and it measures a 'threshold limit' for this particular choice of ϵ. From our properties about absolute value we have that the condition

$$|a_n - a| < \epsilon \iff a - \epsilon < a_n < a + \epsilon$$

can be written in the equivalent form on the right. This is saying for $n > N$ that a_n is within the 'error window' $(a - \epsilon, a + \epsilon)$ of a. So, our definition of convergence is describing the feedback or response relationship between ϵ and N we would intuitively expect if $\{a_n\}$ was clustering around a. In other words, for a convergent sequence $\{a_n\}$, once an ϵ has been chosen, a positive whole number N exists, and it tells you *how far into the terms of the sequence you must go for the remaining terms to be within the chosen error, ϵ, of the limit.*

Now, this may seem like an overly complicated definition of convergence, especially after years of computing limits in a Calculus course. The lie is that the limit exercises in prior Calculus classes were chosen from convergent sequences to build up ones computational skill with limit methods, and finding the destination of something you know converges is very different from guaranteeing a destination exists at all. [23] But the answer for why 'this' definition is really that, once again, we have to get around the notion of infinity. We know $\{\frac{1}{n}\}$ converges to 0, but we really don't have the time to wait for it to 'get there'. [24] We're finite beings, waiting around for a countably infinite process to occur would be very silly indeed.

We get around 'infinity' in this case by stating information about **tails** of a sequence. For a sequence $\{a_n\}_{n=1}^{\infty}$, a *tail* of this sequence is simply $\{a_n\}_{n=N_0}^{\infty}$, i.e. starting the sequence from a later term and then continuing on. From this we can see two other ways that convergence of a sequence is often stated by:

- A sequence $\{a_n\}$ converges to a if for any $\epsilon > 0$ there exists a tail of $\{a_n\}$ that lies in the error window $(a - \epsilon, a + \epsilon)$.
- A sequence $\{a_n\}$ converges to a if for any $\epsilon > 0$, the sequence $\{a_n\}$ is *eventually* in the error window $(a - \epsilon, a + \epsilon)$.

And in any of the three ways of thinking about convergence it is the feedback relationship between ϵ and the threshold N (or start of the tail N) that lets us say where a sequence is going even though we do not have the time to wait for it to 'get there'.

Example 6. *The following are some examples of sequences and their convergence properties:*

a). The constant sequence $\{a_n\} = \{49\}$.

The constant sequence clearly converges to the constant 49. For any $\epsilon \in \mathbb{Q}_+$, for $N = 1$ we have that for all $n > N = 1$,

$$0 = |49 - 49| = |a_n - 49| < \epsilon$$

And so this satisfies our definition of convergence, thus $a_n \to 49$ as $n \to \infty$.

b). The sequence $\{b_n\} = \{\frac{1}{n}\}$.

[22] in fact some people define $N(\epsilon)$ to be a natural number valued function

[23] Bob is being a bit overdramatic here. Intuition built from solving limits will still prove very invaluable here

[24] It never actually does. Some may say it reaches 0 when $n = \infty$ but ∞ is not a natural number. However, and I may put this in an appendix, in a nonarchimedean set like the hyperreals you can make some sense of this limit actuallly reaching it's destination.

We conjecture that this sequence converges to 0. Before we begin with the proof, let us do some scratch-work first. Let us start with the condition we aim to show, which is $|a_n - 0| < \epsilon$ for $n > N$, and see what algebraic relationship it implies between N and ϵ.

$$\frac{1}{n} = |\frac{1}{n} - 0| = |a_n - 0| < \epsilon$$

Thus we see the condition $|a_n - 0| < \epsilon$ will occur when $n > \frac{1}{\epsilon}$. Because of this, a choice of N is any natural number larger than $\frac{1}{\epsilon}$.[25] *Now let us see the proof.*

Proof. Let $\epsilon \in \mathbb{Q}_+$. For this epsilon, by the Archimedean property there exists an $N \in \mathbb{N}$ such that $\frac{1}{\epsilon} < N$. This means that for all $n > N$, we have $\frac{1}{n} < \frac{1}{N} < \epsilon$ and

$$|a_n - 0| = \left|\frac{1}{n} - 0\right| = \frac{1}{n} < \epsilon$$

And since this argument can be done for any choice of $\epsilon \in \mathbb{Q}_+$, we see that $\{a_n\}$ satisfies the definition of convergence to the value 0. □

c). The sequence $\{a_n\} = \{1 + \frac{1}{2n}\}$.

From our intuition in calculus, we know that as $n \to \infty$ that $\frac{1}{2n} \to 0$, thus we conjecture that $\{a_n\} \to 1$. So, first we will perform some scratchwork before writing down the exact proof.

SCRATCHWORK*: For the moment, imagine that ϵ is a positive, fixed, incredibly small number. And let's just start with what we would like to have happen, which is $|a_n - 1| < \epsilon$, and see if this gives a condition or a hint to what N should be.*

$$\begin{aligned} |a_n - 1| &< \epsilon \\ \left|1 + \frac{1}{2n} - 1\right| &< \epsilon \\ \frac{1}{2n} &< \epsilon \\ \frac{1}{2\epsilon} &< n \end{aligned}$$

where I have used that the absolute value of a positive number is itself. But, we have found our condition in our last step that hints at what N should be, which is $N = \lceil \frac{1}{2\epsilon} \rceil$ (this is the ceiling function, i.e. the next whole number larger than 1 over 2ϵ). Now, with the condition found we write out the proof.

Proof. Let $\epsilon > 0$ be a fixed number. Define $N = \lceil \frac{1}{2\epsilon} \rceil$. Then we have that for any $n > N$ that

$$n > N = \left\lceil \frac{1}{2\epsilon} \right\rceil = \frac{1}{2\epsilon}$$

and this means that $|a_n - 1| < \epsilon$ for all $n > N$.

As such an N exists for any chosen $\epsilon > 0$, we have then showed that $\lim_{n\to\infty} a_n = 1$. □

d). The sequence $\{b_n\} = \{2 - \frac{1}{n^2}\}$.

Once again, with our intuition from Calculus we conjecture that this sequence converges to 2.

[25] which the Archimedean property guarantees

Proof. Let ϵ be an arbitrary fixed positive number, and define $N = \lceil \frac{1}{\sqrt{\epsilon}} \rceil$. Then, for all $n > N$ we have that

$$n > N = \left\lceil \frac{1}{\sqrt{\epsilon}} \right\rceil > \frac{1}{\sqrt{\epsilon}}$$

And from this we have

$$n > \frac{1}{\sqrt{\epsilon}} \implies n^2 > \frac{1}{\epsilon}$$
$$\frac{1}{n^2} < \epsilon \implies \left| -\frac{1}{n^2} \right| < \epsilon$$
$$\left| 2 - \frac{1}{n^2} - 2 \right| < \epsilon$$

which is $|b_n - 2| < \epsilon$.

Thus we have shown for this ϵ the existence of N with the property that for all n with $n > N$ that $|b_n - 2| < \epsilon$. And this process can be done for any choice of ϵ, thus it is true for all $\epsilon > 0$.

Thus $\{b_n\} \to 2$. □

e). The sequence $\{c_n\} = \left\{ \frac{2+3n}{5n-1} \right\}$.

SCRATCHWORK*: We know from prior study of rational functions that* $f(x) = \frac{2+3x}{5x-1}$ *has a horizontal aymptote of* $y = \frac{3}{5}$*, so we expect that this is what the sequence* $\{c_n\}$ *tends toward. So for an arbitrary* $\epsilon > 0$*, let us look at what would be necessary of* N *for the condition*

$$|c_n - \frac{3}{5}| < \epsilon$$

to hold for $n > N$*.*

By direct computation we have

$$|c_n - \frac{3}{5}| = \left| \frac{2+3n}{5n-1} - \frac{3}{5} \right| = \left| \frac{2+3n}{5n-1} - \frac{3(n-\frac{1}{5})}{5(n-\frac{1}{5})} \right| = \left| \frac{2+3n-3n+\frac{3}{5}}{5n-1} \right| = \left| \frac{13}{5(5n-1)} \right| = \left| \frac{13}{25(n-\frac{1}{5})} \right|$$

Thus we will have $|c_n - \frac{3}{5}| < \epsilon$ *precisely when*

$$n - \frac{1}{5} > \frac{13}{25\epsilon} \iff n > \frac{13}{25\epsilon} + \frac{1}{5}$$

Thus one could take $N = \lceil \frac{13}{25\epsilon} + \frac{1}{5} \rceil \in \mathbb{N}$.

Proof. Let $\epsilon \in \mathbb{Q}_+$, then there exists $N = \lceil \frac{13}{25\epsilon} + \frac{1}{5} \rceil \in \mathbb{N}$ such that for all $n > N$ we have

$$|c_n - \frac{3}{5}| < \epsilon.$$

As this can be done for any choice of positive rational epsilon, we have that $c_n \to \frac{3}{5}$. □

f). For $0 < r < 1$*, the sequence* $\{f_n\} = \{r^n\}$ *converges to 0.*

This will be left as an exercise.

g). For r with $|r| < 1$, the sequence $\{g_n\}$ defined by

$$g_n = 1 + r + r^2 + r^3 + \cdots + r^n = \sum_{k=0}^{n} r^k$$

converges to $\frac{1}{1-r}$.

This follows from part f). If we look at the following

$$rg_n = r + r^2 + r^3 + \cdots + r^{n+1}$$
$$g_n = 1 + r + r^2 + \cdots + r^n$$

and subtract, we see that $rg_n - g_n = r^{n+1} - 1$, and thus $g_n = \frac{r^{n+1}-1}{r-1}$. And the expression

$$\left| g_n - \frac{1}{1-r} \right| = \left| \frac{r^{n+1}-1}{r-1} + \frac{1}{r-1} \right| = \frac{|r|^{n+1}}{|r-1|}$$

and as $|r|^m \to 0$ as $m \to \infty$ from part f), this suffices to show $g_n \to \frac{1}{1-r}$.

h). The sequence $\{h_n\} = \{(-1)^n\}$ diverges.

We will see the reason for this later in our section on subsequences.

i). The sequence $\{j_n\} = \{n^2\}$ diverges to ∞.

Definition 20. *A sequence $\{x_n\}$* diverges to ∞ *if for all $M > 0$, there exists an $N \in \mathbb{N}$ such that for all $n > N$ we have $x_n > M$. Similarly, a sequences* diverges to $-\infty$ *if for all $M < 0$ there exists an $N \in \mathbb{N}$ such that for all $n > N$ we have $x_n < M$.*

As we saw in i). of the last example, for an arbitrary $M > 0$, if we take $N = \sqrt{M}$, then for all $n > N$, we have that $n^2 > M$. This shows that $\{x_n\}$ diverges to ∞.

Exercises for section 2.1:

1. Find what the following sequences converge to and prove they indeed converge to your claimed value.

   a).
   $$\{x_n\} = \left\{ 3 + \frac{(-1)^n}{2n} \right\}$$

   b).
   $$\{x_n\} = \left\{ \frac{2n-5}{5n+7} \right\}$$

   c).
   $$\{x_n\} = \left\{ \frac{2n^2+3n}{n+n^2} \right\}.$$

   d).
   $$\{x_n\} = \left\{ \frac{1}{n^2} + \frac{2}{n^2} + \cdots + \frac{n}{n^2} \right\}$$

2. (From [FM]) Let $h > 0$ be fixed. Prove (either by induction or using the binomial theorem[26] that
$$(1+h)^n \geq 1 + nh, \qquad n = 1, 2, \ldots$$
Deduce that, if $0 < r < 1$, then $\lim_{n\to\infty} r^n = 0$. [Hint: explain why you can write $r = \frac{1}{1+h}$ for some $h > 0$.]

[26]The binomial theorem states that for all $x, y \in \mathbb{Q}$ and $n \in \mathbb{N}$, $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

2.2 Sequences and Order & Algebraic Limit Rules

In this section we will collect many important results about convergent sequences. Our first result states that if a sequence $\{a_n\}$ converges to a limiting value L, then this value is unique. In other words, it is impossible for a sequence to converge to two distinct values.

Theorem 10. Uniqueness of limits: *If a sequence $\{a_n\}$ converges to a limit L, then this limit is unique.*

Proof. We will assume that $a_n \to L$ and $a_n \to M$. [27] Let us now take a fixed $\epsilon \in \mathbb{Q}_+$. As $a_n \to L$, by the definition of convergence there exists an $N_1 \in \mathbb{N}$ such that for all $n > N_1$

$$|a_n - L| < \frac{\epsilon}{2}.$$

Similarly, as $a_n \to M$ there exists an $N_2 \in \mathbb{N}$ such that for all $n > N_2$

$$|a_n - M| < \frac{\epsilon}{2}.$$

Thus if we take $N = \max(N_1, N_2)$ then we can assume both $|a_n - L|, |a_n - M|$ are less than $\frac{\epsilon}{2}$ when $n > N$. But then, from the triangle inequality we have

$$|L - M| = |L - a_n + a_n - M| \leq |a_n - L| + |a_n - M| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Thus $|L - M| < \epsilon$ and this can be done for any $\epsilon \in \mathbb{Q}_+$, thus it must be that $|L - M| = 0$[28], so $L = M$. □

Lecture 5 - 10/7/24

Let us now see how convergence of sequences interact with the order structure, $<$, on $\mathbb{Q}$.

Theorem 11. *Given two sequences $\{x_n\}$ and $\{y_n\}$ that are convergent with $x_n \to x$ and $y_n \to y$, if $x_n \geq y_n$ for all $n \in \mathbb{N}$, then $x \geq y$.*

Before beginning this proof it should be noted that the condition of $x_n \geq y_n$ for all $n \in \mathbb{N}$ can be laxed. It is only required that $x_n \geq y_n$ on some common tail of both sequences, i.e. that there exists some $M \in \mathbb{N}$ such that $x_n \geq y_n$ for all $n > M$. This is because of the nature of convergent sequences. Once again, any finite number of terms from the beginning of a sequence is effectively meaningless from the perspective of a sequence, its convergent or divergent character is determined by the infinite tails.

Proof. We proceed by contradiction, thus we assume that $x_n \to x$, $y_n \to y$, that $x_n \geq y_n$ for all $n \in \mathbb{N}$ and that $x < y$. As $y > x$, we have that $\frac{y-x}{2}$ is a positive rational number, thus let us take $\epsilon = \frac{y-x}{2}$. By the assumption that $x_n \to x$, for this ϵ there is a $N_1 \in \mathbb{N}$ such that for $n > N_1$

$$|x_n - x| < \frac{y-x}{2}, \iff x - \frac{y-x}{2} < x_n < x + \frac{y-x}{2}$$

[27] technically I should say $L, M \in \mathbb{Q}$ but this proof is true in $\mathbb{R}$ (and actually any metric space)
[28] otherwise this would contradict the Archimedean Principle in $\mathbb{Q}$

Similarly, for this ϵ there is a $N_2 \in \mathbb{N}$ such that for $n > N_2$,

$$|y_n - y| < \frac{y-x}{2}, \iff y - \frac{y-x}{2} < y_n < y + \frac{y-x}{2}$$

However, this implies for any $n > \max(N_1, N_2)$ that

$$x_n < x + \frac{y-x}{2} = \frac{x+y}{2} = y - \frac{y-x}{2} < y_n$$

And so $x_n \geq y_n$ for all $n \in \mathbb{N}$ and $x_n < y_n$ for $n > \max(N_1, N_2)$, which is a clear contradiction. □

Lemma 12. The Squeezing/Sandwich Lemma: *Suppose for two sequences $\{a_n\}$ and $\{b_n\}$ that $0 \leq b_n \leq a_n$ is true for all $n \in \mathbb{N}$.*[29] *If $a_n \to 0$, then $b_n \to 0$.*

The Squeezing Lemma Generalization: *For sequences $\{a_n\}$, $\{b_n\}$, and $\{c_n\}$ with the property that $a_n \leq b_n \leq c_n$ for all $n \in \mathbb{N}$, if $\{a_n\}$ and $\{c_n\}$ are both convergent with $\lim_{n\to\infty} a_n = L = \lim_{n\to\infty} c_n$, then $\{b_n\}$ is convergent and $\lim_{n\to\infty} b_n = L$.*

Proof. Left as an exercise. □

Definition 21. *A sequence $\{x_n\}$ is called* **bounded** *if the terms of sequence, seen as a set*

$$A = \{x_1, x_2, x_3, \dots\}$$

form a bounded set (bounded above and below). Equivalently, the sequence is called bounded if there exists a constant M such that

$$|x_n| \leq M, \quad \forall n \in \mathbb{N}$$

Or, equivalently, a sequence is called bounded if there exists constants M, B such that

$$B \leq x_n \leq M, \quad \forall n \in \mathbb{N}$$

If only M exists, then the sequence is called **bounded above**, *and if only B exists, the sequence is called* **bounded below**.

Theorem 13. *If a sequence $\{a_n\}$ converges, then it is bounded.*

Proof. This will be a constructive argument where we create an explicit bound on the sequence $\{a_n\}$. As we are assuming that $\{a_n\}$ converges, there is a $a \in \mathbb{Q}$ such that $a_n \to a$. Now, take ϵ to be some fixed rationa number that is positive. For this ϵ as $a_n \to a$, the definition of convergence furnishes us with an $N \in \mathbb{N}$ such that for all $n > N$ we have

$$|a_n - a| < \epsilon$$

but this can be equivalently stated

$$a - \epsilon < a_n < a + \epsilon, \quad \forall n > N$$

And this is the crux of our proof, we have just shown that a tail of the sequence is bounded. As we also have that

$$-a - \epsilon < -a_n < \epsilon - a, \quad \forall n > N$$

[29] once again, really only needs to hold on a tail of the sequences

this shows that $|a_n| = \max(a_n, -a_n) < \max(a + \epsilon, \epsilon - a)$ for all $n > N$. Using this we define

$$M = \max(|a_1|, |a_2|, \ldots, |a_N|, \epsilon - a, a + \epsilon)$$

which can be done because this is only a finite number of terms we are searching for the maximum of. With this choice of M we see,

$$|a_n| \leq M, \quad \forall n \in \mathbb{N}$$

and this shows that the sequence is bounded. □

The sequence $\{(-1)^n\}$ is bounded as $|(-1)^n| \leq 2$ for all n, but as we will shortly see, this sequence is not convergent. Thus in general we say that the boundedness of a sequence is a *weaker* notion than convergence, in that convergence implies boundedness but not the other way around. We will see later that is a condition that can be added to boundedness to ensure a sequence converges, and that boundedness of a sequence will ensure the existence of a convergent subsequence.[30]

Theorem 14. (Algebraic Limit Rules) *Given $\{a_n\}$ and $\{b_n\}$ convergent sequences, i.e. $a_n \to a$ and $b_n \to b$, and any constants $\alpha, \beta \in \mathbb{Q}$, we have the following*

i). $(\alpha a_n + \beta b_n) \to \alpha a + \beta b$

ii). $(a_n b_n) \to ab$, and note that this also implies $ca_n \to ca$ for any constant c.

iii). If $b \neq 0$, there exists $N \in \mathbb{N}$ such that for $n > N$, $b_n \neq 0$. Also it then follows that

$$\frac{1}{b_n} \to \frac{1}{b}, \quad \textit{and} \quad \frac{a_n}{b_n} \to \frac{a}{b}$$

Proof. We start with the proof of i.) If both $\alpha = \beta = 0$, then there is nothing to prove. If either α or β is 0, then the proof will follow from part ii). Thus we will assume that both $\alpha, \beta \neq 0$.

We now let $\epsilon > 0$ be a fixed positive rational number. For this ϵ, as $x_n \to x$, there exists $N_1 \in \mathbb{N}$ such that for all $n > N_1$ we have

$$|x_n - x| < \frac{\epsilon}{2|\alpha|}$$

Similarly, for this ϵ, as $y_n \to y$, there exists $N_2 \in \mathbb{N}$ such that for all $n > N_2$ we have

$$|y_n - y| < \frac{\epsilon}{2|\beta|}$$

Then for all $n > \max(N_1, N_2)$, we have

$$\begin{aligned} |\alpha x_n + \beta y_n - (\alpha x + \beta y)| = |\alpha(x_n - x) + \beta(y_n - y)| &\leq |\alpha||x_n - x| + |\beta||y_n - y| \\ &< |\alpha|\frac{\epsilon}{2|\alpha|} + |\beta|\frac{\epsilon}{2|\beta|} = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

No part of this argument relies upon any specific property of ϵ other than it being positive, thus this argument holds for all $\epsilon > 0$, thus we have $(\alpha x_n + \beta y_n) \to (\alpha x + \beta y)$. □

[30] The monotone convergence theorem and the Bolzano-Weierstass theorem respectively

Proof. For part ii). let us first look at the following expression to make a gameplan for our argument.

$$|x_n y_n - xy| = |x_n y_n - x_n y + x_n y - xy| = |x_n(y_n - y) + y(x_n - x)| \leq |x_n||y_n - y| + |y||x_n - x|$$

As we know that $x_n \to x$ and $y_n \to y$ we know we can control the error in $|x_n - x|$ and $|y_n - y|$. Now for the term $|y||x_n - x|$, this can be controlled (made small) as $|y|$ is a fixed quantity. With the term $|x_n||y_n - y|$, both terms in the product are dependent on n, and while we know that $|y_n - y|$ can be made small, this does not guarantee $|x_n||y_n - y|$ is necessarily small.

For example, $\frac{1}{n}$ goes to zero as $n \to \infty$, but $1 = n\left(\frac{1}{n}\right)$ for all $n \in \mathbb{N}$, thus it is very possible for a sequence that is converging to zero to be multiplied by another sequence and never shrink to 0. The saving grace here will be that $\{x_n\}$ is bounded because it is convergent, and this will let us control the expression $|x_n||y_n - y|$. On to the proof.

So, let $\epsilon > 0$. As $x_n \to x$, from a previous theorem we have that $\{x_n\}$ is bounded, thus there is a constant M such that

$$|x_n| \leq M, \quad \forall n \in \mathbb{N}$$

We then take $J = \max(M, |y|)$, the maximum of M and $|y|$, for reasons that will be apparent later.

For this ϵ, there exists an N_1 such that for all $n > N_1$ we have

$$|x_n - x| < \frac{\epsilon}{2J}$$

Similarly, for this $\epsilon > 0$, there exists an N_2 such that for all $n > N_2$ we have

$$|y_n - y| < \frac{\epsilon}{2J}$$

And thus for $n > \max(N_1, N_2)$ we have the following

$$\begin{aligned}
|x_n y_n - xy| &= |x_n y_n - x_n y + x_n y - xy| = |x_n(y_n - y) + y(x_n - x)| \\
&\leq |x_n||y_n - y| + |y||x_n - x| && \text{Triangle Inequality} \\
&\leq M|y_n - y| + |y||x_n - x| && \text{Bound on } \{x_n\} \\
&< M\frac{\epsilon}{2J} + |y|\frac{\epsilon}{2J} && \text{for } n > \max(N_1, N_2) \\
&= \left(\frac{M + |y|}{2J}\right)\epsilon < \epsilon
\end{aligned}$$

No part of this argument relies upon any specific property of ϵ other than it being positive, thus this argument holds for all $\epsilon > 0$, thus we have $x_n y_n \to xy$.

If we take $\{x_n\}$ to be the constant sequence $\{c\}$ for a constant c, then the result we just proved shows that $cx_n \to cx$. □

Proof. For part iii). As $y_n \to y$ for $\epsilon = \frac{|y|}{2}$ we know there exists an $N \in \mathbb{N}$ such that for all $n > N$ we have $|y_n - y| < \frac{|y|}{2}$. From the reverse triangle inequality, we have

$$|y| - |y_n| \leq ||y_n| - |y|| \leq |y_n - y| < \frac{|y|}{2}$$

which shows that $|y_n| > \frac{|y|}{2}$ for all $n > N$. This shows that $y_n \neq 0$ for $n > N$ and completes the first part of the proof. We now have the following computation

$$\left|\frac{1}{y} - \frac{1}{y_n}\right| = \left|\frac{y_n - y}{y_n y}\right| = \frac{1}{|y_n||y|}|y_n - y|$$

Which using the inequality we just found gives

$$\left|\frac{1}{y}-\frac{1}{y_n}\right| < \frac{2}{|y|^2}|y_n - y| \quad \text{for } n > N$$

And by pushing to a perhaps larger value of N we could have $|y_n - y| < \frac{|y|^2\epsilon}{2}$ for all $n > N$ and this would show

$$\left|\frac{1}{y}-\frac{1}{y_n}\right| < \epsilon, \quad \text{for } n > N$$

and this shows the result.

If this is not satisfying to you, the other way this could be seen is that $y_n \to y$ is equivalent to saying $|y_n - y| \to 0$. Using this, the squeeze lemma with

$$\left|\frac{1}{y}-\frac{1}{y_n}\right| < \frac{2}{|y|^2}|y_n - y|$$

shows that $\left|\frac{1}{y}-\frac{1}{y_n}\right| \to 0$ which means $\frac{1}{y_n} \to \frac{1}{y}$.

The general result for $\frac{x_n}{y_n} \to \frac{x}{y}$ now follows from what we have just proven and part ii). with $\frac{x_n}{y_n} = x_n\left(\frac{1}{y_n}\right)$. □

These algebraic limit laws are very useful in making more general arguments of sequence convergence without needing to rely upon the original definition of convergence (the ϵ-N arguments). For example, the result above shows that if $x_n \to x$, then $x_n^2 \to x^2$ and similarly for other powers by an inductive argument. More generally, for a polynomial $p(a)$ with rational coefficients, it is true that $p(x_n) \to p(x)$, and this will be left as a homework exercise (as well as a similar result for rational functions)

Exercises for section 2.2:

1. Prove or disprove with counterexample the following variation of Theorem 11: Given two sequences $\{x_n\}$ and $\{y_n\}$ that are convergent with $x_n \to x$ and $y_n \to y$, if $x_n > y_n$ for all $n \in \mathbb{N}$, then $x > y$.

2. Prove both parts of Lemma 12.

3. Conjecture the value of the following and provide a proof of its convergence (if it does?)

$$\lim_{n\to\infty} \sqrt{n^2+6n} - n$$

Hint: Finding a lower bound on a term in a denominator give an upper bound on its reciprocal, i.e. $f(x) > M$ means $\frac{1}{f(x)} < \frac{1}{M}$.

4. a). Suppose that the sequence $\{x_n\}$ is bounded and that $y_n \to 0$. Prove that $\{x_n y_n\}$ converges to 0.

   b). Give an example in which $y_n \to 0$, and $x_n y_n$ does not converge to 0.

5. a). Suppose that $x_n \to x$, and define y_n to be the sequence given by

$$y_n = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

i.e. the arithmetic mean of the first n terms. Show that $y_n \to x$.

b). Let $\{x_n\} = \{(-1)^n\}$. We know that x_n diverges, but show that y_n in this instance converges. [31]

6. From [FM]

(a) Prove that if $\{x_n\}$ converges to L, then $\{|x_n|\}$ converges to $|L|$.

(b) Is the converse always true ? (is it always the case that if $\{|x_n|\}$ converges, then $\{x_n\}$ converges ?)

7. A polynomial with rational coefficients $p(x)$ is given by

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots a_2 x^2 + a_1 x + a_0, \quad a_0, a_1, a_2, \ldots, a_n \in \mathbb{Q}$$

and a rational function $R(x)$ is a fraction of two polynomials with rational coefficients, i.e. $R(x) = \frac{p(x)}{q(x)}$.

(a) Using induction on the degree of the polynomial and the algebraic limit laws proven in this section, prove why if $x_n \to x$, then $p(x_n) \to p(x)$ for any polynomial $p(x)$ with rational coefficients.

(b) What requirements would need to be placed on the sequence $\{x_n\}$ to guarantee that $R(x_n) \to R(x)$ if $x_n \to x$ and $R(x)$ is some rational function.

8. Provide a direct proof of iii) in Theorem 14 using the definition of convergence (an ϵ-N argument) without using i) or ii) from the algebraic limit rules.

2.3 Subsequences

Lecture 6 - 10/9/24

Definition 22. *A function $f : \mathbb{N} \to \mathbb{N}$ is called* **increasing** *if it respects the order structure on the naturals, i.e. for $n_1 \leq n_2$ we have $f(n_1) \leq f(n_2)$. In particular, by induction, it is true that $f(n) \geq n$ for any increasing function f.*

Given a sequence $\{x_k\}$ and an increasing function f, then $f(k) = n_k$ with $n_k \geq k$ and we call the sequence $\{x_{n_k}\}$ a **subsequence** *of $\{x_k\}$. Put another way, for every infinite subset A of $\mathbb{N}$ there is a subsequence of $\{x_k\}$ in which the index of each term in the subsequence n_k is precisely the elements of A.*

Example 7. *Consider the following*

a). Let us look at the following sequence $\{x_n\} = \{(-1)^n\}$. The collection of terms written in order

$$x_2 = 1, x_4 = 1, x_6 = 1, x_8 = 1, \ \ldots$$

[31] This problem is a short introduction to the notion of Cesáro summation or means

is a subsequence of the sequence. We can label it $\{x_{2n}\}$*. The way we are 'choosing' terms from the original sequence to form this subsequence can be described by the inceasing function* $f(n) = 2n$ *or the the infinite subset of the naturals* $A \subseteq \mathbb{N}$*,* $A = \{2, 4, 6, 8, \dots\}$*, i.e. we are only taking the even terms in the sequence.*

b). Let us look at the following sequence $\{x_n\} = \{(-1)^n\}$*. The collection of terms written in order*

$$x_1 = -1, x_3 = -1, x_5 = -1, x_7 = -1, \ \dots$$

is a subsequence of the sequence. We can label it $\{x_{2n-1}\}$*. The way we are 'choosing' terms from the original sequence to form this subsequence can be described by the inceasing function* $f(n) = 2n - 1$ *or the the infinite subset of the naturals* $A \subseteq \mathbb{N}$*,* $A = \{1, 3, 5, 7, \dots\}$*, i.e. we are only taking the odd terms in the sequence.*

c). Let us look at the sequence given by $\{x_n\} = \{n + 2\}$ *whose terms are written as*

$$x_1 = 3, x_2 = 4, x_3 = 5, x_4 = 6, \ \dots$$

If we select the following terms from the sequence

$$x_2 = 4, x_5 = 7, x_8 = 10, x_{11} = 13, \ \dots$$

this is a subsequence of the sequence. We can label it $\{x_{3n-1}\}$*. The way we are 'choosing' the terms from the original sequence can be described by the increasing function* $f(n) = 3n - 1$ *or the infinite subset of the naturals* A *given by* $A = \{2, 5, 8, 11, 14, \dots\}$*, i.e. we are only taking the terms from the sequence that have a remainder of 2 when divided by 3.*

d). Given an arbitrary sequence $\{x_n\}$*, if we write out terms of the sequence like*

$$x_2, x_1, x_5, x_3, x_{17}, x_{11}, \ \dots$$

then this is NOT a subsequence of $\{x_n\}$ *as the terms listed above are not in an increasing order.*

The following listing of terms

$$x_2, x_3, x_5, x_{117}, x_{2059}, x_{11579}, \ \dots$$

is a subseqeunce of the original sequence (assuming the increasing nature of the terms continues). It does not matter that there is not a discernible pattern in the 'choice' of terms from the sequence that is making the subsequence, all that matters is that the index of the terms is increasing from term to term.

Theorem 15. *A sequence* $\{a_n\}$ *converges to a value* L *if and only if every subsequence converges to* L*.*

Proof. $\Rightarrow$ Assume that $\{a_n\}$ converges to L. Let $\{a_{n_k}\}$ be some subsequence $\{a_n\}$, and let ϵ be greater than 0.

For this choice of ϵ, there exists an $N \in \mathbb{N}$ such that

$$|a_n - L| < \epsilon, \quad \forall n > N$$

Well, $\{a_{n_k}\}$ is a subsequence, i.e.

$$n_1 < n_2 < n_3 < \cdots$$

is a strictly increasing collection of natural numbers, thus $n_{k_1} > N$ for some k_1 (and most certainly $n_N > N$), but then we have

$$|a_{n_k} - L| < \epsilon, \quad \forall k > k_1.$$

as a_{n_k} is a term in the original sequence with $n_k > N$. As this can be done for any $\epsilon > 0$ we have that $\{a_{n_k}\}$ converges to L.

$\Leftarrow$ Assume every subsequence of $\{a_n\}$ converges to L.

This one is simple, as

$$1 < 2 < 3 < 4 \cdots$$

is an increasing collection of natural numbers, any sequence is a subsequence of itself (just like any set is a subset of itself). Thus, as any subsequence of $\{a_n\}$ converges to L, and $\{a_n\}$ is a subsequence of itself, we have that $\{a_n\}$ converges to L. □

The value of this theorem is that it can be very useful in showing that a sequence diverges. In particular, if a sequence $\{x_n\}$ contains two subsequences that converge to two distinct values, then it must be that the original sequence diverges according to this theorem. Thus we can finally give proof of the following.

Example 8. *The sequence $\{x_n\} = \{(-1)^n\}$ diverges.*

The even subsequence $\{x_{2n}\} = \{1\}$ is the constant sequence made of ones and thus $x_{2n} \to 1$, and similarly the odd subsequence $\{x_{2n-1}\} = \{-1\}$ is the constant sequence made of negative ones and thus $x_{2n-1} \to -1$. Thus as two subsequences of $\{x_n\}$ converge to distinct values, the prior theorem means that $\{x_n\}$ is divergent.

Example 9. *Consider the sequence $\{x_n\} = \left\{\cos\left(\frac{n\pi}{2}\right) + \frac{(-1)^n}{n}\right\}$. We will look at the following subsequences.*

Odd terms - We will look at the subsequence formed by odd indices, i.e.

$$x_1, x_3, x_5, x_7, \ldots.$$

The indices of this subsequence are given by the increasing function $f(n) = 2n - 1$, so we will have

$$\{x_{2n-1}\} = \left\{\cos\left(\frac{(2n-1)\pi}{2}\right) + \frac{(-1)^{2n-1}}{2n-1}\right\} = \left\{0 - \frac{1}{2n-1}\right\} = \left\{\frac{-1}{2n-1}\right\}$$

from work we did in prior sections we can see that this subsequence converges to 0, i.e. $\{x_{2n-1}\} \to 0$.

Multiples of 4 - We will look at the subsequence whose indices are multiple of 4,

$$x_4, x_8, x_{12}, x_{16}, \ldots$$

The indices of this subsequence are given by the increasing function $f(n) = 4n$, so we will have

$$\{x_{4n}\} = \left\{\cos\left(\frac{4n\pi}{2}\right) + \frac{(-1)^{4n}}{4n}\right\} = \left\{\cos(2\pi n) + \frac{1}{4n}\right\} = \left\{1 + \frac{1}{4n}\right\}$$

and we can see that this subsequence converges to 1, i.e. $\{x_{4n}\} \to 1$.

Multiples of 4 plus 2 - We will look at the subsequence whose indices are multiples of 4 plus 2,

$$x_2, x_6, x_{10}, x_{14}, \ldots$$

The indices of this subsequence are given by the increasing function $f(n) = 4n-2$, so we will have

$$\{x_{4n-2}\} = \left\{\cos\left(\frac{(4n-2)\pi}{2}\right) + \frac{(-1)^{4n-2}}{4n-2}\right\} = \left\{\cos((2n-1)\pi) + \frac{1}{4n-2}\right\} = \left\{-1 + \frac{1}{4n-2}\right\}$$

and we see that this subsequence converges to -1, i.e. $\{x_{4n-2}\} \to -1$.

As there are two subsequences of $\{x_n\}$ that converge to two different values, the prior theorem gives that the sequence $\{x_n\}$ is divergent.

To close out this section we present two last theorems about sequences and subsequences. These theorems phrase the nature of sequence convergence and subsequence convergence in different ways that will hopefully build your intuition. They will not be on the midterm.

Theorem 16. *A sequence $\{x_n\} \to L$ if and only if for every $\epsilon > 0$ all but a finite number of terms from $\{x_n\}$ are contained within the interval $(L-\epsilon, L+\epsilon)$.*

Proof. $\implies$ Assume that $\{x_n\} \to L$. Thus for any $\epsilon > 0$, we know that there exists $N \in \mathbb{N}$ such that for all $n > N$, $|x_n - L| < \epsilon$. This is equivalent to saying that for $n > N$ we have

$$\begin{aligned} |x_n - L| &< \epsilon \\ -\epsilon < x_n - L &< \epsilon \\ L - \epsilon < x_n &< L + \epsilon \\ x_n \in &(L-\epsilon, L+\epsilon) \end{aligned}$$

Thus, we are guaranteed that $x_n \in (L-\epsilon, L+\epsilon)$ for $n > N$. Thus, the only possible values from the sequence $\{x_n\}$ that are not in $(L-\epsilon, L+\epsilon)$ are

$$\{x_1, x_2, x_3, \ldots, x_{N-1}\}$$

which is a finite number. Thus, this direction is proven.

$\impliedby$ Now, assume that for every $\epsilon > 0$ all but a finite number of terms from $\{x_n\}$ are contained within the interval $(L-\epsilon, L+\epsilon)$. We do not know the exact indices of the sequence terms that are not in $(L-\epsilon, L+\epsilon)$, so let us call them

$$x_{j_1}, x_{j_2}, \ldots, x_{j_M}$$

i.e. M points from the sequence are not in $(L-\epsilon, L+\epsilon)$, but we don't know what the indices are (it could be the first M terms, or the first 3 and then $M-3$ random ones with the highest index being the age of the universe in seconds). But the point is we can now define

$$N = \max\{j_1, j_2, \ldots, j_M\}$$

and this is possible as $\{j_1, j_2, \ldots, j_M\}$ is finite. Now, by definition of N as x_{j_M} is the highest indexed term from the sequence not in $(L-\epsilon, L+\epsilon)$, so for all $n > N$ we must have that $x_n \in (L-\epsilon, L+\epsilon)$. And this then implies that

$$\begin{aligned} x_n \in &(L-\epsilon, L+\epsilon) \\ L - \epsilon < x_n &< L + \epsilon \\ \epsilon < x_n - L &< \epsilon \\ |x_n - L| &< \epsilon \end{aligned}$$

And this can be done for any $\epsilon > 0$, so that we have $\{x_n\} \to L$. $\square$

Definition 23. *For a sequence $\{a_n\}$, a value L is called a* **subsequential limit** *if there exists a subsequence $\{a_{n_k}\}$ of the sequence that converges to L.*

Theorem 17. *The number L is a subsequential limit point of the sequence $\{a_n\}$ if and only if for every $\epsilon > 0$, the interval $(L-\epsilon, L+\epsilon)$ contains infinitely many points of $\{a_n\}$.*

Proof. $\Rightarrow$ Assume that L is a subsequential limit of $\{a_n\}$. Thus by definition, there exists a subsequence $\{a_{n_k}\}$ of $\{a_n\}$ that converges to L. Thus, for any $\epsilon > 0$, we have the existence of some $K \in \mathbb{N}$ such that

$$|a_{n_k} - L| < \epsilon, \quad \forall k > K.$$

But now remember, via the definition of absolute value this means

$$\begin{aligned} |a_{n_k} - L| &< \epsilon, \quad \forall k > K \\ -\epsilon < a_{n_k} - L &< \epsilon, \quad \forall k > K \\ L - \epsilon < a_{n_k} &< L + \epsilon, \quad \forall k > K \\ a_{n_k} &\in (L-\epsilon, L+\epsilon), \quad \forall k > K \end{aligned}$$

This shows that the sequence $\{a_n\}$ contains an infinite number of points within $(L-\epsilon, L+\epsilon)$, and this can be done for any $\epsilon > 0$.

$\Leftarrow$ Assume that for every $\epsilon > 0$ the interval $(L-\epsilon, L+\epsilon)$ contains an infinite number of terms of $\{a_n\}$.

Thus, take $\epsilon = 1$, then we know the set

$$\{a_n \mid a_n \in (L-1, L+1)\}$$

is not empty (and infinite actually). Pick the element of $\{a_n\}$ from this set with the smallest index and call it n_1, i.e. a_{n_1} is the first occurring term of $\{a_n\}$ within $(L-1, L+1)$.

Now, take $\epsilon = \frac{1}{2}$, then we have that the set

$$\{a_n \mid a_n \in (L - \frac{1}{2}, L + \frac{1}{2})\}$$

is nonempty (and infinite) by assumption. Pick the element of $\{a_n\}$ from this set with smallest index that is larger than n_1 and call it n_2, i.e. a_{n_2} is the first occurring term of $\{a_n\}$ after possibly a_{n_1} within $(L - \frac{1}{2}, L + \frac{1}{2})$.

We can now proceed generally. Assume $a_{n_1}, a_{n_2}, \ldots, a_{n_k}$ have all been defined with the property that a_{n_j} is the first term of $\{a_n\}$ after possibly $a_{n_1}, a_{n_2}, \ldots, a_{n_{j-1}}$ within $(L - \frac{1}{j}, L + \frac{1}{j})$. In other words, for every $k \in \mathbb{N}$, we have found a method to define a term a_{n_k} of sequence $\{a_n\}$ with the property that $a_{n_k} \in (L - \frac{1}{k}, L + \frac{1}{k})$.

Thus we have created a subsequence $\{a_{n_m}\}$ of $\{a_n\}$, and by definition this subsequence has the property that

$$a_{n_l} \in (L - \frac{1}{k}, L + \frac{1}{k}) \quad \forall l > k$$

Then for an arbitrary $\epsilon > 0$, there exists some $N \in \mathbb{N}$ with $\frac{1}{N} < \epsilon$ by the Archimedean principle. But then we have

$$a_{n_k} \in (L - \frac{1}{N}, L + \frac{1}{N}) \subseteq (L-\epsilon, L+\epsilon), \quad \forall k > N.$$

But as we saw earlier, this is algebraically equivalent to

$$|a_{n_k} - L| < \epsilon, \quad \forall k > N$$

As this can be done for any $\epsilon > 0$, we have that the subsequence $\{a_{n_k}\}$ converges to L, i.e. L is a subsequential limit. □

Recall that a *tail* of the natural numbers is a set of the form

$$\{k \in \mathbb{N} \mid k \geq N\}$$

for some fixed N. Note that a tail is infinite collection of natural numbers with a finite complement (the numbers not in the tail).

Not all infinite subsets of $\mathbb{N}$ are tails, for example, the even and odd numbers are infinite subsets and neither of them are tails. The indices of our three subsequences in the prior example are all infinite subsets of the natural numbers that are not tails. But the notion of a tail can make it a little clearer how these two theorems explain convergence versus subsequential convergence.

Most mathematicians make a distinction between 'eventually' and 'frequently' when discussing sequential limits and subsequential limits. If a sequence $\{a_n\}$ converges to a value L, then for any ϵ an infinite tail of indices of the sequence will *eventually* be in an error window $(L - \epsilon, L + \epsilon)$ of L, i.e. after some point (term N) every remaining term ($n > N$) will be within ϵ of L.

On the other hand, terms of a sequence $\{a_n\}$ are *frequently* near subsequential limits, i.e. any error window $(L - \epsilon, L + \epsilon)$ of a subsequential limit L will contain an infinite number of points of the sequence, but it will not necessarily be a tail.

Exercises for section 2.3:

1. From [FM] Consider the sequence $u_n = (-1)^n$. Write out the first 5 terms of the subsequence $\{u_{3k+1}\}_{k\geq 1}$.
2. For the sequence $\{x_n\} = \left\{\sin\left(\frac{n\pi}{4}\right) + \frac{1}{n}\right\}$, please write out in simplified form, the following subsequences
   (a) $\{x_{4n}\}$.
   (b) $\{x_{4n+1}\}$.
   (c) $\{x_{4n+2}\}$.
   (d) $\{x_{4n+3}\}$.

   Which of these subsequences converge? Which diverge?
3. (a) Suppose a sequence u_n is such that $\lim_{k\to\infty} u_{2k} = L \in \mathbb{Q}$ and $\lim_{k\to\infty} u_{2k+1} = L$. Prove that $\lim_{n\to\infty} u_n = L$.
   (b) Prove or disprove. If a sequence is such that $\lim_{k\to\infty} u_{3k} = L \in \mathbb{Q}$ and $\lim_{k\to\infty} u_{3k+1} = L$, does this imply that $\lim_{n\to\infty} u_n = L$?

4. Given a sequence $\{x_n\}$, say you have M distinct subsequences of $\{x_n\}$ in that none of the terms in the subsequences overlap. For $M = 3$ this could look like

$$\begin{aligned} &\text{Subsequence 1} \quad x_3, x_6, x_9, \dots \\ &\text{Subsequence 2} \quad x_1, x_4, x_7, \dots \\ &\text{Subsequence 3} \quad x_2, x_5, x_8, \dots \end{aligned}$$

Call A_k the indices of subsequence k, i.e. in terms of the above example $A_1 = \{3, 6, 9, \dots\}$, $A_2 = \{1, 4, 7, \dots\}$ $A_3 = \{2, 5, 8, \dots\}$. If every one of these distinct M subsequences converge to the same value L, what condition is required of

$$\bigcup_{k=1}^{M} A_k$$

to guarantee the original sequence $\{x_n\}$ converges to L.

2.4 Cauchy Sequences

Lecture 7 - 10/11/24

Let us start this section with an example

Example 10. *Consider the sequence $\{a_n\}$ given recursively by*

$$a_1 = 1, \qquad a_{n+1} = \frac{1}{2}\left(a_n + \frac{2}{a_n}\right)$$

Let us first show that $a_n \geq 1$ for all $n \in \mathbb{N}$. This is clearly true for $n = 1$ as $a_1 = 1 \geq 1$. Now assume that $a_n \geq 1$ and let us prove that $a_{n+1} \geq 1$. The following computation

$$\begin{aligned} a_{n+1} - 1 &= \frac{1}{2}\left(a_n + \frac{2}{a_n}\right) - 1 = \frac{1}{2}\left(a_n - 2 + \frac{2}{a_n}\right) = \frac{1}{2a_n}\left(a_n^2 - 2a_n + 1 + 1\right) \\ &= \frac{1}{2a_n}\left((a_n - 1)^2 + 1\right) \geq 0 \end{aligned}$$

as a_n is positive. Thus this shows that $a_{n+1} \geq 1$. And so by induction, we have that $a_m \geq 1$ for all $m \in \mathbb{N}$.

Let us now look at the expression $a_{n+1}^2 - 2$.

$$\begin{aligned} a_{n+1}^2 - 2 &= \left[\frac{1}{2}\left(a_n + \frac{2}{a_n}\right)\right]^2 - 2 = \frac{1}{4}\left(a_n^2 + 4 + \frac{4}{a_n^2}\right) - 2 \\ &= \frac{1}{4}\left(a_n^2 - 4 + \frac{4}{a_n^2}\right) = \frac{1}{4a_n^2}(a_n^4 - 4a_n^2 + 4) = \frac{1}{4a_n^2}(a_n^2 - 2)^2 \end{aligned}$$

And this shows that $a_{n+1}^2 - 2$ is non-negative, thus it equals its own absolute value. As $a_n \geq 1$ for all n we have that

$$|a_{n+1}^2 - 2| \leq \frac{1}{4}(a_n^2 - 2)^2$$

As $a_1 = 1$, this implies that $|a_2^2 - 2| \leq \frac{1}{4}$, and then from the relation above again we have

$$|a_3^2 - 2| \leq \frac{1}{4}(a_2^2 - 2)^2 \leq \frac{1}{4} \cdot \frac{1}{4^2} = \frac{1}{4^3}$$

and continuing in this manner one can see

$$|a_n^2 - 2| \leq \frac{1}{4^{2^{n-1}-1}}$$

and as $2^{n-1} - 1$ is an increasing function, we have $2^{n-1} - 1 \geq n$, thus

$$|a_n^2 - 2| \leq \frac{1}{4^{2^{n-1}-1}} \leq \frac{1}{4^n}$$

As $0 \leq \frac{1}{4} \leq 1$, we know that $\lim_{n\to\infty} \frac{1}{4^n} = 0$, thus the squeeze theorem implies that $\lim_{n\to\infty} |a_n^2 - 2| = 0$. Put another way this shows that $a_n^2 \to 2$. Thus $\{a_n\}$ is a sequence of rational numbers with $a_n^2 \to 2$.

In fact, going further, as $a_n \geq 1$, we have $a_n + \sqrt{2} \geq 2$ for all $n \in \mathbb{N}$, and the following

$$|a_n - \sqrt{2}| = \left|\frac{a_n^2 - 2}{a_n + \sqrt{2}}\right| \leq \frac{1}{2}|a_n^2 - 2|$$

paired with the squeeze theorem gives that a_n converges to $\sqrt{2}$. And that statement will be true in $\mathbb{R}$ as $\sqrt{2}$ exists there, but as we have previously mentioned $\sqrt{2} \notin \mathbb{Q}$, thus $\{a_n\}$ does not converge in $\mathbb{Q}$. And while we will construct the real numbers shortly, and this will remedy the situation with sequences like $\{a_n\}$, there is another issue that naturally arises from the definition of convergence and that is one of practicality.

The definition of convergence of a sequence $\{a_n\}$ depends upon the limit L that the sequence tends to, i.e. $\lim_{n\to\infty} a_n = L$. Our definition of convergence is stated as: For all $\epsilon \in \mathbb{Q}_+$, there exists an $N \in \mathbb{N}$ such that for all $n > N$ we have

$$|a_n - L| < \epsilon.$$

and it implicitly depends on knowledge of what L is. This is a definition we can employ when already know where a sequence is tending or if we have performed some analysis or tests and have a strong candidate in mind for what L should be. But how do we tell if a sequence converges in the first place without knowing what it converges to? This will be explained at the end of the section, but let us start with a definition that is similar to convergence for a sequence but slightly different.

Definition 24. *A sequence $\{x_n\}$ is called* **Cauchy** *if for every $\epsilon \in \mathbb{Q}_+$ there exists an $N \in \mathbb{N}$ such that for all $p, q > N$ (sometimes written $p > q > N$,*

$$|x_p - x_q| < \epsilon$$

or in terms of quantifiers,

$$\forall \epsilon \in \mathbb{Q}_+, \exists N \in \mathbb{N}, \forall p, q > N, |x_p - x_q| < \epsilon$$

Note the definition is similar to convergence but makes no mention of where a sequence is converging (if it does at all[32]). It merely states that a sequence will infinitely cluster in smaller and smaller regions.

[32] In $\mathbb{R}$, it will and this will be mentioned later

Example 11. *i) Let us look at the sequence* $\{x_n\} = \{(-1)^n\}$.

For this sequence we have the following

$$|x_p - x_q| = \begin{cases} 0 & \text{if } p, q \text{ are both odd or both even} \\ 2 & \text{if } p \text{ is odd, } q \text{ is even or vice-versa} \end{cases}$$

And because of this, for any $N \in \mathbb{N}$*, it will be impossible to guarantee* $|x_p - x_q| < \epsilon$ *for all* $p, q > N$ *unless* $\epsilon \geq 2$*. Because of this, we can see that* $\{x_n\}$ *is not a Cauchy sequence.*

ii). Now let us look at the sequence $\{y_n\} = \{\frac{1}{n}\}$*. For* $p > q$ *we will have*

$$|y_p - y_q| = y_q - y_p = \frac{1}{q} - \frac{1}{p} < \frac{1}{q}$$

So if $\epsilon \in \mathbb{Q}_+$ *is some fixed quantity, by the Archimedean property there is a* $N \in \mathbb{N}$ *such that* $\frac{1}{N} < \epsilon$*. Thus for all* $p > q > N$ *we have that*

$$|y_p - y_q| = y_q - y_p = \frac{1}{q} - \frac{1}{p} < \frac{1}{q} < \frac{1}{N} < \epsilon$$

And as this can be done for any ϵ*, we have that* $\{y_n\}$ *is a Cauchy sequence.*

What we can surmise from our examples above is that boundedness of a sequence is not strong enough to guarantee a sequence is Cauchy, but it seems like convergence of a sequence is strong enough to guarantee a sequence is Cauchy. What follows are two theorems, the latter will confirm the second thought, and the first will show that Cauchy sequences are bounded.[33]

Theorem 18. *Cauchy Sequences are bounded*

Proof. Let $\{x_n\}$ be a Cauchy sequence. By definition, if we let $\epsilon \in \mathbb{Q}_+$ be a fixed quantity (like 1), then for this ϵ there is a $N \in \mathbb{N}$ such that for all $p, q > N$ we have

$$|x_p - x_q| < \epsilon$$

So let us fix $q = N + 1$, the above implies that

$$x_{N+1} - \epsilon < x_p < x_{N+1} + \epsilon$$

for all $p > N$, and so $|x_p| < \max(x_{N+1} + \epsilon, \epsilon - x_{N+1})$ for all $p > N$. This gives us a bound on an infinite tail $\{x_n\}_{n=N+1}^{\infty}$ of our sequence, thus we can define

$$M = \max(|x_1|, |x_2|, \ldots, |x_N|, x_{N+1} + \epsilon, \epsilon - x_{N+1})$$

and this will have the property that

$$|x_n| \leq M, \quad \forall n \in \mathbb{N}$$

and thus $\{x_n\}$ is bounded. □

[33] similar to convergent sequences

Theorem 19. *Convergent sequences are Cauchy*

Proof. Assume that $\{x_n\}$ is a converget sequence, i.e. there exists a $L \in \mathbb{Q}$ such that $\lim_{n\to\infty} x_n = L$. Let $\epsilon \in \mathbb{Q}_+$ be a fixed quantity momentarily. The for this ϵ there exists $N \in \mathbb{N}$ such that for all $n > N$,

$$|x_n - L| < \frac{\epsilon}{2}$$

Then for $n, m > N$ we will have

$$\begin{aligned} |x_m - x_n| &= |x_m - L + L - x_n| \\ &\leq |x_m - L| + |x_n - L| && \text{Triangle Inequality} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon && \text{as } n, m > N \end{aligned}$$

And there was nothing special about ϵ other than it being a positive rational number, thus this result holds for any positive rational ϵ. Thus $\{x_n\}$ is a Cauchy sequence. □

Returning to our sequence $\{a_n\}$ from earlier given by

$$a_1 = 1, \qquad a_{n+1} = \frac{1}{2}\left(a_n + \frac{2}{a_n}\right)$$

we saw that $\{a_n^2\} \to 2$, and as this sequence is convergent, we have that $\{a_n^2\}$ is a Cauchy sequence. And similar to a computation from earlier, we have

$$|a_p - a_q| = \left|\frac{a_p^2 - a_q^2}{a_p + a_q}\right| \leq \frac{1}{2}|a_p^2 - a_q^2|$$

and as the right hand side can be made arbitrarily small (less than ϵ) for $p.q > N$ as $\{a_n^2\}$ is Cauchy, we see that the same is true of $\{a_n\}$, thus $\{a_n\}$ is a Cauchy sequence. Of particular to note is that

$$\{a_n\} \text{ is a sequence of rational numbers that is Cauchy but not convergent.}$$

And this is actually another way to phrase the non-completeness of $\mathbb{Q}$. For the real numbers, the theorem above will be one half of a result and that is the Cauchy completeness property.

Cauchy Completeness Property : A set X is called **Cauchy complete** if every Cauchy sequence in X is convergent in X.

So at the moment we have two notions of completeness: Cauchy completeness and the Least Upper Bound property. We will see that these are equivalent notions[34], but it will turn out that many results we will later prove in this course about the real numbers and operations on the real numbers will be equivalent to completeness.[35] [JP]

With all of this said, hopefully we can see the value of Cauchy sequences now. They will play an existential role for us in later sections when paired with the completeness of the reals. In $\mathbb{R}$, as a sequence is Cauchy precisely if it convergent and vice-versa, we can check if a sequence converges to a limit without needing any prior knowledge of what the limit is first. This may seem like a

[34] in sets that have the Archimedean property

[35] i.e. instead of showing $\mathbb{R}$ has the LUB property or Cauchy completeness and deducing these later results, we could start with these later results and deduce the LUP or Cauchy completeness property

silly thing, and the process of checking if a sequence is Cauchy may not bring us any closer or give any more intuition to what the limit of a sequence would be, but it is always best to make sure a destination in mind exists before attempting to reach it.

As one last thing, there is a fairly nice result involving subsequences of a Cauchy sequence. The proof will be left as an exercise, and this theorem will also be used in conjunction with another to show the completeness of $\mathbb{R}$ in different way.

Theorem 20. *If $\{x_n\}$ is a Cauchy sequence that contains a convergent subsequence, i.e. $\{x_{n_k}\} \to L$ for some subsequence $\{x_{n_k}\}$ of $\{x_n\}$, then $\{x_n\}$ is convergent and $\{x_n\} \to L$.*

Proof. Left as an exercise. □

As we saw in the prior section, convergent sequences force all of their subsequences to converge to the same place, and in turn the subsequences of a given sequence can only guarantee the entire sequence converges if every single subsequence converges to the same value. In the latter case, this can be a very large task to prove. The theorem above gives us a much simpler alternative. If you know the sequence 'clusters' (is Cauchy), and that a piece of it converges (the convergent subsequence), then the whole sequence converges. In particular, as the sequnce converges to the subsequential limit in this case, this is often much easier to find as subsequences can ignore initial osciallation in a sequences terms or outlier terms.

Exercises for section 2.4:

1. From [FM]
   (a) Let r be a rational satisfying $0 < r < 1$. Show that for any $p > q$,
   $$\sum_{k=q+1}^{p} r^k \leq \frac{r^{q+1}}{1-r}.$$
   (b) Deduce from (a) that the sequence $v_n := 1 + r + r^2 + \cdots + r^n$ is Cauchy.
   (c) Now let $r = 1/2$ and consider the sequence $\{v_n\}$ defined as in (b). Calling $\{u_n\}$ the sequence defined by
   $$u_n = \sum_{k=0}^{n} \frac{1}{k!} = 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{n!},$$
   show that for all naturals $p > q$,
   $$|u_p - u_q| \leq 2|v_p - v_q|.$$
   (d) Deduce from (c) that the sequence $\{u_n\}$ is Cauchy.
2. [FM] Let a_n a sequence of positive rationals, such that $a_n \geq 1$ for every n.
   (a) Using the identity $a^3 - b^3 = (a-b)(a^2 + ab + b^2)$, true for every rationals a, b,
   $$|a_p - a_q| \leq \frac{1}{3}|a_p^3 - a_q^3|, \qquad \forall p, q \in \mathbb{N}.$$
   (b) Show that if a_n^3 is Cauchy, then the sequence $\{a_n\}$ is Cauchy.
3. Prove theorem 20.

3 The construction of $\mathbb{R}$

Lecture 8 - 10/14/24

So far we have seen that the rational number system $\mathbb{Q}$ has nearly everything we would want in a number system for the study of algebra and calculus, but there are unavoidable 'gaps' in the rationals. We have phrased these gaps in two ways:

- the rationals do not have the least upper bound property, i.e. there are sets in the rationals that are bounded above but do not have a least upper bound.
- the rationals are not Cauchy complete, i.e. there are Cauchy sequences formed of rational numbers that do not converge to rational numbers.

And so now, we will begin our journey towards the construction of the real numbers. Along the way we will need to define real numbers, and show that we have not lost any of the structures we like from $\mathbb{Q}$: field structure, totally ordered, Archimedean Property, etc. But most importantly, we will need to show that $\mathbb{R}$ does not have these same 'gaps' that $\mathbb{Q}$ has. We will do so by showing the reals satisfy the Least Upper Bound and Cauchy completeness properties.[36]

The two typical methods of construction of $\mathbb{R}$ are the method of Dedekind cuts and the method of equivalence classes of Cauchy sequences. They are each individually suited to 'filling the gaps' in $\mathbb{Q}$ by the how the methods themselves seek to 'complete' $\mathbb{Q}$ and turn it into $\mathbb{R}$. To be more clear, the method of Dedekind cuts to build $\mathbb{R}$ has the advantage that proving the Least Upper Bound property of $\mathbb{R}$ becomes almost trivial as it is built into the notion of a 'cut'. Similarly, the method of equivalence classes of Cauchy sequences to build $\mathbb{R}$ has the advantage that Cauchy completeness is easier to prove by nature this construction involving Cauchy sequences to begin with.

In this section I will provide the construction of $\mathbb{R}$ by making use of equivalence classes of Cauchy sequences as this is typically simpler for many students to understand on a first pass. I will try to provide a construction via Dedekind cuts in the appendix for those that are interested.

Lastly, let me explain the general idea behind the construction itself. At the end of the section on Cauchy sequences, we returned to the sequence $\{a_n\}$ given by

$$a_1 = 1, \qquad a_{n+1} = \frac{1}{2}\left(a_n + \frac{2}{a_n}\right)$$

and in that section we saw that this sequence is Cauchy and that it seems to converge to $\sqrt{2}$. So what kept this sequence from converging in $\mathbb{Q}$ was that it's destination, $\sqrt{2}$, does not exist within the rational numbers. Now, this example was chosen as we all have some familiarity with $\sqrt{2}$, even though to us at this moment from a formal standpoint, $\sqrt{2}$ does not exist yet as a number. So, me relying on our prior knowledge of $\sqrt{2}$ to describe the gap in $\mathbb{Q}$ at $\sqrt{2}$ is already fairly circular in that I should not be referencing a real number while defining a real number.

So, how do we fill in these gaps (like $\sqrt{2}$) that the rationals have without giving name to the numbers that make up the gaps? Well, we change our frame of mind. Cauchy sequences have it in their nature to cluster closer and closer together, i.e. they do not diverge by bouncing around or osciallating or diverging off to $\pm\infty$ as we have seen with previous examples. Basically, Cauchy sequences are always converging to something intuitively speaking, and the only thing that will

[36] These will be equivalent properties due to the Archimedean Property

formally keep them from being a convergent sequence is that the value they converge to does not exist in the set in question. The idea for how to fill in these gaps without describing these gaps is: instead of thinking a Cauchy sequence $\{a_n\}$ fails to converge because it's limit L may not be rational, let us describe these values L missing from $\mathbb{Q}$ by describing this number L by the collection of all sequences tending to it. This would like if you had a map with a hole in it but you see multiple roads passing through the hole in your map. You posit or know with a 'gut feeling' that there is a city where the hole in your map is because you see so many paths to it even though you can not yet say what the city is. So instead of us thinking all paths must have a destination, we will define our missing destinations by looking at the paths through these destination gaps.

3.1 Definition of Real Numbers

To begin this section, we need to recall the notion of an equivalence relation and an equivalence class. There will be more information on this topic in an appendix, but quickly, an equivalence relation on a set S is a way of creating a new notion of 'equality' on a set S that is less restrictive that $=$ and will let us see general elements $x, y \in S$ as equivalent due to a property they share even though generally $x \neq y$.

Definition 25. *A relation $\sim$ on a set S is called an* **equivalence relation** *if it is reflexive, symmetric, and transitive.*

- **reflexive** - *A relation $\sim$ is reflexive on S if $x \sim x$ for all $x \in S$.*
- **symmetric** - *A relation $\sim$ is symmetric on S if $x \sim y$ implies $y \sim x$ for all $x, y \in S$.*
- **transitive** - *A relation $\sim$ is transitive on S if $x \sim y$ and $y \sim z$ implies that $x \sim z$ for all $x, y, z \in S$.*

For $\sim$ an equivalence relation on S, the **equivalence class** *of an element x, written $[x]$ is given by*

$$[x] = \{y \in S \mid x \sim y\}$$

which is the collection of all elements equal to x under $\sim$.

We now begin our process of constructing the real numbers. We begin with the following definition. This is us setting the stage with the collection of all sequences of rational numbers that 'cluster' somewhere.

Definition 26. *Let $C_{\mathbb{Q}}$ denote the set of all Cauchy sequences formed by rational numbers.*

The set $C_{\mathbb{Q}}$ acts like an infinite dimensional vector space or really an infinite dimensional algebra in that you can think of the elements/sequences $\{x_n\} \in C_{\mathbb{Q}}$ as (countably) infinite dimensional vectors. We can define termiwse addition and termwise multiplication of sequences by

$$\{x_n\} + \{y_n\} = \{x_n + y_n\}, \qquad \{x_n\} \cdot \{y_n\} = \{x_n \cdot y_n\}$$

but we do need to check that $C_{\mathbb{Q}}$ is *closed* under these operations, and this is what the following theorem will show.

Theorem 21. *For sequences $\{x_n\}, \{y_n\} \in C_{\mathbb{Q}}$ we have $\{x_n + y_n\}, \{x_n \cdot y_n\} \in C_{\mathbb{Q}}$.*

Proof. Let $\epsilon > 0$, as $\{x_n\}$ is Cauchy, there exists N_1 such that for all $p, q > N_1$ we have

$$|x_p - x_q| < \frac{\epsilon}{2}$$

Similarly as $\{y_n\}$ is Cauchy, there exists N_2 such that for all $p, q > N_2$ we have

$$|y_p - y_q| < \frac{\epsilon}{2}$$

Thus for $n > \max(N_1, N_2)$, we have

$$\begin{aligned} |(x_p + y_p) - (x_q + y_q)| &= |(x_p - x_q) + (y_p - y_q)| \\ &\leq |x_p - x_q| + |y_p - y_q| && \text{Triangle inequality} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon && \text{for } n > \max(N_1, N_2) \end{aligned}$$

as this can be done for any positive choice of ϵ we have that $\{x_n + y_n\}$ is a Cauchy sequence, so $\{x_n + y_n\} \in C_{\mathbb{Q}}$.

Now, for the product part. As $\{x_n\}$ and $\{y_n\}$ are both Cauchy sequence, they are bounded, thus there exists L and P such that

$$|x_n| \leq P, \quad |y_n| \leq L, \quad \forall n \in \mathbb{N}$$

We then take $M = \max(P, L)$. Then as $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences, for $\frac{\epsilon}{2M}$ there exists N_1 and N_2 respectively such that

$$|x_p - x_q| < \frac{\epsilon}{2M}, \text{ for } p, q > N_1, \text{ and } |y_p - y_q| < \frac{\epsilon}{2M}, \text{ for } p, q > N_2$$

Then for all $p, q > \max(N_1, N_2)$ we have

$$\begin{aligned} |x_p y_p - x_q y_q| &= |x_p y_p - x_p y_q + x_p y_q - x_q y_q| \\ &= |x_p(y_p - y_q) + y_q(x_p - x_q)| \\ &\leq |x_p||y_p - y_q| + |y_q||x_p - x_q| && \text{Triangle Inequality} \\ &< M|y_p - y_p| + M|x_p - x_q| && \text{Boundedness} \\ &< M\left(\frac{\epsilon}{2M} + \frac{\epsilon}{2M}\right) = \epsilon && \text{for } n > \max(N_1, N_2) \end{aligned}$$

and thus we see that $\{x_n \cdot y_n\}$ is a Cauchy sequence as this can be done for any $\epsilon > 0$. □

Now that we have shown that the space of Cauchy sequences of rational numbers $C_{\mathbb{Q}}$ is closed under termwise addition and termwise multiplication, we can define a real number. Once again, we will define a number/destination in terms of the paths that tend towards 'something'.[37] The only issue is that the collection of all Cauchy sequences of rational numbers is too big in that there can be a lot of sequences that tend/cluster in the same place. To cut this down and cut down the confusion, we create an equivalence relation on the space of Cauchy sequences of rationals that will see two sequences as the 'same' if they cluster/tend to the same place.

Definition 27. *We define the relation $\sim$ on $C_{\mathbb{Q}}$ by saying $\{x_n\} \sim \{y_n\}$ if $\lim_{n\to\infty} |x_n - y_n| = 0$.*

[37] being Cauchy sequences guarantees the clustering to a something, but we can not say where it is tending yet or this would be a circular definition of real numbers

Theorem 22. *The relation $\sim$ forms an equivalence relation on the space of Cauchy sequences of rational numbers, $C_{\mathbb{Q}}$.*

Proof. Left as an exercise. □

Exercises for section 3.1:

1. Provide a proof of Theorem 22.
2. Prove that if a Cauchy sequence of rationals $\{x_n\}$ is modified by changing a finite number of terms, the result is an equivalent Cauchy sequence.
3. Prove that there is an uncountable number of Cauchy sequences of rational numbers equivalent to any given Cauchy sequence of rational numbers. *Hint*: Cantor Diagonalization Argument

3.2 The field structure of $\mathbb{R}$

We have already seen that termwise addition and multiplication is defined upon the space of Cauchy sequences of rationals, $C_{\mathbb{Q}}$, but as we have now put an equivalence relation on this space we must make sure that termwise addition and termwise multiplication remain *well-defined* on the equivalence classes, i.e. we need to check that

$$[\{x_n\}] + [\{y_n\}] = [\{x_n + y_n\}], \quad \text{and} \quad [\{x_n\}] \cdot [\{y_n\}] = [\{x_n \cdot y_n\}]$$

Putting this another way, we need to check that the result we obtain when we add or multiply two classes of Cauchy sequences is independent of the representatives of the classes chosen.

Example 12. *Let $x = [\{x_n\}] = 1$ and $y = [\{y_n\}] = 2$. Then the following are two different representatives of x*

$$\{x_n\} = \{1\}, \qquad \{x'_n\} = \left\{1 + \frac{1}{n}\right\}$$

as $\lim_{n\to\infty} |1 + \frac{1}{n} - 1| = 0$. Similarly the following are two representatives of y

$$\{y_n\} = \{2\}, \qquad \{y'_n\} = \left\{2 + \frac{(-1)^n}{n^2}\right\}$$

as $\lim_{n\to\infty} |2 - (2 + \frac{(-1)^n}{n^2})| = 0$. We can then see that

$$\{x_n + y_n\} = \{3\}, \qquad \{x'_n + y'_n\} = \left\{3 + \frac{1}{n} + \frac{(-1)^n}{n^2}\right\}$$

are equivalent Cauchy sequences as $\lim_{n\to\infty} |3 + \frac{1}{n} + \frac{(-1)^n}{n^2} - 3| = 0$. This means that

$$[\{x_n + y_n\}] = [\{x'_n + y'_n\}]$$

and so the termwise sum of two classes of Cauchy sequences is independent of the representatives chosen.

Theorem 23. *Let $x = [\{x_n\}]$ and $y = [\{y_n\}]$ be real numbers and let $\{x_n\}$, $\{x'_n\}$ be two representatives of x and similarly let $\{y_n\}$, $\{y'_n\}$ be two representatives of y, then*

$$[\{x_n + y_n\}] = [\{x'_n + y'_n\}], \quad \textit{and} \quad [\{x_n \cdot y_n\}] = [\{x'_n \cdot y'_n\}]$$

Thus termwise addition and termwise multiplication of two classes of Cauchy sequences representing real numbers are independent of representatives chosen for each class. This shows that $+$ and $\cdot$ are well-defined operations on $\mathbb{R}$.

Proof. If $\{x_n\}$, $\{x'_n\}$ are two representatives of x then $\lim_{n\to\infty} |x_n - x'_n| = 0$ by definition. Similarly, for the representatives $\{y_n\}$, $\{y'_n\}$ of y we have $\lim_{n\to\infty} |y_n - y'_n| = 0$. Thus for $\epsilon > 0$, we have the existence of $N_1 \in \mathbb{N}$ such that for all $n > N_1$

$$|x_n - x'_n| < \frac{\epsilon}{2}$$

and similarly for this ϵ, the natural number N_2 exists with the property that for all $n > N_2$ we have

$$|y_n - y'_n| < \frac{\epsilon}{2}$$

Because of this, for all $n > \max(N_1, N_2)$ we have

$$\begin{aligned} |(x_n + y_n) - (x'_n + y'_n)| &= |(x_n - x'_n) + (y_n - y'_n)| \\ &\leq |x_n - x'_n| + |y_n - y'_n| && \text{Triangle Inequality} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon && \text{for } n > \max(N_1, N_2) \end{aligned}$$

and this shows that $\lim_{n\to\infty} |(x_n + y_n) - (x'_n + y'_n)| = 0$, which means that $\{x_n + y_n\}$ and $\{x'_n + y'_n\}$ are equivalent sequences, thus

$$[\{x_n + y_n\}] = [\{x'_n + y'_n\}]$$

For the result about products, we first exploit that all of our representatives are Cauchy sequences and are thus bounded. Hence there exists N, V such that

$$|x_n| \leq N, \quad |y'_n| \leq V, \quad \forall n \in \mathbb{N}$$

We then take $M = \max(N, V)$. For $\epsilon > 0$, as $\lim_{n\to\infty} |x_n - x'_n| = 0$ and $\lim_{n\to\infty} |y_n - y'_n| = 0$, we have the existence of N_1 and N_2 such that

$$|x_n - x'_n| < \frac{\epsilon}{2M} \text{ for } n > N_1 \text{ and } |y_n - y'_n| < \frac{\epsilon}{2M} \text{ for } n > N_2$$

and so for $n > \max(N_1, N_2)$ we have

$$\begin{aligned} |x_n y_n - x'_n y'_n| &= |x_n y_n - x_n y'_n + x_n y'_n - x'_n y'_n| \\ &= |x_n(y_n - y'_n) + y'_n(x_n - x'_n)| \\ &\leq |x_n||y_n - y'_n| + |y'_n||x_n - x'_n| && \text{Triangle Inequality} \\ &\leq M|y_n - y'_n| + M|x_n - x'_n| && \text{From boundedness} \\ &< M\left(\frac{\epsilon}{2M} + \frac{\epsilon}{2M}\right) = \epsilon && \text{for } n > \max(N_1, N_2) \end{aligned}$$

And this shows that $\lim_{n\to\infty} |x_n y_n - x'_n y'_n| = 0$. Thus $\{x_n \cdot y_n\}$ and $\{x'_n \cdot y'_n\}$ are equivalent sequences, and so

$$[\{x_n \cdot y_n\}] = [\{x'_n \cdot y'_n\}]$$

□

Now that we have completed this proof, the operations of $+$ and $\cdot$ are defined on $\mathbb{R}$ by

$$[\{x_n\}] + [\{y_n\}] = [\{x_n + y_n\}], \quad \text{and} \quad [\{x_n\}] \cdot [\{y_n\}] = [\{x_n \cdot y_n\}]$$

As a real number x is defined to be an equivalence class of Cauchy sequences of rational numbers, the way we have defined $+$ and $\cdot$ on $\mathbb{R}$ shows that the real numbers inherit many properties from the rational numbers. For example, if $x = [\{x_n\}]$ and $y = [\{y_n\}]$, then

$$x + y = [\{x_n\}] + [\{y_n\}] = [\{x_n + y_n\}] = [\{y_n + x_n\}] = [\{y_n\}] + [\{x_n\}] = y + x$$

where the third equals sign comes from the commutativity of addition on $\mathbb{Q}$. Thus we see how addition on $\mathbb{R}$ inherits commutativity from the commmutativity of addition on $\mathbb{Q}$. Similarly, $\mathbb{R}$ inherits the associative law for $+$ and $\cdot$, the distributive law, and the commutativity of $\cdot$ from $\mathbb{Q}$ as well.

The additive identity of 0 exists in $\mathbb{R}$ as well as the multiplicative identity 1 as constant sequences

$$0 = [\{0\}] = [\{0, 0, 0, \dots\}]$$
$$1 = [\{1\}] = [\{1, 1, 1, \dots\}]$$

and similarly the additive inverse of a real number $x = [\{x_n\}]$ exists by taking the 'negative' sequences, $-x = [\{-x_n\}]$.

$$x + (-x) = [\{x_n\}] + [\{-x_n\}] = [\{x_n - x_n\}] = [\{0\}] = 0$$

Thus, we see that $\mathbb{R}$ so far has all of the field axioms except for the existence of multiplicative inverses for $x \neq 0$. It will turn out that $\mathbb{R}$ does indeed have this property, but it will take a good amount of work to show.

Theorem 24. *For a real number $x = [\{x_n\}]$ with $x \neq 0$ there exists a sequence $\{a_n\} \in x$ with $a_n \neq 0$ for all $n \in \mathbb{N}$. Furthermore $\{\frac{1}{a_n}\}$ is a Cauchy sequence and $\frac{1}{x} = \left[\left\{\frac{1}{a_n}\right\}\right]$.*

Proof. We have that $x = [\{x_n\}] \neq 0 = [\{0\}]$, thus $\lim_{n\to\infty} |x_n - 0| \neq 0$. Because of this we have the negation of convergence

$$\neg(\forall \epsilon > 0, \exists N \in \mathbb{N}, s.t. \forall n > N, |x_n - 0| < \epsilon)$$
$$\exists \epsilon > 0, \forall N \in \mathbb{N}, s.t. \exists n > N, |x_n| \geq \epsilon$$

So there exists a specific ϵ such that for any $N \in \mathbb{N}$, there is some n larger than N with $|x_n| \geq \epsilon$. So let ρ equal this specific value of ϵ for which this happens.

Now, by definition the sequence $\{x_n\}$ is Cauchy, and because of this, for any choice of $\epsilon > 0$ there exists some $N \in \mathbb{N}$ for the choice of ϵ such that for all $p, q > N$, we have

$$|x_p - x_q| < \epsilon$$

And the point of me re-stating the definition of Cauchy is that because $\{x_n\}$ is Cauchy we are free to choose the value of ϵ as long as it is positive. So, we choose $\epsilon = \frac{\rho}{2}$. For this choice of ϵ there is a $N_1 \in \mathbb{N}$ such that for all $p, q > N_1$

$$|x_p - x_q| < \frac{\rho}{2}$$

From what we saw earlier, there is some $n_1 > N_1$ such that $|x_{n_1}| \geq \rho$, and thus for all $p > N_1$ we have

$$|x_p| = |x_{n_1} - (x_{n_1} - x_p)| \geq |x_{n_1}| - |x_{n_1} - x_p| > \rho - \frac{\rho}{2} = \frac{\rho}{2} > 0$$

Thus we have just shown that $|x_p| > 0$ for all $p > N_1$, i.e. that there is a tail of the sequence $\{x_n\}$ in which all terms of the sequence in the tail are non-zero. And so, the only terms of this sequence that can be zero must be x_m for $1 \leq m \leq N_1$, which is a finite number of terms. Replacing a finite number of terms in a sequence will not change it's convergent or Cauchy nature as these properties depend upon the infinite tails of a sequence. Thus for any term $x_m = 0$ for $1 \leq m \leq N_1$ simply replace it with $x_m = 1$ and name this new sequence $\{a_n\}$, i.e. the terms in this new sequence will be

$$a_n = \begin{cases} x_n & \text{if } x_n \neq 0,\ 1 \leq n \leq N_1 \\ 1 & \text{if } x_n = 0,\ 1 \leq n \leq N_1 \\ x_n & \text{if } n > N_1 \end{cases}$$

Then this sequence will have the property that $\{a_n\} \sim \{x_n\}$, $a_n \neq 0$ for all $n \in \mathbb{N}$ and $|a_n| \geq \frac{\rho}{2}$ for $n > N_1$.

As none of the terms in the sequence $\{a_n\}$ are equal to zero, we can define the sequence $\left\{\frac{1}{a_n}\right\}$. As $|a_n| \geq \frac{\rho}{2}$ for $n > N_1$, we see that for $p, q > N_1$

$$\left|\frac{1}{a_p} - \frac{1}{a_q}\right| = \frac{|a_p - a_q|}{|a_p||a_q|} \leq \frac{4}{\rho^2}|a_p - a_q|$$

As $\{a_n\}$ is Cauchy, for $\frac{\rho^2\epsilon}{4}$ there is a N_2 such that for $p, q > N_2$ we have

$$|a_p - a_q| < \frac{\rho^2\epsilon}{4}$$

and because of this for $p, q > \max(N_1, N_2)$ we have

$$\left|\frac{1}{a_p} - \frac{1}{a_q}\right| \leq \frac{4}{\rho^2}|a_p - a_q| < \frac{4}{\rho^2} \cdot \frac{\rho^2\epsilon}{4} = \epsilon$$

And this shows that $\left\{\frac{1}{a_n}\right\}$ is a Cauchy sequence. Lastly as

$$\left\{\frac{1}{a_n}\right\} \cdot \{a_n\} = \{1\}$$

and $x = [\{a_n\}]$ and $1 = [\{1\}]$ we have that $\left[\left\{\frac{1}{a_n}\right\}\right]$ is the multiplicative inverse of x. □

With everything we have shown, we can close this section by saying that $(\mathbb{R}, +, \cdot)$ has the structure of a *field*.

Lecture 9 - 10/16/24

3.3 The order structure of $\mathbb{R}$

As mentioned earlier, some of the advantages of the Cauchy sequence construction of $\mathbb{R}$ is that the field axioms of $\mathbb{R}$ and the Cauchy completeness of the reals are relatively easy to prove. With that said, there is always a cost that comes due with any choice that is made, and what we will see is that the order structure on $\mathbb{R}$ is considerably more difficult to discuss because of real numbers being defined as equivalence classes of Cauchy sequences.

We begin with a fairly long theorem that will make defining $<$ on $\mathbb{R}$ possible.

Theorem 25. *If $x = [\{x_n\}] \in \mathbb{R}$ is nonzero, $x \neq 0$, then there exists $\delta \in \mathbb{Q}_+$ such that for any choice of representative $\{x_n\} \in x$ there is a $N \in \mathbb{N}$ such that for $n > N$, we have $|x_n| > \delta$.*

Going further for $x = [\{x_n\}] \in \mathbb{R}$, precisely one of the following hold:

a). For any $\{x_n\} \in x$ there exists $N \in \mathbb{N}$ such that for all $n > N$ we have $x_n > \delta$.

b). For any $\{x_n\} \in x$ there exists $N \in \mathbb{N}$ such that for all $n > N$ we have $x_n < -\delta$.

Before we give the proof of this result, let us consider what it is saying. If a real number $x \neq 0$, then any representative $\{x_n\}$ of $x = [\{x_n\}]$ eventually 'stays away from 0' in that for every choice of representative $\{x_n\}$ there is some point N in which the terms in the sequence stay away from 0. ($|x_n| > \delta$ for $n > N$) Do note that N, the point at which the sequence $\{x_n\}$ stays away from zero, can be different from representative to representative. If $\{x'_n\}$ was another representative of x, the value N' for which $|x'_n| > \delta$ for $n > N'$ could be larger or smaller than N.

Proof. We have that $x = [\{x_n\}] \neq 0 = [\{0\}]$, thus $\lim_{n\to\infty} |x_n - 0| \neq 0$. Because of this we have the negation of convergence

$$\neg(\forall \epsilon > 0, \exists N \in \mathbb{N}, s.t. \forall n > N, |x_n - 0| < \epsilon)$$
$$\exists \epsilon > 0, \forall N \in \mathbb{N}, s.t. \exists n > N, |x_n| \geq \epsilon$$

Thus for our initial choice of representative $\{x_n\}$ we have the existence of a specific value of $\epsilon \in \mathbb{Q}_+$ which we will call 4δ, in which for any $N \in \mathbb{N}$ there is some $n > N$ with $|x_n| > 4\delta$.

Note this is not saying a tail of the sequence $\{x_n\}$ has $|x_n| > 4\delta$ for all $n > N$, but this is saying that no matter which term you are at in the sequence x_p, there is some later term ($p < n$) with $|x_n| > 4\delta$. However, due to the nature of $\{x_n\}$ being a Cauchy sequence, we will be able to bootstrap our way from this result to the fact that a tail of $\{x_n\}$ stays bounded away from zero.

As $\{x_n\}$ is a Cauchy sequence, for $\epsilon = 2\delta$, we know there exists an $N_1 \in \mathbb{N}$ such that for all $p, q > N_1$ we have

$$|x_p - x_q| < 2\delta$$

Because of the condition two paragraphs above, for this N_1 there is a $n > N_1$ with $|x_n| > 4\delta$. And because of the Cauchy condition above we also have $|x_p - x_n| < 2\delta$ for $p > N_1$. Thus by the reverse triangle inequality, we have

$$|x_p| = |x_n - (x_n - x_p)| \geq ||x_n| - |x_n - x_p|| \geq |x_n| - |x_n - x_p| > 4\delta - 2\delta = 2\delta$$

for $p > N_1$. Thus we have just showed that a tail of the representative $\{x_n\}$ stays bounded away from 0. We now bootstrap from this result to why we can make a similar claim for some tail of any representative sequence for x.

If $\{x'_n\}$ is another representative sequence of x, then by definition we have that $\lim_{n\to\infty} |x_n - x'_n| = 0$. Thus taking $\epsilon = \delta$ there is some $N_2 \in \mathbb{N}$ such that for all $n > N_2$ we have

$$|x_n - x'_n| < \delta$$

By then taking $n > \max(N_1, N_2) = N'$ and using a similar argument with the reverse triangle inequality we have

$$|x'_n| = |x_n - (x_n - x'_n)| \geq ||x_n| - |x_n - x'_n|| \geq |x_n| - |x_n - x'_n| > 2\delta - \delta = \delta$$

And so we have $|x'_n| > \delta$ for $n > N'$. As this N' can be found for any choice of representative sequence $\{x'_n\}$ for x we have proven the first part of our claim.

Let us now prove the 'going further' part. We will follow a similar idea as our proof above in that we will prove the result for a single representative $\{x_n\}$ of x and then show how this enforces the result for any representative of x.

So, take $\{x_n\}$ to be a representative of $x \neq 0$. Because of what we have proven above there is a $\delta \in \mathbb{Q}_+$ and an N such that for all $n > N$ we have $|x_n| > \delta$. Put another way, we have

$$x_n > \delta \quad \text{or} \quad x_n < -\delta \quad \text{for } n > N$$

If we label these conditions *i*). $x_n > \delta$ and *ii*). $x_n < -\delta$, then as $n > N$ contains an infinite number of natural numbers we have the following possibilities:

1. Either *i*). happens for an infinite number of $n > N$ and *ii*). happens finitely often or vice-versa *ii*). happens infinitely oftern and *i*). happens finitely.
2. Both *i*). and *ii*). happen for an infinite number of terms $n > N$.

In case 1, if *ii*). happens finitely often, then just take

$$M = \max(n \mid x_n < -\delta)$$

and then for $n > \max(N, M)$ we have $x_n > \delta$. The argument is the same for the vice-versa case with the exception that $x_n < -\delta$ for $n > \max(N, M)$ and M is the maximum of the indices with $x_n > \delta$.

In case 2, if *i*). and *ii*). happen infinitely often, then for any $N \in \mathbb{N}$ there exists $p > N$ with $x_p > \delta$ and there exists $q > N$ with $x_q < -\delta$, but then we have

$$|x_p - x_q| = x_p - x_q > 2\delta$$

and as this can be done for any N, this contradicts that $\{x_n\}$ is Cauchy. So it must be that case 2 does not happen. Thus we have that the representative $\{x_n\}$ satisfies precisely one of a). and b). Without loss of generality, assume that $\{x_n\}$ satisfies a). and we will now show why any other representative $\{x'_n\}$ of x satisfies a). as well.

If $\{x'_n\}$ is another representative for x, then by the work above we know that $\{x'_n\}$ satisfies precisely one of a). or b). If $\{x'_n\}$ satisfies b). then we know that there exists N_1 such that $x_n > \delta$ for $n > N_1$ as $\{x_n\}$ satisfies a). and there exists N_2 such that $x'_n < -\delta$ for all $n > N_2$. But then we have

$$|x_n - x'_n| = x_n - x'_n > 2\delta$$

for all $n > \max(N_1, N_2)$, but this contradicts the fact that $\lim_{n\to\infty} |x_n - x'_n| = 0$, so it must be that $\{x'_n\}$ satisfies a). as well. $\square$

Now that the proof of this theorem is out of the way we are capable of defining the order structure on $\mathbb{R}$.

Definition 28. *For a non-zero real number $x = [\{x_n\}] \neq 0$, we say that x is* **positive** *and write $x > 0$ if there exists a $\delta \in \mathbb{Q}_+$ such that for any representative $\{x_n\} \in x$ there exists $N \in \mathbb{N}$ such that $x_n > \delta$ for all $n > N$.*

Similarly, we say that x is **negative** *and write $x < 0$ if there exists $\delta \in \mathbb{Q}_+$ such that for any representative $\{x_n\} \in x$ there exists $N \in \mathbb{N}$ such that $x_n < -\delta$ for all $n > N$.*

Lastly, we write $x > y$ for two real numbers x, y as shorthand for $x - y > 0$.

Now in the context of this definition the prior theorem was nothing more than the **trichotomy law** of $<$ for $\mathbb{R}$: i.e. that for two real numbers x, y precisely one of the following hold

$$x < y, \qquad x = y, \qquad x > y$$

Thus to see that $<$ forms a total order on $\mathbb{R}$ we must only check that $<$ is transitive.

Theorem 26. *The relation $<$ defined on $\mathbb{R}$ above satisfies the transitive property, i.e. if $x < y$ and $y < z$ then $x < z$.*

Proof. Assume that $x < y$ and $y < z$, then by definition we have $y - x > 0$ and $z - y > 0$. Take $\{y_n - x_n\}$ to be a representative of $y - x$ and $\{z'_n - y'_n$ to be a representative of $z - y$. Then there exists $\delta_1, \delta_2 \in \mathbb{Q}_+$ and $N_1, N_2 \in \mathbb{N}$ such that

$$y_n - x_n > \delta_1, \text{ for } n > N_1, \qquad z'_n - y'_n > \delta_2, \text{ for } n > N_2$$

Thus for $n > \max(N_1, N_2)$ we have that

$$z'_n - x_n + y_n - y'_n > \delta_1 + \delta_2$$

As $\{y_n\}$ and $\{y'_n\}$ are both representatives of y, we have that $\lim_{n\to\infty} |y_n - y'_n| = 0$. Thus for $\epsilon = \frac{\delta_1+\delta_2}{2}$ we have the existence of N_3 such that for all $n > N_3$,

$$|y_n - y'_n| < \frac{\delta_1 + \delta_2}{2} \iff -\frac{\delta_1 + \delta_2}{2} < y_n - y'_n < \frac{\delta_1 + \delta_2}{2}$$

We then have for $n > \max(N_1, N_2, N_3)$ that

$$z'_n - x_n + \frac{\delta_1 + \delta_2}{2} > z'_n - x_n + y_n - y'_n > \delta_1 + \delta_2$$

Thus we have

$$z'_n - x_n > \frac{\delta_1 + \delta_2}{2}, \text{ for } n > \max(N_1, N_2, N_3)$$

At this point we will use a common trick we have used many times involving the reverse triangle inequality. If $\{z''_n - x'_n\}$ was any other representative of $z - x$, then for $\epsilon = \frac{\delta_1+\delta_2}{4}$ there exists a $N_4 \in \mathbb{N}$ such that for all $n > N_4$ we have

$$|(z'_n - x_n) - (z''_n - x'_n)| < \frac{\delta_1 + \delta_2}{4}$$

we then have for $n > \max(N_1, N_2, N_3, N_4)$ that

$$\begin{aligned} |z_n'' - x_n'| &= |(z_n' - x_n) - [(z_n' - x_n) - (z_n'' - x_n')]| \\ &\geq |z_n' - x_n| - |(z_n' - x_n) - (z_n'' - x_n')| \\ &= (z_n' - x_n) - |(z_n' - x_n) - (z_n'' - x_n')| \\ &> \frac{\delta_1 + \delta_2}{2} - \frac{\delta_1 + \delta_2}{4} = \frac{\delta_1 + \delta_2}{4} \end{aligned}$$

and from our the proof of our prior theorem it must be that $z_n'' - x_n' > \frac{\delta_1+\delta_2}{4}$ as all representatives satisfy either condition a). or b). in the last theorem. Thus the constant $\frac{\delta_1+\delta_2}{4} \in \mathbb{Q}_+$ is a lower bound on all representative sequences of $z - x$ and thus $z - x > 0$ and so $x < z$. □

With this theorem concluded we have shown that $<$ is a total order structure on $\mathbb{R}$. It is left to see that $<$ obeys the properties that make $\mathbb{R}$ into a *totally ordered field*. Recall that the conditions of a field X being a totally ordered field were

1. If $a < b$ and $c \in X$ implies that $a + c < b + c$.
2. If $a > 0$ and $b > 0$, then $ab > 0$.

However, we will instead show that the positive real numbers are closed under addition and multiplication

1*. If $a > 0$ and $b > 0$, then $a + b > 0$.
2. If $a > 0$ and $b > 0$, then $ab > 0$.

To see that this equivalent to the ordered field properties, we must check that the two versions of statement 1 are equivalent.

Let us assume 1 and that $a > 0$ and $b > 0$, then as a holds, we can add b to both sides of $a > 0$ and we have

$$0 < b = b + 0 < b + a$$

and this shows that $a + b > 0$ when $a > 0$ and $b > 0$, thus $1 \implies 1^*$.

Now let us assume 1^* and assume that $a < b$. We argue by contradiction, thus assume 1 does not hold, i.e. there exists $c \in X$ such that $a + c \geq b + c$. If $a + c = b + c$, then it must be $a = b$ by the field cancellation laws but this contradicts $a < b$, thus it must be that $a + c > b + c$. We thus have $b - a > 0$ and $(a + c) - (b + c) > 0$, and thus by 1^* it must be that

$$(b - a) + (a + c) - (b + c) > 0$$

But after distributing and cancelling on the left side we end with $0 > 0$ which is a contradiction, thus it must be that $1^* \implies 1$.

Theorem 27. *For real numbers x, y we have the following*

i). If $x > 0$ and $y > 0$, then $x + y > 0$.

ii). If $x > 0$ and $y > 0$, then $xy > 0$.

Proof. For part i). let $\{x_n\}$ be a representative of x and $\{y_n\}$ a representative of y. Then there exists $\delta_1, \delta_2 \in \mathbb{Q}_+$ and $N_1, N_2 \in \mathbb{N}$ respectively such that

$$x_n > \delta_1, \text{ for } n > N_1, \qquad y_n > \delta_2, \text{ for } n > N_2$$

Thus for $n > \max(N_1, N_2)$ we have $x_n + y_n > \delta_1 + \delta_2$. At this point the standard argument involving the reverse triangle inequality can be used to show why

$$x'_n + y'_n > \frac{\delta_1 + \delta_2}{2}$$

for $n > N$ for some N if $\{x'_n + y'_n\}$ is any other representative of $x + y$. Thus $x + y > 0$.

For part ii). for $n > \max(N_1, N_2)$ we have that

$$x_n y_n > \delta_1 \delta_2$$

Similarly, at this point the standard argument involving the reverse triangle inequality can be used to show why

$$x'_n y'_n > \frac{\delta_1 \delta_2}{2}$$

for $n > N$ for some N if $\{x'_n y'_n\}$ is any other representative of xy. Thus $xy > 0$. □

Thus up to this point we have shown that $\mathbb{R}$ is a *totally ordered field.*

Lecture 10 - 10/18/24

3.4 The density of $\mathbb{Q}$ and the Archimedean Property

Lecture 11 - 10/21/24

Before we move onto the completeness of the real number system, we collect two other important properties that the reals have. The first is that the reals will inherit the Archimedean property from the rational numbers and so there are no infinitely large or infinitesimally small numbers in $\mathbb{R}$. The second is that the rational numbers $\mathbb{Q}$ are **dense** inside of $\mathbb{R}$ in that we can find a rational number arbitrarily close to any real number we pick. This will prove to be incredibly useful in future arguments.

Before we prove these properties, in this section we will begin proving statements about real numbers themselves, not representatives of these real numbers. In particular, our proofs will rely upon the absolute value function and the triangle inequality, which held in all prior sections up to now as they were applied to rational numbers. But now we would like to apply these tools to real numbers. This will be no problem. If one looks at our definition of absolute value and proof of the triangle inequality for $\mathbb{Q}$, we required nothing more than the rationals being a totally ordered field. As our last section finished with $\mathbb{R}$ being a totally ordered field, we have that $|x|$ is defined[38] for real numbers and that the triangle inequality also holds for $\mathbb{R}$.

One may remember from previous math classes that we typically require a certain amount of 'similarity' between objects if we are to perform operations between them, i.e. for example you

[38] using the same definition from $\mathbb{Q}$

can not add a vector to a matrix or add two vectors coming from different dimensional spaces[39]. In a similar manner, we are about to compare rational numbers to real numbers, and a rational number is just that, a number $q \in \mathbb{Q}$. But to us, real numbers x are equivalence classes of Cauchy sequences of real numbers $x = [\{x_n\}]$. Because of this we can not directly add q and x. We remedy this by seeing that there is a 'copy' of $\mathbb{Q}$ sitting inside of $\mathbb{R}$ by sending q to the class of the constant sequence $[\{q\}_{n=1}^{\infty}]$.

Definition 29. *The function $i : \mathbb{Q} \to \mathbb{R}$ given by $i(q) = [\{q\}_{n=1}^{\infty}]$, which maps each rational number q to the equivalence class of the constant sequence $\{q\}$ is often called the* **injection mapping** *from the rational numbers to the reals. Also note that*

$$i(q+r) = i(q) + i(r), \quad \text{and} \quad i(qr) = i(q)i(r)$$

from our previous results about the well-definedness of addition and multiplication in $\mathbb{R}$.

Before we prove the Archimedean Property and the density of $\mathbb{Q}$ inside of $\mathbb{R}$ we will prove a lemma that will come up often in this section and the next.

Lemma 28. *For real numbers $x = [\{x_n\}]$ and $y = [\{y_n\}]$ if there exists representatives $\{x_n\}$,$\{y_n\}$ of x, y respectively and an $N \in \mathbb{N}$ such that for $n > N$ it is true that $x_n \geq y_n$, then $x \geq y$.*

Proof. We argue by contradiction, assume that $x_n \geq y_n$ for $n > N$ and $\{x_n\}, \{y_n\}$ representatives of x, y and that $x < y$. Thus $y - x > 0$ and so there exists $\delta \in \mathbb{Q}_+$ such that for any representative $\{y_n - x_n\}$ there is some $N \in \mathbb{N}$ such that

$$y_n - x_n > \delta$$

for $n > N$. However, this is true for our initial choices of representatives of x, y as well, thus there is some N such that for all $n > N$ we have

$$x_n \geq y_n > x_n + \delta$$

and this is a contradiction as it implies that $\delta < 0$. □

Theorem 29. (The Archimedean Principle of $\mathbb{R}$)*: For every real number x with $x > 0$ there exists a natural number N such that $0 < \frac{1}{N} < x$.*

Recall, as we saw in section 1.4 that there are two equivalent versions of the Archimedean principle. We will prove this one, the equivalence of the second version is the same as it was in that prior section.

Proof. As $x > 0$, there exists a $\delta \in \mathbb{Q}_+$ such that for any representative $\{x_n\}$ of x there is an $N \in \mathbb{N}$ such that for all $n > N$ we have $x_n > 2\delta$. Making use of the lemma we just proved we then have that $2i(\delta) \leq x$.

From the Archimedean Principle on $\mathbb{Q}$, for $\delta > 0$ we know the existence of a natural number N with the property that $0 < \frac{1}{N} < \delta$. Thus our lemma again gives that

$$0 < i\left(\frac{1}{N}\right) \leq i(\delta)$$

[39] at least without embedding the smaller space in the larger

And so putting both results together we have that

$$0 < i\left(\frac{1}{N}\right) < x$$

□

Remark 3. *It is an unfortunate bad habit that oftentimes in mathematics the correct formal notation is dropped or forgotten for the sake of shorthand or convenience. There is value in this after formal arguments that are completed. With that said, people do not typically write the injection mapping for a rational number. For $q \in \mathbb{Q}$ it is more formal to say $i(q)$ is the version of q in $\mathbb{R}$, but after finishing the construction of $\mathbb{R}$ we will do away with the equivalence classes of Cauchy sequences and simply refer to objects in $\mathbb{R}$ as numbers, and in this sense $i(q) = q$. Hence the reason the statement of the Archimedean principle above was given in the form above.*

Theorem 30. (Density of $\mathbb{Q}$ in $\mathbb{R}$)*: For every real number x and any $\epsilon > 0$ there exists a rational number q such that*

$$|x - q| < \epsilon$$

i.e. any real number x has rational numbers that are arbitrarily close to it.

Proof. Let $\epsilon > 0$, by the Archimedean property, there is a natural number N such that $0 < i\left(\frac{1}{N}\right) < \epsilon$. As $\{x_n\}$ is Cauchy there is another rational number N' such that for $p, q > N'$ we have

$$|x_p - x_q| < \frac{1}{N}, \quad x_q - \frac{1}{N} < x_p < x_q + \frac{1}{N}$$

Fix q to be equal to $N' + 1$, and for the sake of simplicity call $r = x_{N'+1}$. Thus for $p > N'$ we have

$$r - \frac{1}{N} < x_p < r + \frac{1}{N}$$

Thinking of $r - \frac{1}{N}$ and $r + \frac{1}{N}$ as the tails of two constant sequences, using our lemma above we have that

$$i(r) - i\left(\frac{1}{N}\right) \leq x \leq i(r) + i\left(\frac{1}{N}\right)$$

which is equivalent to

$$|x - i(r)| \leq i\left(\frac{1}{N}\right) < \epsilon$$

and this proves the result □

3.5 The completeness of $\mathbb{R}$

At this point we will consider sequences of real numbers, $\{x_n\}$ with $x_n \in \mathbb{R}$ for all $n \in \mathbb{N}$. Our definitions of Cauchy and convergence will remain the same with the exception that we will say $\epsilon > 0$ instead of $\epsilon \in \mathbb{Q}_+$, i.e. we will allow ϵ to be any positive real value. Many of our previous theorems about sequences of rational numbers will pass through to theorems about sequences of real numbers, in particular the Algebraic Limit Rules, The Squeeze Lemma, and the interaction of sequences with order $<$ will pass through.

As we make our way towards showing that the real numbers are complete[40] we first prove a result that we have been dancing around for awhile. We created the reals, or the specific numbers in the reals, as equivalence classes of Cauchy sequences, i.e. a collection of roads traveling to the same place without giving name to the destination. But we have created the real numbers now, so these destinations now exist and as such, we can speak about them. The next theorem simply verifies something we believed or intuited to be true, but does so formally with the structures we have created.

Theorem 31. *For a real number $x = [\{x_n\}]$ and any representative $\{x_n\}$ of x, the injection map $i : \mathbb{Q} \to \mathbb{R}$ given by $i(r) = [\{r\}]$ defines a sequence of real numbers $\{i(x_n)\}$ and this sequence converges to x.*

Proof. Let $\epsilon > 0$, and without loss of generality by picking a smaller value or making use of the Archimedean property assume that $\epsilon \in \mathbb{Q}_+$. As $\{x_n\}$ is a Cauchy sequence, there is an $N \in \mathbb{N}$ such that for $p, q > N$ we have

$$|x_p - x_q| < \frac{\epsilon}{2}$$

and this gives

$$x_q - \frac{\epsilon}{2} < x_p < x_q + \frac{\epsilon}{2}, \quad \text{for } p, q > N$$

The sequence $\left\{x_q - \frac{\epsilon}{2}\right\}$ is a representative for $x - \frac{\epsilon}{2}$ and similarly $\left\{x_q + \frac{\epsilon}{2}\right\}$ is a representative for $x + \frac{\epsilon}{2}$. Using the injection map, the constant sequence $\{x_p\}_n$ is a representative of $i(x_p) = [\{x_p\}_n]$, and so Lemma 28 gives that

$$x - \frac{\epsilon}{2} \leq i(x_p) \leq x + \frac{\epsilon}{2}, \quad \text{for } p > N$$

which is equivalent to

$$|i(x_p) - x| \leq \frac{\epsilon}{2} < \epsilon, \quad \text{for } p > N$$

and since this can be done for any $\epsilon > 0$, we have that $\{i(x_n)\} \to x$. □

We already knew that Cauchy sequences 'clustered infinitely', and as we saw with $\mathbb{Q}$, what kept a Cauchy sequence from converging was that it's limit did not exist in $\mathbb{Q}$. The theorem above simply says that our construction of the reals fills in those specific 'gaps', i.e. that for any rational sequence $\{r_n\}$ in the equivalence class of x converges to x in $\mathbb{R}$.

However, considering that we have created a new number system $\mathbb{R}$, we can consider sequences of real numbers, i.e. $\{x_n\}$ with $x_n \in \mathbb{R}$ for all $n \in \mathbb{N}$. The next theorem proves the Cauchy completeness of $\mathbb{R}$. In context of our current discussion, this means when we move to $\mathbb{R}$ there are no new kinds of 'gaps' in the reals, i.e. that any Cauchy sequence of real numbers will converge to a real number.

Theorem 32. (The Cauchy Completeness of $\mathbb{R}$) - *A sequence $\{x_n\}$ of real numbers converges if and only if it is a Cauchy sequence.*

Proof. $\implies$ The direction of $\{x_n\}$ convergent implies that $\{x_n\}$ is Cauchy was proven earlier for sequences of rational numbers and the proof is identical for sequences of real numbers.

[40] Both Cauchy complete and satisfying the Least Upper Bound property

$\Longleftarrow$ Assume that $\{x_n\}$ is a Cauchy sequence. For each x_m in the sequence $\{x_n\}$ by the density of $\mathbb{Q}$ in $\mathbb{R}$ there exists a rational number r_m such that

$$|x_m - i(r_m)| < \frac{1}{m}$$

We make the argument that $\{r_n\}$ is a Cauchy sequence of rational numbers. We first write the following

$$\begin{aligned} |r_p - r_q| = |i(r_p) - i(r_q)| &= |i(r_p) - x_p + x_p - x_q + x_q - i(r_q)| \\ &\leq |i(r_p) - x_p| + |x_p - x_q| + |x_q - i(r_q)| && \text{Triangle Inequality} \\ &< \frac{1}{p} + |x_p - x_q| + \frac{1}{q} \end{aligned}$$

Taking $\epsilon > 0$ to be an arbitrary fixed error term, we know as $\{x_n\}$ is Cauchy that there is an $N \in \mathbb{N}$ such that for all $p, q > N$ we have

$$|x_p - x_q| < \frac{\epsilon}{3}$$

By possibly increasing N if need be, we can assume by the Archimedean Principle that $\frac{1}{p}, \frac{1}{q} < \frac{\epsilon}{3}$ for $p, q > N$ as well. Thus for p, q larger than this N we have

$$|r_p - r_q| < \frac{1}{p} + |x_p - x_q| + \frac{1}{q} < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

And as this can be done for any choice of ϵ we have that $\{r_n\}$ is Cauchy. However, $\{r_n\}$ being Cauchy means that it is a representative of a real number, $x = [\{r_n\}]$. We claim that $\{x_n\} \to x$. Looking at the following expression

$$|x_n - x| = |x_n - i(r_n) + i(r_n) - x| \leq |x_n - i(r_n)| + |i(r_n) - x| < \frac{1}{n} + |i(r_n) - x|$$

where the first inequality comes from the Triangle inequality and the second comes from how we choose r_n from the density of the rationals. By our previous theorem, we have that as $x = [\{r_n\}]$ that $\{i(r_n)\} \to x$. Thus there is an $N \in \mathbb{N}$ such that $n > N$ implies that

$$|i(r_n) - x| < \frac{\epsilon}{2}$$

and by possibly increasing N we can assume that $\frac{1}{n} < \frac{\epsilon}{2}$ for $n > N$ by the Archimedean Principle. And so

$$|x_n - x| < \frac{1}{n} + |i(r_n) - x| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

And as this can be done for any choice of ϵ we have $\{x_n\} \to x$ and thus is convergent. □

To close off this section, we prove that $\mathbb{R}$ has the least upper bound property as well. This is beyond the scope of this course, but generally Cauchy completeness of a set X is weaker than completeness of X in the sense of having the least upper bound property. However, when X has the Archimedean property, these notions collapse and become equivalent. With that said, see if you can spot where the Archimedean property is required in the next proof.

Lemma 33. (Finite Geometric Sum)*: For any real number x, we have the following*

$$\sum_{k=0}^{n-1} x^k = \frac{1 - x^n}{1 - x}$$

Proof. Write out the left hand side as

$$S = 1 + x + x^2 + \cdots + x^{n-1}$$

if we multiply this by x and subtract we see

$$\begin{aligned} S &= 1 + x + x^2 + \cdots + x^{n-1} \\ -xS &= \quad -x - x^2 - \cdots - x^{n-1} - x^n \\ S - xS &= 1 - x^n \end{aligned}$$

and so by factoring the left side as $(1-x)S$ and dividing we see the result. □

Lecture 12 - 10/23/24

Theorem 34. (The L.U.B. property)- *The reals, $\mathbb{R}$, satisfy the least upper bound property, i.e. for $A \subseteq \mathbb{R}$ a nonempty subset that is bounded above, then* $\sup A$ *exists within $\mathbb{R}$.*

Proof. Assume that A is a non-empty subset of $\mathbb{R}$ that is bounded above. As A is bounded above, there exists a $y \in \mathbb{R}$ such that $y \geq a$ for all $a \in A$. Without loss of generality, by taking a value larger than this specific y, we can assume that y is rational. Similarly, as $A \neq \varnothing$, there is some $x \in A$. By possibly taking a value smaller than x we can assume that x is rational.

We will define two sequences, $\{a_n\}$ and $\{b_n\}$ and we initialize their values as $a_1 = x$ and $b_1 = y$. We then look at their midpoint $m_2 = \frac{a_1+b_1}{2}$:

- If m_2 is an upper bound of A then we define $a_2 = a_1 = x$ and $b_2 = m_2$.
- If m_2 is not an upper bound of A, then we define $b_2 = b_1 = y$ and $a_2 = m_2$.

and we continue onward in such a manner, i.e. we define $m_{n+1} = \frac{a_n+b_n}{2}$ and define:

- If m_{n+1} is an upper bound of A we define $a_{n+1} = a_n$ and $b_{n+1} = m_{n+1}$.
- If m_{n+1} is not an upper bound of A we define $a_{n+1} = m_{n+1}$ and $b_{n+1} = b_n$.

Using this, we have created sequences $\{a_n\}$ and $\{b_n\}$.

Generally, we have

$$b_n - a_n = \begin{cases} b_{n-1} - m_{n-1} \\ m_{n-1} - a_{n-1} \end{cases} = \frac{b_{n-1} - a_{n-1}}{2}$$

and from this we see that

$$b_n - a_n = \frac{y-x}{2^{n-1}}$$

which shows that $\{b_n - a_n\}$ converges to 0.

We now show that $\{a_n\}$ is a Cauchy sequence. Assuming that $p > q$ we first have

$$|a_p - a_q| = \left| \sum_{k=0}^{p-q-1} a_{q+k+1} - a_{q+k} \right| \leq \sum_{k=0}^{p-q-1} |a_{q+k+1} - a_{q+k}|$$

by the general triangle inequality. Now the term $a_{q+k+1} - a_{q+k}$ is one of two things

$$a_{q+k+1} - a_{q+k} = \begin{cases} 0 \\ m_{q+k+1} - a_{q+k} = \frac{1}{2}(b_{q+k} - a_{q+k}) = \frac{y-x}{2^{q+k}} \end{cases}$$

And because of this we have

$$|a_p - a_q| \leq \left(\frac{y-x}{2^q}\right) \sum_{k=0}^{p-q-1} \frac{1}{2^k}$$

and by using the formula for a finite geometric series we have

$$|a_p - a_q| \leq \left(\frac{y-x}{2^q}\right)\left(\frac{1 - \frac{1}{2^{p-q}}}{1 - \frac{1}{2}}\right) = (y-x)\left(\frac{1}{2^{q-1}} - \frac{1}{2^{p-1}}\right)$$

From a previous homework, we have that $\lim_{n\to\infty} r^n = 0$ for $0 \leq r < 1$ and this shows why the right hand side of the above can be made arbitrarily small for p, q large enough (as y, x are fixed). Thus $\{a_n\}$ is Cauchy. And by our previous result, $\{a_n\}$ is convergent and thus converges to $a \in \mathbb{R}$. Using our algebraic limit laws, as

$$b_n = (b_n - a_n) + a_n$$

we have that $\{b_n\}$ is convergent, say to b, and as $\lim_{n\to\infty}(b_n - a_n) = 0$, we have that $a = b$.

To finish our proof, we claim that $b = \sup A$. If b was not an upper bound of A, there would exist $s \in A$ with $b < s$. Taking $\epsilon = \frac{s-b}{2}$, as $\{b_n\} \to b$, there exists $N \in \mathbb{N}$ such that for all $n > N$

$$|b_n - b| < \frac{s-b}{2}, \iff b_n < \frac{s+b}{2} < s$$

and this is a contradiction as in the construction of sequence $\{b_n\}$, every term b_n was an upper bound of A. Thus we have that b is an upper bound of A, and we only need to argue that b is the least upper bound of A.

If p was another upper bound of A with $p < b$, then taking $\epsilon = \frac{b-p}{2}$, as $\{a_n\} \to b$, there exists $N \in \mathbb{N}$ such that for all $n > N$ we have

$$|a_n - b| < \frac{b-p}{2}, \iff p < \frac{p+b}{2} < a_n$$

and in the construction of the sequence $\{a_n\}$, each term a_n was not an upper bound of A. Thus $p < a_{N+1}$ and as a_{N+1} is not an upper bound of A, we have that p is not an upper bound of A, which is a contradiction. Thus it must be that b is the least upper bound of A, thus $b = \sup A \in \mathbb{R}$. □

Exercises for section 3.5:

1. For $x \in \mathbb{R}$ with $x > 0$ prove there exists a unique $y > 0$ with $y^2 = x$. (i.e. this is the existence of the square root for positive x.)

   Hint: For this argument, follow the proof of the least upper bound property. If $x < 1$, then $x^2 < x < 1$ and label $y_1 = x$ and $z_1 = 1$. If $x > 1$, then $1 < x < x^2$ and label $y_1 = 1$ and $z_1 = x$. Call $m_2 = \frac{y_1+z_1}{2}$ and think about how one should define y_2, z_2 if $m_2^2 > x$ or $m_2^2 < x$ respectively.

2. In reference to number 1. can you extrapolate this argument to show how for $x \in \mathbb{R}$ with $x > 0$ the nth root of x exists?

4 Sequences (Reprise)

Lecture 13 - 10/28/24

Now that we have constructed the real numbers $\mathbb{R}$ we would like to see some further properties about sequences of real numbers. In particular, for any sequence of real numbers, $\{x_n\}$, a set can be formed that is made up precisely of the terms of this sequence

$$A = \{x_n \mid n \in \mathbb{N}\}$$

For the sequences $\{\frac{1}{n}\}_{n=1}^{\infty}$ and $\{1\}_{n=1}^{\infty}$, the set A would be $\{1, \frac{1}{2}, \frac{1}{3}, \dots\}$ and $\{1\}$ respectively. In a similar vein, it will be possible to create sequences out of elements from a specified set, B. The goal of this section is to explore some initial connections between sets formed from sequences and sequences formed from sets. We will primarily focus on sequences in this section (hence the title), but in a later section on the topology of $\mathbb{R}$ we will explore the other direction in more detail.

4.1 Properties of Supremum and Infinimum Operations

In this section we collect some theorems about the supremum and infimum of sets. Recall that the supremum of a set is the least upper bound of a set and the infimum is the greatest lower bound. The operations of finding the supremum or infimum of a set generalize the operations of finding a maximum or minimum of a set. For example, the set

$$A = \left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\right\}$$

fails to have a maximum value. The reason for this can be thought of in two ways, the first is that the maximum function is only guaranteed to output a value when there is a finite number of inputs. Or phrased in terms of a computer program, it is believable to algorithmically find the largest element of a finite number of elements in a finite amount of time, but searching through an infinite list could literally take forever. The second is that the elements in A are tending to a value, namely 1, and 1 would be the maximum of A if it existed within A but it does not.

However A does have a supremum within $\mathbb{R}$ as it is a subset that is bounded above, and as we will see shortly $\sup A = 1$. Thus the supremum operation can help us generalize the maximum function in that it will tell us the next best thing, which is the smallest number that bounds every term in A.

Theorem 35. *The supremum and infimum of a set (if they exist) are unique.*

Proof. Assume A is a set that is bounded above, and that α and β are both $\sup A$. As β is an upper bound of A, and α is the least upper bound of A we have that $\alpha \leq \beta$. As α is an upper bound of A, and β is the least upper bound of A we have that $\beta \leq \alpha$. Thus $\alpha = \beta$, and therefore the supremum of A is unique.

The proof of the uniqueness of infimum is similar. □

Theorem 36. *For a set A that is bounded above, a number α equals $\sup A$ if and only if α is an upper bound of A, and for all $\epsilon > 0$, there exists an $x \in A$ such that $x > \alpha - \epsilon$.*

Proof. $\implies$ Assume that $\alpha = \sup A$.

Then by definition, α is an upper bound of A as it is the least upper bound of A. To prove the second part of the claim, let us proceed by way of contradiction and assume there is some $\epsilon > 0$ for which $x > \alpha - \epsilon$ for no $x \in A$, i.e. for this $\epsilon > 0$ we have $x \leq \alpha - \epsilon$ for all $x \in A$.

But then $\alpha - \epsilon$ is an upper bound of A that is smaller than α, which is the least upper bound of A, and this is a clear contradiction.

$\impliedby$ Assume that α is an upper bound of A, and that for every $\epsilon > 0$ there exists an $x \in A$ such that $x > \alpha - \epsilon$.

By assumption, α is an upper bound A. We need to prove that α is the least upper bound of A, i.e. that if β is any other upper bound of A, then $\alpha \leq \beta$.

Thus, by way of contradiction, let us assume that β is an upper bound of A with $\beta < \alpha$. Then define $\epsilon = \alpha - \beta$. By assumption, for this ϵ, there exists an $x \in A$ such that

$$\begin{aligned} x &> \alpha - \epsilon = \alpha - (\alpha - \beta) \\ x &> \beta \end{aligned}$$

But then β is not an upper bound of A, thus we have a contradiction.

Thus α is the least upper bound of A, so $\alpha = \sup A$. □

An analogous result involving the infimum of a set that is bounded below will be left for the exercises.

Example 13. *Let us find the supremum of the following set*

$$A = \left\{ \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots \right\}$$

The set is of the form

$$A = \left\{ \frac{n}{n+1} \mid n \in \mathbb{N} \right\} = \left\{ 1 - \frac{1}{n+1} \mid n \in \mathbb{N} \right\}$$

From this we can conjecture that 1 is the supremum of this set. Let us use our prior theorem to prove this.

As each $x \in A$ is of the form $x = 1 - \frac{1}{n+1}$ for n a natural number, it is clear that 1 is an upper bound as

$$1 - \frac{1}{n+1} < 1, \qquad \forall n \in \mathbb{N}$$

Next, we need to prove that for all $\epsilon > 0$, there is some $x \in A$ with $x > 1 - \epsilon$. So, let ϵ be a some fixed arbitrarily small number, and let us perform some algebra to see what the condition $x > 1 - \epsilon$ is equivalent to.

$$\begin{aligned} x &> 1 - \epsilon \\ 1 - \frac{1}{n+1} &> 1 - \epsilon \\ \frac{1}{n+1} &< \epsilon \\ \frac{1}{\epsilon} - 1 &< n \end{aligned}$$

And this is clearly feasible. As small as ϵ is, as it is not equal to zero, it's reciprocal is some finite large number. And as the natural numbers go on forever, there is some n with this property. [41]

So, back to our proof. For a chosen $\epsilon > 0$, take $n \in \mathbb{N}$ to be a natural number with $\frac{1}{\epsilon} - 1 < n$, then

$$x = 1 - \frac{1}{n+1} > 1 - \epsilon$$

Thus, there exists $x \in A$ with $x > 1 - \epsilon$. And this can be done for all $\epsilon > 0$.

Therefore, by our prior theorem, we have that $\sup A = 1$.

The next theorem is a slight generalization of the previous theorem with a subtly different initial criteria. It does hint at connection between sets and sequences that I will explain more after the proof.

Theorem 37. *Assume that A is a set in the reals that is bounded above, let $\alpha = \sup A$ and assume that $\alpha \notin A$. Then for any $\epsilon > 0$ there is an infinite number of elements of A contained within the interval $(\alpha - \epsilon, \alpha)$.*

Proof. By way of contradiction, we will assume for a fixed given $\epsilon > 0$ there is only a finite number of values from A contained within $(\alpha - \epsilon, \alpha)$.

Since the number of elements from A is finite, there is a fixed $N \in \mathbb{N}$ in which we can label all the elements from A in $(\alpha - \epsilon, \alpha)$ by $x_1, x_2, x_3, x_4, \ldots, x_N$.

Without loss of generality, we can further assume that these values from A are listed in order, i.e.

$$\alpha - \epsilon < x_1 < x_2 < \cdots < x_N < \alpha$$

Thus, it is implicit here that x_N is the element in A closest to α.

Now, take $\epsilon_1 = \alpha - x_N > 0$. By our previous theorem, for this ϵ_1, there exists some $y \in A$ with $\alpha > y > \alpha - \epsilon_1$.

$$y > \alpha - \epsilon_1 = \alpha - (\alpha - x_N) = x_N.$$

But then y is an element of A within $(\alpha - \epsilon, \alpha)$ closer to α than x_N, and this is a contradiction. □

Much like our example from the start of the section, the maximum of a set that is bounded above may not exist, but the supremum of the set will exist in $\mathbb{R}$, and the theorem above states that if the maximum does not exist within the set then there is at least an infinite number of elements in the set clustering near its supremum. In particular, when paired with the Axiom of Choice, for a set A that is bounded above but $\sup A \notin A$, this theorem lays out a process of creating a sequence $\{x_n\}$ with $x_n \in A$ for all $n \in \mathbb{N}$ with $\{x_n\} \to \sup A$.[42] Once again, there is a similar result for the infimum of a set that is bounded below, but it will be left for the exercises.

Example 14. *Let $A = \{-1, 2, 4\}$.*

Then it is clear that $\sup A = 4$ *and* $\inf A = -1$. *We also have*

$$3A = \{-3, 6, 12\}, \qquad -2A = \{2, -4, -8\}$$

and that $\sup 3A = 12, \inf 3A = -3, \sup(-2A) = 2$ *and* $\inf(-2A) = -8$.

[41] Archimedean Property

[42] this will be a theorem in the next section

In particular we have the following theorem

Theorem 38. *For A a bounded set of the reals, we have that*

$$\sup(cA) = c\sup A, \qquad \inf(cA) = c\inf(A)$$

for a real constant $c > 0$, and

$$\sup(cA) = c\inf(A), \qquad \inf(cA) = c\sup A$$

for a real constant $c < 0$.

Proof. We will provide proof for the result when $c > 0$, the result for $c < 0$ will be left as an exercise.

Let $\alpha = \sup(cA)$ and $\beta = \sup(A)$. As β is an upper bound of A we have $\beta \geq a$ for all $a \in A$. Thus $c\beta \geq ca$ for all $a \in A$, thus $c\beta$ is an upper bound of cA. Taking $\epsilon > 0$, then $\frac{\epsilon}{c} > 0$, and thus as $\beta = \sup A$, there exists some $b \in A$ such that

$$\beta - \frac{\epsilon}{c} < b$$

Multiplying both sides by c we have that

$$c\beta - \epsilon < cb \in cA$$

As this can be done for any choice of $\epsilon > 0$, by our prior theorem it must be that $c\beta = \alpha$.

For the proof involving infimums, similarly let $\alpha = \inf(cA)$ and $\beta = \inf(A)$. We have that $\beta \leq a$ for all $a \in A$ and so $c\beta \leq ca$ for all $a \in A$, thus $c\beta$ is a lower bound of cA. And then taking $\epsilon > 0$, $\frac{\epsilon}{c} > 0$, and as $\beta = \inf(A)$ we know there exists a $y \in A$ with

$$y < \beta + \frac{\epsilon}{c}$$

and so by multiplying by c we have

$$cy < c\beta + \epsilon$$

And as this can be done for any choice of $\epsilon > 0$ this gives that $c\beta = \alpha$. □

Exercises for section 4.1:

1. What is the analogous theorem to theorem 36 for infimums of sets. Please provide proof of this result.
2. What is the analogous theorem to theorem 37 for infimums of sets. Please provide proof of this result.
3. Prove the $c < 0$ case of theorem 38.

4.2 The Monotone Convergence Theorem

In previous sections we saw that the convergence of a sequence implied that it was bounded and similarly that a sequence being Cauchy implied that it was bounded. We also saw examples of bounded sequences that were neither convergent nor Cauchy (in particular the sequence $\{(-1)^n\}$), and so 'boundedness' of a sequence is a weaker condition than that of convergence or Cauchy. However, we will see in this section that with one extra condition on a bounded sequence, we can guarantee convergence.

Definition 30. *A sequence $\{x_n\}$ is called* **monotonically increasing** *(resp.* **monotonically decreasing***) if $x_{n+1} \geq x_n$ (resp. $x_{n+1} \leq x_n$) holds for all $n \in \mathbb{N}$. A sequence is called* **strictly increasing** *(resp.* **strictly decreasing***) if $x_{n+1} > x_n$ (resp. $x_{n+1} < x_n$) holds for all $n \in \mathbb{N}$.*

Example 15. *Let us look at the following sequences.*

a). The sequence $\{1 - \frac{1}{n}\}$ is strictly increasing.

b). The sequence $\{\frac{1}{2^n}\}$ is strictly decreasing.

c). The sequence $\{(-1)^n\}$ is neither increasing or decreasing, and it does not converge.

d). The sequence $\{2 + \frac{(-1)^n}{n}\}$ is neither increasing nor decreasing, but it does converge to 2.

The even terms of this sequence are of the form $2 + \frac{1}{2n}$, and the odd terms are of the form $2 - \frac{1}{2n+1}$. Thus the even terms are always above 2 (and actually decreasing down to 2), and the odd terms are always below 2 (and actually increase up to 2).

Theorem 39. (Monotone Convergence Theorem) - *A bounded montone sequence converges. More specifically, a monotonically (strictly) increasing sequence that is bounded above converges, and a monotonically (strictly) decreasing sequence that is bounded below converges.*

Proof. Without loss of generality, assume that the sequence $\{x_n\}$ we are working with is monotonically increasing (the proof for it being monotonically decreasing is similar). Define the following set

$$A = \{x_1, x_2, x_3, \dots\}$$

i.e. A is just the elements of the sequence listed out. By assumption, as the sequence is bounded, A is bounded above, and therefore $\sup A$ exists. So, call $\alpha = \sup A$.

Now, let $\epsilon > 0$ be an arbitrarily chosen fixed number. For this ϵ, via a previous theorem about suprema, we know that there exists an $x \in A$ with the property that $x > \alpha - \epsilon$. This means there exists $N \in \mathbb{N}$ such that $x_N > \alpha - \epsilon$ as every element of A is some term in the sequence.

By the assumption of monotonicity (monotonically increasing) we have that $x_n \geq x_N$ for all $n > N$. So we have

$$\alpha - \epsilon < x_N \leq x_n \leq \alpha < \alpha + \epsilon$$

for $n > N$. (The second to last inequality holds as α is by definition an upper bound of A). So, for all $n > N$ we have

$$\begin{aligned} \alpha - \epsilon < x_n < \alpha + \epsilon \\ -\epsilon < x_n - \alpha < \epsilon \\ |x_n - \alpha| < \epsilon \end{aligned}$$

As this can be done for any choice of $\epsilon > 0$ we have that $\{x_n\} \to \alpha$, and thus the sequence converges. □

As we saw in the proof of the theorem, for a sequence $\{x_n\}$, if the sequence is bounded above and increasing then we know it converges to the supremum of the set of terms in the sequence. And similarly if the sequence is bounded below and decreasing it converges to the infinimum of the set of terms in the sequence. Once again, this gives some perspective of how sup/inf generalize max/min as the max/min of a increasing/decreasing sequence may not exist, but we can say the sequence approaches the 'celing'/'floor' of the collection of terms and in a sense is tending towards a largest/smallest value.

This theorem is valuable as it gives a characterization of convergence using only two simple properties to check: boundedness and monotonicity. It is also an existential result, as it tells us a sequence converges if it meets those two criteria, but it does not tell us what the sequence converges to exactly.[43] With that said, it is always good before setting off for a destination to make sure it exists first, so this theorem will come of use when we want to know there is an answer in some cases before looking for one.[44]

Specifically, consider example a). above. It is clear that this sequence is increasing and bounded above by 1. Intuition or knowledge from Calculus 1 makes it clear what this sequence will converge to. So, here's an example that is less obvious.

Lecture 14 - 10/30/24

Example 16. *Show the sequence* $\{x_n\} = \{(1+\frac{1}{n})^n\}$ *converges.*

To do this we will show that the sequence is bounded above and monotonically increasing. The argument for boundedness and monotonicty will make use of the binomial theorem. The proof of the binomial theorem can be found in the appendix on induction for those curious.

Boundedness

Here will will show that each term of the sequence is indeed bounded above. So, let $x_n = (1+\frac{1}{n})^n$. *By the binomial theorem, this equals*

$$x_n = \left(1+\frac{1}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{n}\right)^k 1^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \frac{1}{n^k}$$

Now, let us look at the term inside of the sum for an arbitrary k, *i.e. the term*

$$\binom{n}{k} \frac{1}{n^k}$$

First of this can be written as

$$\binom{n}{k} \frac{1}{n^k} = \frac{n!}{k!(n-k)!} \frac{1}{n^k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!n^k}$$

by canceling $(n-k)!$ *from the numerator. Now the numerator has* k *terms in it, making use of* n^k *being* n *written* k *times we have*

$$\binom{n}{k} \frac{1}{n^k} = \frac{1}{k!}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)\cdots\left(\frac{n-k+1}{n}\right) < \frac{1}{k!}$$

[43] we know it converges to the sup or inf of the terms of the sequence, but these are symbolic answers and may not give any idea of what the actual value the sequence is converging to

[44] as these two properties are sometime easier to check for a sequence than it being Cauchy

and the inequality holds because every fraction in parenthesis above is less than 1, so their product is as well. And so this shows that

$$x_n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \frac{1}{n^k} < \sum_{k=0}^{n} \frac{1}{k!}$$

Making use of $0! = 1! = 1$*, we have that*

$$x_n < 2 + \sum_{k=2}^{n} \frac{1}{k!}$$

Now for $k \geq 2$ *we have that*

$$k! = k(k-1)(k-2)\cdots 3 \cdot 2 \cdot 1 > k(k-1)$$

and this implies that

$$\frac{1}{k!} < \frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}$$

And so we have that

$$x_n < 2 + \sum_{k=2}^{n} \frac{1}{k-1} - \frac{1}{k}$$

And this sum is a telescoping sum (i.e. many of the terms cancel out as you can see)

$$x_n < 2 + \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots \left(\frac{1}{n-1} - \frac{1}{n}\right)$$
$$x_n < 3 - \frac{1}{n} < 3$$

Thus every term in the sequence is bounded by 3, so the sequence is bounded.

Monotonicity

Here we will show that the sequence $\{x_n = (1 + \frac{1}{n})^n\}$ *is monotonically increasing.*

For this, let us start with $x_{n+1} = (1 + \frac{1}{n+1})^{n+1}$*. Much like the last portion, this will be*

$$x_{n+1} = \sum_{k=0}^{n+1} \left(\begin{array}{c} n+1 \\ k \end{array} \right) \frac{1}{(n+1)^k}$$

If we look at individual terms in the sum and follow a similar simplification process to what we did in the section above we find

$$\left(\begin{array}{c} n+1 \\ k \end{array} \right) \frac{1}{(n+1)^k} = \frac{1}{k!} \left(\frac{n+1}{n+1}\right) \left(\frac{n}{n+1}\right) \left(\frac{n-1}{n+1}\right) \cdots \left(\frac{n-k+2}{n+1}\right)$$

And we have k *terms in large sets of parenthesis above, each of the form*

$$\frac{n+1-j}{n+1}$$

for $0 \leq j \leq k-1$*. And the following*

$$\frac{n+1-j}{n+1} \geq \frac{n-j}{n}$$

is true for all $0 \leq j \leq k-1$. *But this means that*

$$\begin{aligned}\binom{n+1}{k}\frac{1}{(n+1)^k} &= \frac{1}{k!}\left(\frac{n+1}{n+1}\right)\left(\frac{n}{n+1}\right)\left(\frac{n-1}{n+1}\right)\cdots\left(\frac{n-k+2}{n+1}\right) \\ &\geq \frac{1}{k!}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)\cdots\left(\frac{n-k+1}{n}\right) = \binom{n}{k}\frac{1}{n^k}\end{aligned}$$

And so, we have that

$$x_{n+1} = \sum_{k=0}^{n+1}\binom{n+1}{k}\frac{1}{(n+1)^k} \geq \sum_{k=0}^{n}\binom{n}{k}\frac{1}{n^k} = x_n$$

Thus the sequence $\{x_n\}$ *is monotonically increasing.*

Now, by our theorem we know that the sequence $\{x_n\}$ *converges. Later in the course we will show that*

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

i.e. this sequence converges to the transcendental number e.

Example 17. *Determine the convergence of the following sequence* $\{x_n\}$ *defined recursively by*

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right), \qquad x_1 > \sqrt{a}, \quad a > 0$$

We will show that this sequence is bounded from below and monotonically decreasing.

Let us see how it is bounded below at first

$$\begin{aligned}x_{n+1}^2 - a &= \left[\frac{1}{2}\left(x_n + \frac{a}{x_n}\right)\right]^2 - a \\ &= \frac{1}{4}\left(x_n^2 + 2a + \frac{a^2}{x_n^2}\right) - a \\ &= \frac{1}{4}\left(x_n^2 + 2a + \frac{a^2}{x_n^2} - 4a\right) \\ &= \frac{1}{4}\left(x_n^2 - 2a + \frac{a^2}{x_n^2}\right) \\ &= \left[\frac{1}{2}\left(x_n - \frac{a}{x_n}\right)\right]^2 \geq 0\end{aligned}$$

And this shows that $x_{n+1} \geq \sqrt{a}$ *for all* $n \in \mathbb{N}$. *Thus* $\{x_n\}$ *is bounded from below.*

Next let us look at

$$\begin{aligned}x_n - x_{n+1} &= x_n - \frac{1}{2}\left(x_n + \frac{a}{x_n}\right) \\ &= \frac{x_n}{2} - \frac{a}{2x_n} \\ &= \frac{x_n^2 - a}{2x_n} \geq 0\end{aligned}$$

as the numerator is positive due to the lower bound found earlier. Thus we have $x_n \geq x_{n+1}$ for all $n \in \mathbb{N}$. So, the sequence is monotonically decreasing.

Thus, by the monotone convergence theorem, we know that $\{x_n\}$ converges to some value L.

To find L, we exploit the recursive relation

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$$

Taking limits of both sides we have

$$L = \frac{1}{2}\left(L + \frac{a}{L}\right)$$

and this implies that

$$\begin{aligned} \frac{L}{2} &= \frac{a}{2L} \\ L^2 &= a \\ L &= \pm\sqrt{a} \end{aligned}$$

As the terms in the sequence are positive, we have that $L = \sqrt{a}$. Thus we have shown that the sequence defined by

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right), \qquad x_1 > 0, \quad a > 0$$

converges to $\sqrt{a}$.

Our previous theorem showed that a bounded monotone sequence converges, in particular, in the proof we specifically saw that a monotonically increasing sequence that is bounded above converges to the supremum of the set of terms in the sequence. Here we provide proof of a remark from the previous section that said any set that is bounded above has a sequence of terms within it that approximate the supremum. (with a similar statement for the infimum of a set that is bounded below)

Theorem 40. *Let A be a non-empty subset of $\mathbb{R}$ that is bounded above, then there exists a sequence $\{x_n\}$ with $x_n \in A$ for all $n \in \mathbb{N}$ with $\{x_n\} \to \sup A$.*

Do note that the proof of this theorem will require the Axiom of Choice.

Proof. First, let $\alpha = \sup A$, which we know exists as A is bounded above and the completeness of $\mathbb{R}$. Let us recall another theorem from the prior section. For α we know that for any choice of $\epsilon > 0$, the existence of an $x \in A$ such that $x > \alpha - \epsilon$.

So, let $\epsilon = 1$, by the above, there then exists an $x_1 \in A$ with the property that $x_1 > \alpha - 1$, and we also have

$$\alpha - 1 < x_1 < \alpha < \alpha + 1$$

Next, as our theorem holds for any choice of ϵ, let $\epsilon = \frac{1}{2}$, by the above, there then exists an $x_2 \in A$ with the property that $x_2 > \alpha - \frac{1}{2}$, and we also have

$$\alpha - \frac{1}{2} < x_2 < \alpha < \alpha + \frac{1}{2}$$

And again, as our theorem holds for any choice of ϵ, let $\epsilon = \frac{1}{3}$, by the above, there then exists an $x_3 \in A$ with the property that $x_3 > \alpha - \frac{1}{3}$, and we also have

$$\alpha - \frac{1}{3} < x_3 < \alpha < \alpha + \frac{1}{3}$$

And hopefully it is clear that we can keep going with this process, i.e. for each $n \in \mathbb{N}$ there is an $x_n \in A$ with

$$\alpha - \frac{1}{n} < x_n < \alpha < \alpha + \frac{1}{n}.$$

And this shows that

$$\begin{aligned} \alpha - \frac{1}{n} < x_n < \alpha + \frac{1}{n} \\ -\frac{1}{n} < x_n - \alpha < \frac{1}{n} \\ |x_n - \alpha| < \frac{1}{n}. \end{aligned}$$

And then, the squeeze theorem states that if $0 \leq a_n \leq b_n$ holds for all $n \in \mathbb{N}$ and $\{b_n\} \to 0$, then $\{a_n\} \to 0$.

As $\{\frac{1}{n}\} \to 0$, what we've shown above then says that $\{x_n - \alpha\} \to 0$, thus $\{x_n\} \to \alpha = \sup A$. □

Exercises for section 4.2:

1. Provide a proof of the Monotone Convergence Theorem 39 for a decreasing sequence that is bounded below.
2. Proof the analogous version of theorem 40 where A is a non-empty subset of the reals that is bounded below.
3. Define a sequence $\{a_n\}$ by $a_1 = \sqrt{2}$ and $a_{n+1} = \sqrt{2 + a_n}$.
   a). Show that $a_n \leq 2$ for every n.
   b). Show that $\{a_n\}$ is an increasing sequence. And then explain why $\{a_n\}$ converges.
   c). Show that $\lim a_n = 2$.
4. Let $k > 1$ be a constant, and define a sequence $\{a_n\}$ by $a_1 = 1$ and
$$a_{n+1} = \frac{k(1 + a_n)}{k + a_n}$$
   a). Show that $\{a_n\}$ converges. (Either show the sequence is Cauchy or satisfies the conditions in the monotone convergence theorem)
   b). Find $\lim a_n$.
5. From [R], fix $\alpha > 1$ and take $x_1 > \sqrt{\alpha}$ and define
$$x_{n+1} = \frac{\alpha + x_n}{1 + x_n}.$$

a). Prove that $x_1 > x_3 > x_5 > \cdots$.

b). Prove that $x_2 < x_4 < x_6 < \cdots$.

c). Find what $\{x_n\}$ converges to and prove your claim.

6. For $x_1 > 0$ and $a > 0$ define the sequence $\{x_n\}$ recursively by

$$x_{n+1} = \frac{p-1}{p} x_n + \frac{\alpha}{p x_n^{p-1}}$$

where p is a fixed positive integer. Prove what $\{x_n\}$ converges to. *Hint: Monotone Convergence theorem and the arithmetic mean vs geometric mean inequality will help here*

4.3 The Bolzano-Weierstrass Theorem

In this last subsection in our reprise chapter on Sequences, I want to present a companion piece to the Monotone Convergence Theorem. In that section, we saw that the additional condition of monotonicity for a sequence gives boundedness of a sequence the strength to guarantee the convergence of the sequence. However, the point of this section is that even without monotonicity the boundedness of a sequence does give some convergent information, in that it does guarantee the existence of at least one convergent subsequence. This is precisely the content of the Bolzano-Weierstrass theorem.

This section also exists as a small glimpse at a future section on **compactness** when we speak more about the topology of $\mathbb{R}$. In that section we will see the Bolzano-Weierstrass theorem again from the viewpoint of sets.

To start, let us recall a definition from our first section on sequences

Definition 31. *For a sequence $\{x_n\}$, a number L is a* **subsequential limit** *of $\{x_n\}$ if for every $\epsilon > 0$ and every $N \in \mathbb{N}$ there is some $n > N$ with*

$$|x_n - L| < \epsilon$$

This lines up with the idea that a sequence is 'frequently' in the error window of a subsequential limit. At any index in the sequence N, there is some index past it $n > N$ with $|x_n - L| < \epsilon$. (But this does not mean a 'tail' is in the error window!) Also note that due to theorem 17 this definition can be equivalently phrased as saying L is a subsequential limit of $\{x_n\}$ if for any $\epsilon > 0$ we have that the interval $(L - \epsilon, L + \epsilon)$ contains an infinite number of terms from the sequence.

Theorem 41. (Bolzano-Weierstrass)*: A bounded sequence $\{x_n\}$ has at least one convergent subsequence.*

Proof. Define the set A to be the set of terms in the sequence $\{x_n\}$, i.e.

$$A = \{x_1, x_2, x_3, \dots\}$$

As we do not allow repeated elements within sets, there is two possible cases. Either A is finite or A is infinite.

Case 1: A is finite. If A is finite, then there exists an $N \in \mathbb{N}$ and real numbers $y_1, y_2, \dots, y_N$ such that

$$A = \{y_1, y_2, \dots, y_N\}$$

From this we can define the following subsets of the natural numbers.

$$B_k = \{n \in \mathbb{N} \mid x_n = y_k\}$$

i.e. each B_k just collects the indices from all terms in the sequence $\{x_n\}$ that equal y_k. Well then,

$$B_1 \cup B_2 \cup \cdots \cup B_N = \mathbb{N}$$

and so it must be that one of the B_k, $1 \leq k \leq N$ is an infinite set. Say B_j is an infinite set, then there is an infinite number of terms from $\{x_n\}$ that equal y_j, thus for any $\epsilon > 0$ the interval $(y_j - \epsilon, y_j + \epsilon)$ contains an infinite number of terms from $\{x_n\}$, and thus y_j is a subsequential limit of $\{x_n\}$ and the theorem is proven in this case.

Case 2: A is infinite, i.e. the sequence $\{x_n\}$ contains an infinite number of distinct terms. As $\{x_n\}$ is a bounded sequence, we have that A is a counded set. As A is bounded, there exists a positive real number M such that

$$|a| \leq M, \qquad \forall a \in A$$

Put another way, the set A fits within the interval $[-M, M]$.

If we cut the interval $[-M, M]$ into two equal pieces, we have the intervals $[-M, 0]$ and $[0, M]$, both of length M. As A is an infinite set, at least one of these intervals $[-M, 0]$ or $[0, M]$ contains an infinite number of points from A. Pick one of these that contains an infinite number of terms and call this A_1.

Now bisect A_1 into two equal intervals of length $\frac{M}{2}$. Similar to above, one of these intervals (or both) contains an infinite number of points from A, pick one of these intervals and call this A_2. Note that $A_2 \subseteq A_1$.

Proceed onward in this same manner, bisecting A_2 to find A_3 and so on and so on. What we will find over all is a definition of A_n for any natural number n, with the length of A_n given as $\frac{M}{2^{n-1}}$, and the property that

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq A_4 \cdots$$

i.e. the sets are all nested (each A_n contains the next A_{n+1}). We now come to a lemma

Lemma 42. *If $I_n = [a_n, b_n]$ is a sequence of nested intervals $I_1 \supseteq I_2 \supseteq I_3 \cdots$ whose lengths converge to 0, then $\bigcap_{n=1}^{\infty} I_n$ consists of exactly one point.*[45]

Proof. As all intervals are contained within the first interval I_1, we have that the left endpoints of each interval

$$a_n \leq b_1 \qquad \forall n \in \mathbb{N}$$

i.e. b_1 is an upper bound for the sequence $\{a_n\}$. By the way each interval is nested, we also have

$$a_1 \leq a_2 \leq a_3 \leq \cdots$$

thus the sequence $\{a_n\}$ is bounded and monotonic. So, by the bounded monotone convergence theorem, we know the sequence $\{a_n\}$ converges. Let us say $\{a_n\}$ converges to a.

This same argument can be applied to see that the right endpoints form a sequence $\{b_n\}$ that is bounded below and monotonically decreasing, thus by the bounded monotone convergence theorem this sequence also converges, let us say to a value b. So, the sequence $\{b_n\}$ converges to b.

[45] this is a specialized version of the Nested Interval Property, which will be seen in the later section on compactness

As each interval I_n is of the form $I_n = [a_n, b_n]$, the length of I_n is $b_n - a_n$. By assumption, the lengths of these intervals tend to 0, i.e. $\lim_{n\to\infty}(b_n - a_n) = 0$, so

$$\begin{aligned} 0 &= \lim_{n\to\infty}(b_n - a_n) \\ &= \lim_{n\to\infty} b_n - \lim_{n\to\infty} a_n \\ &= b - a \end{aligned}$$

where in the second line we used our algebraic properties of limits. Thus we see that $a = b$, i.e. the left endpoint sequence $\{a_n\}$ and the right endpoint sequence $\{b_n\}$ converge to the same place (just from different directions) As the sequences are monotonic, we have that

$$a_n \leq a, \qquad b \leq b_n, \qquad \forall n \in \mathbb{N}$$

and as $a = b$, this implies that $a \in I_n$ for all $n \in \mathbb{N}$. Thus $\bigcap_{n=1}^{\infty} I_n$ is not empty and contains at least one point. We need to justify now why it only contains one point.

So, let c be some real number distinct from a, i.e. $c \neq a$. Without loss of generality, assume that $c > a$, and thus the distance from a to c is $c - a > 0$. As $\{b_n\}$ converges to a (from above), using $\epsilon = \frac{c-a}{2}$, there is a $N \in \mathbb{N}$ such that for all $n > N$

$$|b_n - a| < \frac{c-a}{2}$$

As $\{b_n\}$ approaches a from above, we have that $|b_n - a| = b_n - a$. So then, for $n > N$,

$$\begin{aligned} b_n - a &< \frac{c-a}{2} \\ b_n &< \frac{c-a}{2} + a = \frac{c+a}{2} < c \end{aligned}$$

as $\frac{a+c}{2}$ is the midpoint between a and c. The key is that $b_n < c$ for $n > N$. But this then means that $c \notin I_n$ for $n > N$, so

$$c \notin \bigcap_{n=1}^{\infty} I_n.$$

Thus, because of this we finally have that

$$\bigcap_{n=1}^{\infty} I_n = \{a\}.$$

□

Back to our proof of Bolzano-Weierstrass: We have a definition of A_n for any natural number n, with the length of A_n given as $\frac{M}{2^{n-1}}$, and the property that

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq A_4 \cdots$$

i.e. the sets are all nested (each A_n contains the next A_{n+1}). And every A_n is of the form $[a_n, b_n]$ for real numbers a_n, b_n. As

$$\lim_{n\to\infty} \frac{M}{2^{n-1}} = 0,$$

the lengths of the A_n go to 0.

Thus we can apply our lemma, and so we know there is a point p with the property that

$$\bigcap_{n=1}^{\infty} A_n = \{p\}$$

Now let $\epsilon > 0$ be arbitrary, as the sequence of lengths $\{\frac{M}{2^{n-1}}\}$ goes to zero, for this ϵ there is some $N \in \mathbb{N}$ such that for all $n > N$,

$$|\frac{M}{2^{n-1}} - 0| < \epsilon$$

As this holds for all $n > N$, let us take $n = N + 1$, and the above implies that $\frac{M}{2^N} < \epsilon$, and recall that $\frac{M}{2^N}$ is the length of A_{N+1}. Well, as $p \in A_{N+1}$, this means that

$$p \in A_{N+1} \subseteq (p - \epsilon, p + \epsilon)$$

and by definition A_N contains an infinite number of points from A. As $\epsilon > 0$ was arbitrary, we have shown by theorem 17 that p is a subsequential limit point of $\{x_n\}$. □

Once again, the Bolzano-Weierstrass theorem will come up again in a later section on compactness and at that time it will say that a bounded closed set contains at least one limit point.[46] This is a direct consequence of closed and bounded sets being **compact**, which will be a topic for a later section.

Even though we are not discussing compactness as of yet, there is a common phrase of intuition that typically follows a study of compactness and that is that 'compactness generalizes finiteness', and I'd like to give some context to this notion every time that I can.

You likely recall the pidgeonhole principle from a prior math class, it simply states that if you have a collection of k boxes that you will place n objects into, then you know that at least one box must contain $\lceil \frac{n}{k} \rceil$ many elements. As an example if there are 3 boxes and 10 objects to be placed in these boxes then you know at least one box contains 4 elements. Well, what if you had n boxes and an infinite number of elements. This is precisely the situation of case 1 in the proof above, and we saw that at least one box contains an infinite number of elements. Put another way, when the values the terms in a sequence can take on is finite, then there must be at least one constant subsequence.

Bolzano-Weierstrass generalizes this idea with a bounded sequence that contains an infinite collection of distinct terms. As there are an infinite number of 'boxes' and 'objects' in this case, we can no longer deduce that at least one box has an infinite number of elements, but we can deduce that there is a convergent subsequence. In this analogy, we can no longer say at least one box contains an infinite number of elements, but we can say that there is one box that has an infinite number of objects arbitrarily close to it.

4.4 Limsup & Liminf (Optional)

limit sets of sequences

limit point of sequence set if subsequential limit (we've done this)

example of rational sequence where limit set is R

[46] a limit point of a set is a point where elements from the set cluster infinitely, we will see a formal definition later

5 Series

5.1 Introduction & Definitions

Definition 32. *For a sequence $\{a_n\}$ we can define a new sequence by*

$$\begin{aligned} s_1 &= a_1 \\ s_2 &= a_1 + a_2 \\ &\vdots \\ s_k &= \sum_{j=1}^{k} a_j = a_1 + a_2 + \cdots + a_k \end{aligned}$$

this is called the **sequence of partial sums** *of $\{a_n\}$. If the sequence of partial sums $\{s_k\}$ converges to a real number L, then we define the* **series** *of $\{a_n\}$ as*

$$\lim_{k\to\infty} s_k = \sum_{k=1}^{\infty} a_k = L.$$

We say this series diverges if the sequence of partial sums diverges. We say the series diverges to ∞ or $-\infty$ is the sequence of partial sums does so. When talking about the series of $\{a_n\}$ as shorthand the notation $\sum a_n$ and $\sum_n a_n$ is often used.

As $\mathbb{R}$ is Cauchy complete and thus every sequence is convergent if and only if it is Cauchy, we have that a series converges if the sequence of partial sums is Cauchy, i.e. if $\forall \epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $m, n > N$ we have

$$|s_n - s_m| < \epsilon, \quad \forall m, n > N.$$

Without loss of generality (WLOG) if we take $n > m > N$, then the condition above turns into

$$\begin{aligned} |s_n - s_m| &< \epsilon \\ \left| \sum_{k=1}^{n} a_k - \sum_{k=1}^{m} a_k \right| &< \epsilon \\ \left| \sum_{k=m+1}^{n} a_k \right| &< \epsilon \end{aligned}$$

And so this leads to the following definition

Definition 33. (Cauchy Condition for Series) *For a sequence $\{a_n\}$, the sequence of partial sums $\{s_n\}$ converges if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n, m > N$,*

$$\left| \sum_{k=m+1}^{n} a_k \right| < \epsilon, \qquad \forall n > m > N.$$

As we saw with sequences, changing a finite number of terms from s sequence $\{x_n\}$ does not effect its convergent behavior as this is dependent upon infinite tails of the sequence. Similarly,

changing out or replacing a finite number of terms from a series will not effect it's convergent behavior.

What follows is a very useful theorem. It gives a necessary condition on the convergence of a series. When working with a new series and trying to determine its convergence behavior, this theorem is the first test that should be applied.

Theorem 43. *If $\sum a_n$ converges, then the sequence of the terms $\{a_n\}$ must converge to 0, i.e. $\lim_{n\to\infty} a_n = 0$.*

Proof. Let $s_n = \sum_{k=1}^{n} a_n$ denote the sequence of partial sums of a_n. As $\sum a_n$ converges, we know that the sequence $\{s_n\}$ is convergent, and therefore Cauchy. Thus, by the condition above for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n, m > N$,

$$\left| \sum_{k=m+1}^{n} a_k \right| < \epsilon, \qquad \forall n > m > N.$$

If we take $n = N + 2$ and $m = N + 1$, then this implies

$$|a_{N+2}| = \left| \sum_{k=N+1+1}^{N+2} a_k \right| < \epsilon$$

In fact taking $n = N + k + 1$ and $m = N + k$, this Cauchy Condition gives

$$|a_{N+k+1}| < \epsilon,$$

for all $k \in \mathbb{N}$. As this can be done for every $\epsilon > 0$, this shows that $\{a_n\}$ converges to 0, i.e. $\lim a_n = 0$. □

As mentioned in the comment before the theorem, this result is incredibly useful for checking the convergence of a series. The contrapositive of the theorem states that if $\lim_{n\to\infty} a_n \neq 0$, then the series $\sum a_n$ must diverge. However do be careful, the converse of the theorem is not true, i.e. the terms $\{a_n\}$ decaying to 0 as $n \to \infty$ is not sufficient enough to guarantee the sum converges.

Example 18. *The harmonic series is given by*

$$H = \sum_{n=1}^{\infty} \frac{1}{n}$$

i.e. the sum of all reciprocals of natural numbers. If we assume that the harmonic series actually converges to a value that we will call H as well we can see

$$\begin{aligned} H &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \cdots \\ H &> 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{6} + \frac{1}{6} + \cdots \\ H &> \frac{1}{2} + 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots \\ H &> \frac{1}{2} + H \end{aligned}$$

and this leaves us with a pretty clear contradiction that $\frac{1}{2} < 0$. Thus it must be that H diverges and thus diverges to ∞. This shows that the converse to the previous theorem is not true as $\lim_{n\to\infty} \frac{1}{n} = 0$ but H diverges.

Definition 34. *A* **geometric series** *is a series of the specific form*

$$a + ar + ar^2 + ar^3 + \cdots = \sum_{n=0}^{\infty} ar^n$$

Theorem 44. *Given a geometric series*

$$a + ar + ar^2 + ar^3 + \cdots = \sum_{n=0}^{\infty} ar^n$$

then we have the following

$$\sum_{n=0}^{\infty} ar^n = \begin{cases} 0, & \text{if } a = 0 \\ \frac{a}{1-r}, & \text{if } a \neq 0,\ |r| < 1 \\ \text{diverges}, & \text{if } a \neq 0, |r| \geq 1 \end{cases}$$

Proof. Let $s_n = \sum_{k=0}^{n} ar^k = a\sum_{k=0}^{n} r^k$ be the sequence of partial sums of $\{a_n\}$. If $a = 0$, then $s_n = 0$ for all $n \in \mathbb{N}$, and thus the series converges to 0.

Now assume that $a \neq 0$, $|r| < 1$, and without loss of generality assume that $a > 0$. (This will not matter much in the proof and the proof when $a < 0$ is very similar). We make use of the following

$$s_n = a\sum_{k=0}^{n} r^k = a + ar + ar^2 + \cdots + ar^n$$
$$rs_n = ar\sum_{k=0}^{n} r^k = ar + ar^2 + ar^3 + \cdots + ar^{n+1}$$

Then

$$s_n - rs_n = a - ar^{n+1},$$

and this implies that

$$s_n = \frac{a(1 - r^{n+1})}{1 - r}$$

What this then tells us is that

$$\left| s_n - \frac{a}{1-r} \right| = \frac{|-ar^{n+1}|}{|1-r|} = \frac{a}{1-r} \cdot |r|^{n+1}$$

as $a > 0$ and $|r| < 1$. Thus we see that $\{s_n\}$ will converge to $\frac{a}{1-r}$, precisely when $|r|^{n+1}$ converges to 0, and this occurs when $|r| < 1$.

On the other hand, when $|r| \geq 1$, then nth term in the series $a_n = ar^n$. But then $\lim a_n \neq 0$, and so by the contrapositive of the prior theorem, we know that the series in this case diverges. □

As we have seen, when $|r| < 1$ we have the following formulas that come up often

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}, \qquad \sum_{n=1}^{\infty} ar^n = \frac{ar}{1-r}$$

Example 19. *The geometric series is very useful when trying to find closed forms of decimal expansions that are periodic in nature.*

a).

$$44.4444\cdots = 40 + 4 + 0.4 + 0.04 + \cdots =$$

This is equivalent to

$$40\left(1 + \frac{1}{10} + \frac{1}{10^2} + \cdots\right)$$

And so we see this is a geometric series with $a = 40$ *and* $r = \frac{1}{10}$*, thus this converges to*

$$\frac{40}{1 - \frac{1}{10}} = \frac{40}{\frac{9}{10}} = \frac{400}{9}.$$

b).

$$1 - 4 + 16 - 64 + \cdots$$

This is a geometric series with $a = 1$ *and* $r = -4$*. As* $|r| = |-4| = 4 \geq 1$*, this geometric series does not converge.*

Exercises for section 5.1:

1. From [FM], pick a decimal expansion of a real number and suppose that the sequence of decimals is *periodic* in the sense that it takes the repeating form

   $$x = 0.a_1\ a_2\ a_3\ \cdots\ a_n\ a_1\ a_2\ a_3\ \cdots\ a_n\ \cdots$$

   Show that x must be rational.

5.2 Series with nonnegative terms

Instead of jumping directly to the study of fully abstract series $\sum a_n$ with $a_n \in \mathbb{R}$ we will first collect a number of results specifically about series made up of nonnegative terms, i.e. $\sum a_n$ with $a_n \geq 0$ for all $n \in \mathbb{N}$.

Theorem 45. *Let* $\sum a_n$ *be a series of nonnegative terms, i.e.* $a_k \geq 0$ *for all* $k \in \mathbb{N}$*, then the series converges if and only if the partial sums are a bounded sequence.*

Proof. $\implies$ Let $\{s_n\}$ be the sequence of partial sums for the series $\sum a_n$. If we are assuming that the series converges, then there exists some $L \in \mathbb{R}$ such that

$$\sum_{k=1}^{\infty} a_k = \lim_{n\to\infty} s_n = L$$

Thus the sequence of partial sums are a convergent sequence and hence bounded as convergent sequences are bounded.

$\impliedby$ Assume that the sequence of partial sums $\{s_n\}$ are bounded, i.e. bounded above as the series is made up of nonnegative terms. Thus there exists an $M > 0$ such that

$$s_n \leq M, \qquad \forall n \in \mathbb{N}$$

and similarly

$$s_{n+1} = \sum_{k=1}^{n+1} a_k = s_n + a_{n+1} \geq s_n$$

as the terms a_n are nonnegative. Thus $\{s_n\}$ is a monotonically increasing sequence. Thus, by the monotone convergence theorem as $\{s_n\}$ is monotonically increasing and bounded above, $\{s_n\}$ is a convergent sequence and thus the series converges. □

Before we move on, as a quick corollary to the theorem above we have

Corollary 46. *Let $\sum a_n$ be a series of nonnegative terms. Then $\sum a_n$ either converges or diverges to ∞.*

The next result is often called the 'sparseness relation' for series as it makes equivalent the convergence of one series with another that is made up with certain multiples of sparse terms within the original series. This theorem will immediately have use in our analysis of p-series and is another result that be used to quickly show certain types of series converge or diverge with little effort.

Theorem 47. (Sparseness Relation) *Let $\{a_n\}$ be a monotonically decreasing sequence of nonnegative terms, then $\sum a_n$ converges if and only if*

$$\sum_{k=0}^{\infty} 2^k a_{2^k}$$

converges.

Proof. Let s_n denote the partial sum of the original series and t_k denotes the partial sum of the 'sparse' series, so

$$\begin{aligned} s_n &= a_1 + a_2 + a_3 + \cdots + a_n \\ t_k &= a_1 + 2a_2 + 4a_4 + \cdots + 2^k a_{2^k} \end{aligned}$$

Then if $n < 2^k$ we have that

$$\begin{aligned} s_n &= a_1 + a_2 + a_3 + \cdots + a_n \\ &\leq a_1 + a_2 + a_3 + \cdots + a_n + \cdots a_{2^k} + \cdots + a_{2^{k+1}-1} \\ &\leq a_1 + a_2 + a_2 + a_4 + a_4 + a_4 + a_4 + \cdots + \underbrace{a_{2^k} + \cdots + a_{2^k}}_{2^k \text{ times}} \\ &= a_1 + 2a_2 + 4a_4 + \cdots + 2^k a_{2^k} = t_k \end{aligned}$$

And so we see that $s_n \leq t_k$ for $n < 2^k$. For any n there is a k in which this condition will hold. Thus if

$$\sum_{k=1}^{\infty} 2^k a_{2^k} \text{ converges,}$$

then the sequence $\{t_k\}$ is bounded, which implies that $\{s_n\}$ is bounded and thus the original series $\sum a_n$ converges by the previous theorem.

Similarly, for $n > 2^k$ we have

$$\begin{aligned} s_n &= a_1 + a_2 + \cdots + a_n \\ &\geq a_1 + a_2 + \cdots + a_{2^k} \\ &\geq \frac{1}{2}a_1 + a_2 + a_4 + a_4 + \cdots + \underbrace{a_{2^k} + \cdots + a_{2^k}}_{2^{k-1} \text{ times}} \\ &= \frac{1}{2}a_1 + a_2 + 2a_4 + \cdots + 2^{k-1}a_{2^k} = \frac{1}{2}t_k \end{aligned}$$

So $s_n \geq \frac{1}{2}t_k$ for $n > 2^k$, and for any k we can find an n so this condition holds. Thus if the original series $\sum a_n$ converges, then $\{s_n\}$ is bounded and thus $\{t_k\}$ is a bounded sequence. Thus

$$\sum_{k=1}^{\infty} 2^k a_{2^k}$$

converges by the previous theorem. □

Theorem 48. (p-series)*: We have the following*

$$\sum_{n=1}^{\infty} \frac{1}{n^p} = \begin{cases} \textit{converges if } p > 1 \\ \textit{diverges if } p \leq 1 \end{cases}$$

Proof. The individual terms in the series are $a_n = \frac{1}{n^p}$, and so, if $p < 0$ then $p = -r$ for some $r > 0$ and then

$$a_n = \frac{1}{n^p} = n^r$$

and as $n \to \infty$ we have that $n^r \to \infty$. Thus as the terms a_n do not go to zero, it must be that $\sum \frac{1}{n^p}$ diverges in this case.

If $p = 0$, then $a_n = 1$ for all $n \in \mathbb{N}$ and similarly the terms a_n do not go to zero, thus the series diverges. When $p > 0$, we have $n^p < (n+1)^p$ and thus we are in the case of our previous theorem as $a_n = \frac{1}{n^p}$ forms a monotonically decreasing sequence. Thus the original series will converge precisely when

$$\sum_{k=0}^{\infty} 2^k \frac{1}{(2^k)^p} = \sum_{k=1}^{\infty} 2^{k(1-p)} = \sum_{k=0}^{\infty} \left(\frac{1}{2^{p-1}}\right)^k$$

does. And we can see that this series is a geometric series. Thus it will converge precisely when $\frac{1}{2^{p-1}} < 1$, which only happens when $p - 1 > 0$. □

Example 20. *Let us use the sparseness relation to analyze the convergence or divergence of the following series*

$$\sum_{n=2}^{\infty} \frac{1}{n \ln n}$$

Our terms are of the form $a_n = \frac{1}{n \ln n}$, and thus we have

$$2^n a_{2^n} = 2^n \cdot \frac{1}{2^n \ln 2^n} = \frac{1}{\ln 2^n} = \frac{1}{n \ln 2}$$

and thus

$$\sum_{k=1}^{\infty} 2^k a_{2^k} = \sum_{k=1}^{\infty} \frac{1}{k \ln 2} = \frac{1}{\ln 2} H$$

where H is the harmonic series. Thus the sparse series diverges and so by the sparseness relation we have that

$$\sum_{n=2}^{\infty} \frac{1}{n \ln n} \text{ diverges}$$

Exercises for section 5.2:

1. Give an example of a series $s_n = \sum_{k=0}^{n} a_k$ such that $\{s_n\}_n$ diverges and $\{s_n\}_n$ is bounded.
2. Let $\{\sum_{k=1}^{n} a_k\}_n$ be a series with non-negative terms.
   (a) If $\sum_{k=1}^{n} a_k$ converges, does this imply $\sum_{k=1}^{n} a_k^2$ converges ? Prove or disprove.
   (b) If $\sum_{k=1}^{n} a_k^2$ converges, does this imply $\sum_{k=1}^{n} a_k$ converges ? Prove or disprove.

5.3 Convergence Tests

In this section we collect a few theorems that are known as convergence tests. Besides the geometric series, it is very difficult to find the exact value that a series converges to.[47] Because of this we often will have to settle for an existential result of knowing a series converges and leaving its convergent value as some representative symbol. The theorems in this section will help you 'test' a series you are handed for its convergent nature.

Theorem 49. (The Comparison Theorem)

a). If there exists $N \in \mathbb{N}$ such that $|a_n| \leq c_n$ for all $n > N$ and $\sum c_n$ converges, then $\sum a_n$ converges.

b). If there exists $N \in \mathbb{N}$ such that $a_n \geq d_n \geq 0$ for $n > N$ and $\sum d_n$ diverges, then $\sum a_n$ diverges.

Proof. Let's begin with the proof of a). By making use of the general triangle inequality

$$\left| \sum_{k=m+1}^{n} a_k \right| \leq \sum_{k=m+1}^{n} |a_k| \leq \sum_{k=m+1}^{n} c_k$$

As the terms c_n are nonnegative, we have that

$$\sum_{k=m+1}^{n} c_k = \left| \sum_{k=m+1}^{n} c_k \right|$$

From the computations above, we see that if $\sum c_n$ satisfies the Cauchy condition, then so does $\sum a_n$. This suffices to prove part a).

For the proof of part b). from part a). we see that if $\sum a_n$ was convergent then we would have that $\sum d_n$ is convergent. Thus by the contrapositive, if $\sum d_n$ diverges, we have that $\sum a_n$ diverges. □

[47] usually involving power series or Fourier series methods

Theorem 50. (Ratio Test) *Let $\sum a_n$ be a series, and assume that $a_n = 0$ for only finitely many terms. Let*

$$L = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|$$

Then

- *If $L < 1$, then $\sum a_n$ converges.*
- *If $L > 1$, then $\sum a_n$ diverges.*
- *If $L = 1$, then the test is inconclusive.*

Proof. Suppose, at first, that $L < 1$, and take $\epsilon = \frac{1-L}{2} > 0$. Then as

$$\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = L$$

Then for this $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n > N$

$$\left|\left|\frac{a_{n+1}}{a_n}\right| - L\right| < \frac{1-L}{2}.$$

And by using the reverse triangle inequality, we find for $n > N$,

$$\left|\frac{a_{n+1}}{a_n}\right| - L = \left|\frac{a_{n+1}}{a_n}\right| - |L| \leq \left|\frac{a_{n+1}}{a_n} - L\right| < \frac{1-L}{2}.$$

So, for $n > N$

$$\frac{|a_{n+1}|}{|a_n|} < \frac{1+L}{2}.$$

Now, as $L < 1$, the midpoint between L and 1, $M = \frac{1+L}{2} < 1$, and we found for all $n > N$ that

$$\frac{|a_{n+1}|}{|a_n|} < M < 1.$$

If we set $n = N + 1$, we find

$$|a_{N+2}| < M|a_{N+1}|$$

and then

$$|a_{N+3}| < M|a_{N+2}| < M^2|a_{N+1}|$$

In general, for any $k \in \mathbb{N}$, we have

$$|a_{N+k}| < M^k|a_{N+1}|$$

It is because of this, that we can compare our series to a geometric series.

$$\begin{aligned}
\sum_{n=1}^{\infty} |a_n| &= \sum_{n=1}^{N} |a_n| + \sum_{n=N+1}^{\infty} |a_n| = \sum_{n=1}^{N} |a_n| + \sum_{k=0}^{\infty} |a_{N+1+k}| \\
&< \sum_{n=1}^{N} |a_n| + \sum_{k=0}^{\infty} M^k |a_{N+1}| = \sum_{n=1}^{N} |a_n| + |a_{N+1}| \sum_{k=0}^{\infty} M^k \\
&= \sum_{n=1}^{N} |a_n| + |a_{N+1}| \left[\frac{1}{1-M}\right] = \text{something finite}
\end{aligned}$$

Thus making use of the comparison test with $a_n \leq |a_n|$, we have that $\sum a_n$ converges.

If $L > 1$, take $\epsilon = \frac{L-1}{2} > 0$. Similar to the process above using the reverse triangle inequality, we find a $N \in \mathbb{N}$ with the property that for $n > N$ that

$$\frac{|a_{n+1}|}{|a_n|} > \frac{L+1}{2} > 1.$$

But this implies that

$$|a_N| < |a_{N+1}| < |a_{N+2}| < \cdots$$

And this shows that $\lim_{n\to\infty} a_n \neq 0$, and thus by the contrapositive of theorem 43, we have that $\sum a_n$ diverges. □

Theorem 51. (Root Test) *Let $\sum a_n$ be a series and let*

$$L = \lim_{n\to\infty} \sqrt[n]{|a_n|}$$

Then

- *If $L < 1$, then $\sum a_n$ converges.*
- *If $L > 1$, then $\sum a_n$ diverges.*
- *If $L = 1$, then the test is inconclusive.*

This proof is very similar to the proof of the ratio test.

Proof. Suppose, at first, that $L < 1$, and take $\epsilon = \frac{1-L}{2} > 0$. Then as

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} = L$$

Then for this $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n > N$

$$\left| \sqrt[n]{|a_n|} - L \right| < \frac{1-L}{2}.$$

And by using the reverse triangle inequality, we find for $n > N$,

$$\sqrt[n]{|a_n|} - L = \left| \sqrt[n]{|a_n|} \right| - |L| \leq \left| \sqrt[n]{|a_n|} - L \right| < \frac{1-L}{2}.$$

So, for $n > N$

$$\sqrt[n]{|a_n|} < \frac{1+L}{2}.$$

Now, as $L < 1$, the midpoint between L and 1, $M = \frac{1+L}{2} < 1$, and we found for all $n > N$ that

$$\sqrt[n]{|a_n|} < M$$

If we set $n = N + 1$, we find

$$|a_{N+1}| < M^{N+1}$$

and generally that

$$|a_k| < M^k$$

for all $k > N$. As $M < 1$, we may follow the proof of the ratio test, and compare $\sum a_n$ to a geometric series that converges, thus $\sum a_n$ converges.

If $L > 1$, take $\epsilon = \frac{L-1}{2} > 0$. Similar to the process above using the reverse triangle inequality, we find a $N \in \mathbb{N}$ with the property that for $n > N$ that

$$\sqrt[n]{|a_n|} > \frac{L+1}{2} > 1.$$

But this implies that

$$|a_k| > 1$$

for $k > N$ and this shows that $\lim_{n\to\infty} a_n \neq 0$, and thus by the contrapositive of theorem 43, we have that $\sum a_n$ diverges. □

Example 21. *Determine the convergence or divergence of the following series.*

a).

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

Just directly we see

$$a_n = \frac{n \cdot n \cdot n \cdots n}{n \cdot (n-1) \cdot (n-2) \cdots 1} = \left(\frac{n}{n}\right)\left(\frac{n}{n-1}\right) \cdots \left(\frac{n}{2}\right) \cdot \left(\frac{n}{1}\right)$$

Thus a_n is the product of n terms that are each greater or equal to 1, so $\lim_{n\to\infty} a_n \neq 0$. And thus the series diverges.

b).

$$\sum_{n=1}^{\infty} \frac{2^n}{n!}$$

Use the Ratio Test

$$\frac{a_{n+1}}{a_n} = \frac{2^{n+1}}{(n+1)!} \cdot \frac{n!}{2^n} = \frac{2}{n+1}.$$

As

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 0 < 1$$

we have that the series converges by the ratio test.

c).

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

Use the root test

$$\sqrt[n]{a_n} = \sqrt[n]{\frac{e^n}{n^n}} = \frac{e}{n}.$$

As

$$\lim_{n\to\infty} \sqrt[n]{a_n} = 0 < 1$$

the series converges.

d).

$$\sum_{n=2}^{\infty} \frac{\ln n}{e^{\sqrt{n}}}$$

Use the comparison test, as

$$\lim_{x\to\infty} \frac{3\ln x}{\sqrt{x}} = 0,$$

(Use L'hopitals rule to see this). This means that at some point $\sqrt{x} \geq 3\ln x$, i.e. there is some natural number N such that for all $n > N$, we have $\sqrt{n} \geq 3\ln n$. And for this we have

$$\begin{aligned} \sqrt{n} &\geq \ln n^3 \\ e^{\sqrt{n}} &\geq n^3 \\ \frac{1}{n^3} &\leq \frac{1}{e^{\sqrt{n}}} \end{aligned}$$

As $\ln n \leq n$, we see that for $n > N$ we have

$$\frac{\ln n}{e^{\sqrt{n}}} \leq \frac{n}{n^3} = \frac{1}{n^2}.$$

Via the p-series test, we know $\sum \frac{1}{n^2}$ converges, thus the series converges by the comparison test.

Lecture 15 - 7/26/24

Exercises for section 5.3:

1. Deduce the convergence or divergence of the following series using the convergence tests in this section:
   (a) $\sum_{n\geq 1} \frac{1}{\sqrt{n^2+3n+1}}$.
   (b) $\sum_{n\geq 1} \frac{3n+4}{n^3+7}$.
   (c) $\sum_{n\geq 1} \frac{n^4}{4^n}$.
   (d) $\sum_{n\geq 1} \frac{2^{-n}+3n}{3^{-n}+n^2}$
   (e) $\sum_{n\geq 1} \sqrt{n^4+3n^{5/4}} - n^2$
   (f) $\sum_{n\geq 1} \frac{a_n}{n}$, $a_n := n \mod 5$.
   (g) $\sum_{n\geq 1} \frac{a_n}{n^2}$, $a_n := n \mod 5$.
   (h) $\sum_{n\geq 1} \frac{(n+1)^2}{(n+2)!}$.
   (i) $\sum_{n\geq 1} 4^{-n}(n \mod 4)^n$.
2. Determine the convergence or divergence of a series whose nth term is given by
   (a)

$$a_n = \frac{n^2}{3n^2-1}$$

(b)

$$a_n = \frac{e^{n^2}}{n! n^n}$$

(c)

$$a_n = \frac{1}{n^3 + 4\sqrt{n}}$$

(d)

$$a_n = \frac{\ln n}{n}, \quad n \geq 2$$

5.4 Absolute & Conditional Convergence

defining rearrangements

theorem about rearrangement not effecting absolute convergence

divergence of parts of conditionally convergent sums

riemann rearrangement

finite rearrangements change nothing

In this section we will look at series made up of general terms, i.e. we are no longer restricting ourselves to series of nonnegative terms. We saw in the previous section that we can apply most of our convergence tests to series of general terms, but this leads us now to the following: every series $\sum a_n$ of general terms has an associated series of nonnegative terms $\sum |a_n|$. In this section we will look at the connections and differences that can arise in studying a series and its associated series of absolute value terms.

Definition 35. *A series $\sum a_n$ is said to be* **absolutely convergent** *if $\sum |a_n|$ converges. If $\sum a_n$ converges but $\sum |a_n|$ diverges, then the series is called* **conditionally convergent**.

The first result that we come to is that absolute convergence is a very strong condition in the sense that if $\sum a_n$ converges absolutely, then the series $\sum a_n$ converges outright. This is one reason for the naming of conditionally convergent series as 'conditional' as there are no absolutely convergent series that are not convergent.

Theorem 52. *If a series is absolutely convergent, then it is convergent.*

Proof. Assume that $\sum a_n$ is an absolutely convergent series. Let $s_n = \sum_{k=1}^{n} a_n$ denote the sequence of partial sums of the series $\sum a_n$, and $t_n = \sum_{k=1}^{n} |a_n|$ denote the sequence of partial sums of the series $\sum |a_n|$.

By assumption, $\sum |a_n|$ converges, hence $\{t_n\}$ converges, hence $\{t_n\}$ satisfies the Cauchy condition. Thus, let $\epsilon > 0$, and for this choice of epsilon, there is an $N \in \mathbb{N}$ such that

$$|t_m - t_n| < \epsilon, \quad \forall m, n > N.$$

For this same m and n (both greater than N) look at the following (Without loss of generality assume that $n > m$ as well)

$$\begin{aligned}|s_n - s_m| &= \left|\sum_{k=1}^{n} a_k - \sum_{k=1}^{m} a_k\right| = \left|\sum_{k=m+1}^{n} a_k\right| \\ &= |a_{m+1} + a_{m+2} + \cdots + a_n| \\ &\leq |a_{m+1}| + |a_{m+2}| + \cdots + |a_n| = \sum_{k=m+1}^{n} |a_k| \\ &= t_n - t_m = |t_m - t_n| < \epsilon\end{aligned}$$

Note: the third line came from the triangle inequality. Thus, this shows that the sequence $\{s_n\}$ satisfies the Cauchy condition, thus $\{s_n\}$ converges, thus the series $\sum a_n$ converges. □

Real quick, let's take a quick inventory of a few facts. We saw that for $\sum a_n$ to converge a necessary condition is that $\lim_{n\to\infty} a_n = 0$, and from the comparison test, p-series, and harmonic series, we know that the decay to 0 on the terms a_n must be faster than $\frac{1}{n}$. The theorem we have just proven states that for a series $\sum a_n$, if its associated series of nonnegative terms $\sum |a_n|$ converges then the original series converges. So this leaves very little room for what a conditionally convergent series can look like. In particular, everything we have seen suggests that a conditionally convergent series can not be made up of nonnegative terms[48], and thus there is some internal cancellation or deconstructive interference in a conditionally convergent series that lets convergence happen. Because of this, the following theorem will help us determine convergence of conditionally convergent series.

Definition 36. *If $a_n > 0$ for every n, the series $\sum(-1)^n a_n$ and $\sum(-1)^{n+1} a_n$ are called* **alternating series**.

Theorem 53. (Alternating Series Test) *Let $\sum(-1)^{n+1} a_n$ be an alternating series such that*

i). $a_n \geq a_{n+1} > 0$ for every n.

ii). $\lim_{n\to\infty} a_n = 0$

Then $\sum(-1)^{n+1} a_n$ and $\sum(-1)^n a_n$ converge.

Do note that from theorem 43 that a series not meeting ii). of the criteria above are divergent.

Proof. We will prove the convergence of the series $\sum(-1)^{n+1} a_n$. The convergence of $\sum(-1)^n a_n$ follows by a similar proof. Let $\{s_n\}$ denote the sequence of partial sums of $\sum(-1)^{n+1} a_n$, and in fact we will look at the subsequence of this sequence, given by $\{s_{2n}\}$.

[48] we will soon see this to be a fact

We first notice that

$$\begin{aligned} s_{2n} &= \sum_{k=1}^{2n}(-1)^{k+1}a_k = a_1 - a_2 + \cdots - a_{2n} \\ &= (a_1 - a_2) + (a_3 - a_4) + \cdots + (a_{2n-1} - a_{2n}) \\ &= \sum_{j=1}^{n}(a_{2j-1} - a_{2j}) \\ &= \sum_{j=1}^{n-1}(a_{2j-1} - a_{2j}) + (a_{2n-1} - a_{2n}) = s_{2(n-1)} + (a_{2n-1} - a_{2n}) \\ &\geq s_{2(n-1)} \end{aligned}$$

The inequality follows due to the assumption that $a_{2n} \leq a_{2n-1}$. So, we see that the subsequence $\{s_{2n}\}$ is an increasing sequence.

We can also see,

$$\begin{aligned} s_{2n} &= a_1 - a_2 + a_3 - a_4 \cdots - a_{2n} \\ &= a_1 - (a_2 - a_3) - (a_4 - a_5) \cdots - (a_{2n-2} - a_{2n-1}) - a_{2n} \\ &\leq a_1 \end{aligned}$$

And the inequality here follows from, once again, the assumption that $a_k \geq a_{k+1}$, thus all the terms we are subtracting off in the line above are positive (and the last term we are subtracting off is positive by assumption) This shows that the sequence $\{s_{2n}\}$ is bounded.

So, by the monotone convergence theorem $\{s_{2n}\}$ converges to some real number, let us call it α. But then we have

$$\begin{aligned} \lim_{n\to\infty} s_{2n+1} &= \lim_{n\to\infty} s_{2n} + a_{2n+1} \\ &= \lim_{n\to\infty} s_{2n} + \lim_{n\to\infty} a_{2n+1} \\ &= \alpha + 0 = \alpha \end{aligned}$$

By the assumption that $\lim a_n = 0$. This means the subsequence $\{s_{2n+1}\}$ converges to the same limit as $\{s_{2n}\}$. Thus, by a previous homework problem (section 2.3 3a).), we have that $\{s_n\}$ converges. Thus, the alternating series $\sum(-1)^{n+1}a_n$ converges. □

Example 22. *Let us look at the series*

$$\rho = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$$

we have already seen that its associated series of absolute value terms is

$$H = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

which diverges. But ρ is of the form $\sum(-1)^{n+1}a_n$ with $a_n = \frac{1}{n}$ and we see that $\{a_n\}$ satisfies the two conditions of the alternating series test. Thus ρ is a conditionally convergent series. In fact later we will see that

$$\rho = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = \ln 2$$

in our study of Taylor series. However, we will see shortly that the terms in this sum must be in this specific order.

Example 23. *Determine whether the following are absolutely convergent, conditionally convergent, or neither.*

a).

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

For absolute convergence we will see if the following series converges

$$\sum_{n=1}^{\infty} \frac{1}{2n^2+1}$$

As $n^2 \leq 2n^2+1$ *we have that*

$$\sum_{n=1}^{\infty} \frac{1}{2n^2+1} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Thus, by the comparison test, this series converges. Thus the original series converges absolutely.

b).

$$\sum_{n=2}^{\infty} \frac{(-1)^n}{\ln n}$$

For absolute convergence we will determine whether or not the series

$$\sum_{n=2}^{\infty} \frac{1}{\ln n}$$

converges. As $\ln n$ *grows slower than any polynomial we have* $\ln n \leq n$ *for all* $n \in \mathbb{N}$*. Thus, we have*

$$\sum_{n=2}^{\infty} \frac{1}{n} \leq \sum_{n=2}^{\infty} \frac{1}{\ln n}$$

Thus by the comparison test and the fact that the harmonic series diverges, we know that $\sum_{n=2}^{\infty} \frac{1}{\ln n}$ *diverges. So the series does not converge absolutely, but perhaps it converges conditionally.*

The original series is an alternating series so we apply the alternating series test. The terms are $a_n = \frac{1}{\ln n}$*.*

- *As natural log is an increasing function, we have* $\ln n \leq \ln(n+1)$ *for all* $n \in \mathbb{N}$*. Thus*

$$a_{n+1} = \frac{1}{\ln(n+1)} \leq \frac{1}{\ln n} = a_n, \quad \forall n \in \mathbb{N} \cap [2, \infty)$$

- *Secondly we have*

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1}{\ln n} = 0$$

Thus by the alternating series test the original series does converge. So this series is conditionally convergent.

c).

$$\sum_{n=1}^{\infty} \frac{(-1)^n n}{n+4}$$

For this series

$$\sum_{n=1}^{\infty} \frac{(-1)^n n}{n+4}$$

to check for absolute convergence we look at the series

$$\sum_{n=1}^{\infty} \frac{n}{n+4}$$

However as

$$\lim_{n\to\infty} \frac{n}{n+4} = 1 \neq 0$$

this series can not converge. Thus the series does not converge absolutely.

Also, while this is an alternating series, because the limit of the terms do not go to 0 this series will not conditionally converge either.

Now we come to the big difference between absolutely and conditionally convergent series, and the second reason the word 'conditional' is an adequate name for non-absolutely convergent yet convergent series. The big thing is this. Arithmetic with an infinite number of terms present is tricky. When adding up a finite number of terms, the order of the terms does not matter, but when adding up an infinite number of terms the order of the terms can determine everything. This is not a failure of commutativity of addition directly. It is simply due to the fact that infinite sets can contain subsets as numerous as the sets themselves, for example, we add up all the even terms of a sequence and then the odd terms, we may end up in an entirely different place then adding them directly. If anything while the commutative law lets you interchange two terms or a finite number of terms, it can fail in the case of interchanging an infinite number of terms. Before we move on, let us be a little more formal about what we mean when we say 'interchanging' terms.

The series $\sum a_n$ is shorthand for

$$a_1 + a_2 + a_3 + a_4 + a_5 + \cdots$$

A rearrangement would look like, for example,

$$a_{237} + a_{14} + a_1 + a_{1331331} + a_2 + \cdots$$

Thus a rearrangement is just listing all the terms in a sum in a different order and adding them up in a different order. You can think of this as a reindexing of terms or permuting the indices on the terms a_n. (so a permutation of the natural numbers $\mathbb{N}$)

Definition 37. *For a series $\sum a_n$, a* **rearrangement** *of this series will be an bijective function $f : \mathbb{N} \to \mathbb{N}$, and the actual rearrangement is*

$$\sum_{n=1}^{\infty} a_{f(n)}$$

Before we delve into the difference between an absolutely convergent and conditionally convergent series, let us finish a promise on a remark following the definition of the Cauchy condition in section 5.1. In particular, it was said that altering or changing a finite number of terms within a series will not effect its convergent behavior. In the same manner here let us see why a rearrangement that only permutes a finite number of terms in a series does not effect its convergent value (if the series converges).

Theorem 54. *Let $\sum a_n$ be a convergent series (either absolute or conditional) and $L \in \mathbb{R}$ such that*

$$\sum_{n=1}^{\infty} a_n = L$$

and let $f : \mathbb{N} \to \mathbb{N}$ be a rearrangement. Define the fixed point set of f as

$$F = \{n \in \mathbb{N} \mid n = f(n)\}$$

If the compliment of F in $\mathbb{N}$ is finite, then

$$\sum_{n=1}^{\infty} a_{f(n)} = L$$

Thus any rearrangement that only rearranges a finite number of terms does not effect the value the series to.

Proof. Let $\{s_n\}$ denote the partial sum of the original series, thus $\lim_{n\to\infty} s_n = L$. Let $\{t_n\}$ denote the partial sums of the rearranged series, i.e.

$$t_n = \sum_{k=1}^{n} a_{f(k)}$$

By assumption we have that F^c is a finite set, thus $F^c = \{n_1, n_2, \ldots, n_m\}$ for some $m \in \mathbb{N}$. Take

$$M = \max f(F^c) = \max(f(n_1), f(n_2), \ldots, f(n_m))$$

thus we can be sure that $\{f(n_1), f(n_2), \ldots, f(n_m)\} \subseteq \{1, 2, \ldots, M\}$. And for any $n > M$, $n \in F$ and thus $n = f(n)$. Thus for $n > M$ we have

$$\begin{aligned}
t_n - s_n &= \sum_{k=1}^{n} a_{f(n)} - \sum_{k=1}^{n} a_n = \left(\sum_{k=1}^{M} a_{f(n)} + \sum_{k=M+1}^{n} a_{f(n)}\right) - \left(\sum_{k=1}^{M} a_n + \sum_{k=M+1}^{n} a_n\right) \\
&= \sum_{k=1}^{M} a_{f(n)} + \sum_{k=M+1}^{n} a_n - \sum_{k=1}^{M} a_n - \sum_{k=M+1}^{n} a_n = \sum_{k=1}^{M} a_{f(n)} - \sum_{k=1}^{M} a_n \\
&= \sum_{k\in F^c} a_{f(n)} + \sum_{k\in F\cap\{1,2,\ldots,M\}} a_{f(n)} - \sum_{k\in F^c} a_n - \sum_{k\in F\cap\{1,2,\ldots,M\}} a_n \\
&= \sum_{k\in F^c} a_{f(n)} + \sum_{k\in F\cap\{1,2,\ldots,M\}} a_n - \sum_{k\in F^c} a_n - \sum_{k\in F\cap\{1,2,\ldots,M\}} a_n = \sum_{k\in F^c} a_{f(n)} - a_n = 0
\end{aligned}$$

as F^c is a finite set and a finite sum can have its terms reordered, and f is a permutation on F^c since it fixes all elements of F.[49] Thus we see why $\lim_{n\to\infty} |t_n - s_n| = 0$ and so $\lim_{n\to\infty} t_n = L$ by our algebraic limit rules, thus we have

$$\sum_{n=1}^{\infty} a_{f(n)} = L$$

as we wanted to show. □

As we have seen from this theorem we only need to focus on rearrangements that permute some infinite set of terms from $\mathbb{N}$. What we see is that absolute convergence is a strong enough condition to make infinite series act like finite sums in that a rearrangement of terms in an absolutely convergent series will not effect the value the series converges to. But for a conditionally convergent series it will turn out that for any real number there exists a rearrangement of the series that will converge to that specified value. Thus the order of summing terms in a conditionally convergent series must be done with care if one hopes to get a particular value, and conditionally convergent series go against our intuition that the order of summation of terms should not effect the result of the sum.

Theorem 55. *Let $\sum a_n$ be absolutely convergent, then any rearrangement of $\sum a_n$ converges to the same value.*

Proof. Let $\sum a_n$ be an absolutely convergent series, and let $f : \mathbb{N} \to \mathbb{N}$ be a rearrangement of the series. Let us call

$$S_n = \sum_{k=1}^{n} a_k, \qquad T_n = \sum_{k=1}^{n} a_{f(k)},$$

the sequences of partial sums of the original series and the rearranged series, and

$$|S_n| = \sum_{k=1}^{n} |a_k|, \qquad |T_n| = \sum_{k=1}^{n} |a_{f(k)}|,$$

be the sequences of partial sums of the absolute value of the terms from the original series and the rearranged series respectively.

By our assumption, $\{|S_n|\}$ is a convergent sequence and thus is bounded, i.e. there exists a $M > 0$ such that

$$|S_n| < M, \qquad \forall n \in \mathbb{N}.$$

As

$$|T_{n+1}| = \sum_{k=1}^{n+1} |a_{f(k)}| = \sum_{k=1}^{n} |a_{f(k)}| + |a_{f(n+1)}| = |T_n| + |a_{f(n+1)}| \geq |T_n|$$

We see that the sequence $\{|T_n|\}$ is increasing.

Now, let us define the following

$$N_1 = \max\{f(j) \mid 1 \leq j \leq N\}$$

[49] yes, I know I could have appealed to reordering finite sums for any fixed n, I just wanted to make explicit the cancellation on F^c from f being a permutation here.

(this exists as the set is finite),and we have the following

$$|T_N| = \sum_{k=1}^{N} |a_{f(k)}| \leq \sum_{k=1}^{N_1} |a_k| = |S_{N_1}| < M$$

This holds because

- Due to the injectivity of the rearrangement f, we have

  $$f(\{1, 2, \ldots, N\}) \subseteq \{1, 2, \ldots, N_1\}$$

  with equality in these sets only if f is the identity map on $\{1, 2, \ldots, N\}$ (i.e. if f does not rearrange the first N terms)

- The fact that $\{|S_n|\}$ is bounded, so the bound M can be used for any term in the sequence $\{|S_n|\}$.

This shows that the sequence $\{|T_n|\}$ is bounded.

Now, as the sequence $\{|T_n|\}$ is increasing and bounded, by the monotone convergence theorem this sequence converges. This means that the rearrangement is absolutely convergent, and by a theorem from last class, we have that the rearranged series converges.

So, what we have so far is that $\sum a_n$ converges (by assumption), and the rearranged series $\sum a_{f(n)}$ is convergent as well. Equivalently, we have that $\{S_n\}$ converges, and $\{T_n\}$ converges. Now assume that

$$\lim_{n\to\infty} S_n = \alpha,$$

We need to now show that $\lim_{n\to\infty} T_n = \alpha$.

Let us look at what we wish to prove. We would like to show that for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n > N$,

$$|T_n - \alpha| < \epsilon,$$

So, let $\epsilon > 0$ be arbitrary. As $\{S_n\}$ converges to α, there exists an $N_1 \in \mathbb{N}$ such that for all $n > N_1$ we have

$$|S_n - \alpha| = \left| \sum_{k=1}^{n} a_k - \alpha \right| < \frac{\epsilon}{2}.$$

As $\sum a_n$ is absolutely convergent, for this epsilon, there exists $N_2 \in \mathbb{N}$ such that

$$\sum_{k=N_2}^{\infty} |a_k| < \frac{\epsilon}{2}.$$

i.e. as the sum of the absolute value terms converges, the 'tails' of the sum must get arbitrarily small as the 'tail' is the difference between the partial sum and the series.[50] Now take $N = \max\{N_1, N_2\}$, and define

$$M = \max\{f^{-1}(\{1, 2, \ldots, N\})\}$$

[50] this follows from the Cauchy condition

We can choose this M because the set is finite due to the injectivity of f. We pick this M so that we are guaranteed that if we go M terms in the rearranged sum then the rearranged sum contains the first N terms of the original sum. One last definition, call

$$B_n = \{1, 2, \ldots, n\} \backslash f^{-1}(\{1, 2, \ldots, N\})$$

We then have, for $n > M$

$$\begin{aligned} |T_n - \alpha| &= \left| \sum_{k=1}^{n} a_{f(k)} - \alpha \right| \\ &= \left| \sum_{k \in f^{-1}(\{1,\ldots,N\})} a_{f(k)} - \alpha + \sum_{k \in B_n} a_{f(k)} \right| \\ &\leq \left| \sum_{k=1}^{N} a_k - \alpha \right| + \left| \sum_{k \in B_n} a_{f(k)} \right| \end{aligned}$$

where the last line came from the triangle inequality. For $k \in B_n$, $f(k) > N$, thus the second term above has the property

$$\left| \sum_{k \in B_n} a_{f(k)} \right| \leq \sum_{k \in B_n} |a_{f(k)}| \leq \sum_{k=N+1}^{\infty} |a_k|$$

from the triangle inequality and the fact that adding in positive terms still obeys the inequality above. So, we have,

$$|T_n - \alpha| \leq \left| \sum_{k=1}^{N} a_k - \alpha \right| + \sum_{k=N+1}^{\infty} |a_k|$$

and by our definition of N, this means

$$|T_n - \alpha| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus for $n > M$ we have that $|T_n - \alpha| < \epsilon$. As ϵ can be taken arbitrarily, we have

$$\lim_{n \to \infty} T_n = \alpha = \lim_{n \to \infty} S_n,$$

thus the rearranged series converges to the same value as the original series. □

So we have seen that the value of an absolute convergent series does not change when the terms in the series are rearranged. Before we move onto Riemann's rearrangement theorem for conditionally convergent series, we have the next theorem. This theorem gives an answer to what I was hinting at while motivating the alternating series test. While not every conditionally convergent series is necessarily in the form of an alternating series directly, this theorem assures us that a conditionally convergent series contains two divergent positive and negative series.

Theorem 56. *Let $\sum a_n$ be a conditionally convergent series. Choose the positive terms of the series $\sum a_n$ and form a series $\sum b_n$ from these terms. Call the negative terms of the series $\sum c_n$, then $\sum b_n$ diverges to ∞ and $\sum c_n$ diverges to $-\infty$.*

Proof. If we suppose that $\sum b_n$ converges to $M > 0$ and $\sum c_n$ converges to $K < 0$, then we would have

$$\sum_{k=1}^{n} |a_k| = \sum_{k=1}^{n} |b_k| + \sum_{k=1}^{n} |c_k| = \sum_{k=1}^{n} b_k - \sum_{k=1}^{n} c_k \leq M - K$$

This would then imply that the sequence of partial sums of $\sum |a_n|$ is increasing and bounded above, hence by the monotone convergence theorem convergent. This would imply $\sum a_n$ is absolutely convergent, not conditionally convergent. Thus this is not possible.

Thus is must be the case that at least one of $\sum b_n$ or $\sum c_n$ diverges (to ∞ or $-\infty$ respectively). Thus, let us assume $\sum b_n$ diverges to ∞ and that $\sum c_n$ converges to $K < 0$. As $\sum b_n$ diverges to ∞, we have for any $M - K \in \mathbb{R}$, the existence of an N such that

$$\sum_{k=1}^{N} b_k \geq M - K$$

(i.e. the sum will always grow bigger than an arbitrary choice of constant.) But then we have

$$\sum_{k=1}^{N_1} a_k = \sum_{k=1}^{N} b_k + \sum_{k=1}^{N_2} c_k \geq M - K + K = M$$

for some N_2 and $N_1 > N$. As we can find such an N_1 for any constant real M, we have shown that $\sum a_n$ diverges to ∞, but this contradicts the fact that $\sum a_n$ is conditionally convergent. Thus it must be the case that $\sum b_n$ diverges to ∞ and $\sum c_n$ diverges to $-\infty$. □

Theorem 57. (Riemann Rearrangement Theorem) *Let $\sum a_n$ be a conditionally convergent series. Then given any number $L \in \mathbb{R}$, it is possible to find a rearrangement $f : \mathbb{N} \to \mathbb{N}$ such that*

$$\sum_{n=1}^{\infty} a_{f(n)} = L$$

Proof. Let $\sum a_n$ be a conditionally convergent series. Let $\sum b_n$ be the series created from the positive terms of $\{a_n\}$, and $\sum c_n$ be the series created by the negative terms of $\{a_n\}$. By the previous theorem we have that $\sum b_n$ diverges to ∞, and $\sum c_n$ diverges to $-\infty$.

We construct the rearrangement of $\{a_n\}$ in the following manner. Without loss of generality, assume $L > 0$. Define the following subset of $\mathbb{N}$

$$B_1 = \left\{ m \in \mathbb{N} \mid \sum_{k=1}^{m} b_k \geq L \right\}$$

Then as the series $\sum b_n$ diverges to ∞ we have that B_1 is not empty, and thus by the well-ordering principle of $\mathbb{N}$, B_1 has a least element, let us call it n_1. Thus we have

$$b_1 + b_2 + \cdots + b_{n_1} \geq L$$
$$b_1 + b_2 + \cdots + b_{n_1 - 1} < L$$

and the second line comes from n_1 being the least element of B_1, i.e. n_1 is the least number of terms we must take from $\{b_n\}$ to pass the threshold of L from below.

We have currently overshot L, now we need to get back down under it. Define the following set

$$C_2 = \left\{ m \in \mathbb{N} \mid \sum_{k=1}^{n_1} b_k + \sum_{k=1}^{m} c_k \leq L \right\}$$

As $\sum_{k=1}^{n_1} b_k$ is a finite sum of positive terms and hence finite and $\sum c_n$ diverges to $-\infty$ we have that C_2 is not empty and thus by the well ordering principle of $\mathbb{N}$, C_2 has a least element that we call n_2, and we have

$$b_1 + \cdots + b_{n_1} + c_1 + \cdots + c_{n_2} \leq L$$
$$b_1 + \cdots + b_{n_1} + c_1 + \cdots + c_{n_2 - 1} > L$$

Generally we define n_k as least elements of the following sets depending on if k is even or odd

$$B_{2p-1} = \left\{ m \in \mathbb{N} \mid \sum_{k=1}^{m} b_k + \sum_{k=1}^{n_{2p-2}} c_k \geq L,\ m > n_{2p-3} \right\}$$
$$C_{2p} = \left\{ m \in \mathbb{N} \mid \sum_{k=1}^{n_{2p-1}} b_k + \sum_{k=1}^{m} c_k \leq L,\ m > n_{2p-2} \right\}$$

And so we continue to do this over and over, going back and forth between overshooting L and undershooting L.

We now use this process to construct the rearrangement function $f : \mathbb{N} \to \mathbb{N}$,

- Start by having f map the set $\{1, 2, \ldots, n_1\}$ to the indices of a_n such that $a_n \in \{b_1, b_2, \ldots, b_{n_1}\}$. Thus
$$\sum_{k=1}^{n_1} a_{f(k)} = \sum_{k=1}^{n_1} b_k \geq L$$
- Then have f map the set $\{n_1 + 1, n_1 + 2, \ldots, n_1 + n_2\}$ to the indices of a_n such that $a_n \in \{c_1, c_2, \ldots, c_{n_2}\}$ for $n_1 < n \leq n_1 + n_2$. Then
$$\sum_{k=1}^{n_1+n_2} a_{f(k)} = \sum_{k=1}^{n_1} b_k + \sum_{k=1}^{n_2} c_k \leq L$$
- At the general stage, let $N = n_1 + n_2 + \cdots n_m$, and assume that f has been defined on $\{1, 2, \ldots, N\}$, we then define f on $\{N+1, N+2, \ldots, N+n_{m+1}\}$ to map to the indices of the next step of b_k if m is even, and the indices of the next step of c_k if m is odd. In particular we have
$$\sum_{k=1}^{N+n_{m+1}} a_{f(k)} = \begin{cases} \sum_{k=1}^{n_{m+1}} b_k + \sum_{k=1}^{n_m} c_k \geq L, & \text{if } m \text{ is even} \\ \sum_{k=1}^{n_m} b_k + \sum_{k=1}^{n_{m+1}} c_k \leq L, & \text{if } m \text{ is odd} \end{cases}$$
and due to how each n_k was chosen we have that
$$\sum_{k=1}^{N+n_{m+1}-1} a_{f(k)} = \begin{cases} \sum_{k=1}^{n_{m+1}-1} b_k + \sum_{k=1}^{n_m} c_k < L, & \text{if } m \text{ is even} \\ \sum_{k=1}^{n_m} b_k + \sum_{k=1}^{n_{m+1}-1} c_k > L, & \text{if } m \text{ is odd} \end{cases}$$

If N is contained within the following set

$$N \in \left\{ \sum_{k=1}^{m} n_k \mid m \text{ is even} \right\}$$

then we have the following

$$\begin{aligned} \sum_{k=1}^{N-1} a_{f(k)} &< L \\ L \leq \sum_{k=1}^{N} a_{f(k)} &< L + b_{n_m} \qquad \text{adding } b_{n_m} \text{ to both sides} \end{aligned}$$

and thus

$$\left| \sum_{k=1}^{N} a_{f(k)} - L \right| < b_{n_m}$$

As n_m is an increasing function on the naturals (in terms of m), and each term in b_k equals some term in a_n, i.e. $b_{n_m} = a_k$ for some k, we have that $\lim_{m\to\infty} b_{n_m} = 0$ as we know $\lim_{n\to\infty} a_n = 0$ since the original series is conditionally convergent.[51] And this shows that

$$\sum_{n=1}^{\infty} a_{f(n)} = L$$

□

Exercises for section 5.4:

1. Determine whether the series whose nth term is given below converges absolutely, conditionally, or diverges:

   (a)
   $$a_n = \frac{(-1)^n}{\sqrt{n}}$$

   (b)
   $$a_n = \frac{(-1)^n}{n(\ln n)^3}, \quad n \geq 2,$$

   (c)
   $$a_n = \frac{n}{(-2)^n}$$

   (d)
   $$a_n = \frac{(-1)^n}{n^{\frac{1}{4}}}$$

[51] just as a note we could have finished this argument with m being odd and the terms c_{n_m} as well

5.5 Addition & Multiplication of Series (Optional)

Supplemental stuff, add/mult of series, things from rudin, defining e properly. limsup and liminf, general versions of convergence tests.

3.41, 3.42, 3.47, 3.48, 3.49, 3.50, 3.51, dirichlet condition, try and do abels theorem, wait until power series

Supplement on addition and product of series, Abel's theorem, maybe this is better for a later class with power series more involved.

Exercises for section 5.5:

1. (a) Give an example of two convergent series $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ such that $\sum_{k=1}^{\infty} a_k b_k$ diverges.
   (b) Can this happen if one of the series is absolutely convergent ?
2. If $\sum a_n$ converges, and if $\{b_n\}$ is monotonic and bounded, prove that $\sum a_n b_n$ converges.
3. Prove that the Cauchy product of two absolutely convergent series converges absolutely.

5.6 The Exponential Function (Optional)

Two versions of e, how they are equivalent in MCT section. Law of exponents and maybe irrationality and transcendence of e. Maybe make this a supplement and use pg 22 Rudin problems 6 and 7

Doing e with series as this time, nth roots parts with e as well

5.7 More General Convergence Theorems (Optional)

more general convergence tests, integral test maybe.

More general versions of these theorems and 3.37 from rudin in supp. (or make optional section)

Thm: (*Integral Test*) Let $\sum a_n$ be a series of positive numbers with

$$a_1 \geq a_2 \geq a_3 \geq \cdots$$

Let $f(x)$ be a nonincreasing continuous function on $(0, \infty)$ such that $f(n) = a_n$ for each positive integer n. Then $\sum a_n$ converges if and only if the improper integral $\int_1^{\infty} f(x)dx$ converges.

Proof. If we return to calculus 1 for a moment, and recall how the Riemann integral is defined as a limit of Riemann sums, we recall the definitions:

- The *right endpoint approximation* of the Riemann integral of f over an interval (a, b) is denoted by $R(f, \Delta x)$. This is found by summing the areas of rectangles of width Δx and height $f(x_k)$ where x_k is the right endpoint of a subinterval of (a, b) found by partitioning (a, b) into subintervals of length Δx.

- The *left endpoint approximation* of the Riemann integral of f over an interval (a, b) is denoted by $L(f, \Delta x)$. This is found by summing the areas of rectangles of width Δx and height $f(x_k)$ where x_k is the left endpoint of a subinterval of (a, b) found by partitioning (a, b) into subintervals of length Δx.

It is clear from a picture, that for a non increasing function $f(x)$ that is integrable over (a, b) that

$$R(f, \Delta x) \leq \int_a^b f(x)dx \leq L(f, \Delta x)$$

The proof simply relies on the following, for the interval $(1, \infty)$, and $\Delta x = 1$, we have

$$L(f, 1) = \sum_{n=1}^{\infty} a_n, \qquad R(f, 1) = \sum_{n=2}^{\infty} a_n$$

Then, we have if $\sum a_n$ converges, then

$$\int_1^{\infty} f(x)dx \leq L(f, 1) = \sum a_n,$$

so the improper integral converges. If the integral converges, then

$$\sum_{n=2}^{\infty} a_n = R(f, 1) \leq \int_1^{\infty} f(x)dx.$$

implies that $\sum_{n=2}^{\infty} a_n$ converges, so adding a_1 will still be convergent, thus $\sum a_n$ converges. $\square$

Lecture 16 - 7/29/24

6 Topology of $\mathbb{R}$

Before we jump into the subsections about specific kinds of sets in $\mathbb{R}$, let us have a quick reminder about interval notation. Recall the following definitions for $a, b \in \mathbb{R}$ with $a < b$.

$$\begin{aligned}(a,b) &= \{x \in \mathbb{R} \mid a < x < b\}\\ [a,b] &= \{x \in \mathbb{R} \mid a \leq x \leq b\}\\ [a,b) &= \{x \in \mathbb{R} \mid a \leq x < b\}\\ (a,b] &= \{x \in \mathbb{R} \mid a < x \leq b\}\end{aligned}$$

Intervals of the form (a,b) are often called open intervals and intervals of the form $[a,b]$ are called closed intervals. The reason for these names will be justified in first two sections.

6.1 Open Sets

Definition 38. *For $x \in \mathbb{R}$, a* **neighborhood** *of x, (often simply referred to as an 'nhood' of x), is any interval of the form (a,b) that contains x. However, we usually take this to mean an interval of the form $(x-\delta, x+\delta)$ for some $\delta > 0$ so the interval is symmetric about x or centered about x.*

A set A in the reals is called **open** *if every $x \in A$ has a neighborhood that is entirely contained within A, or equivalently, for each $x \in A$, there exists an $\epsilon > 0$ such that*

$$(x-\epsilon, x+\epsilon) \subseteq A$$

Example 24. *Determine if the following sets are open.*

a). $A = (0,1)$. For $x \in (0,1)$ as this means $0 < x < 1$ if we simply take the midpoint between x and 0 and x and 1 respectively, we see that

$$x \in \left(\frac{x}{2}, \frac{x+1}{2}\right) \subseteq (0,1)$$

Thus every $x \in (0,1)$ has a neighborhood that stays within $(0,1)$, and so $(0,1)$ is an open set.

b). $B = [0,1]$. This set is not open. Take $x = 0$. Any neihborhood of this point is of the form $(-\epsilon, \epsilon)$ for $\epsilon > 0$. Because of this, no neighborhood of 0 is entirely contained within $[0,1]$ and thus $[0,1]$ is not open.

c). $C = (0,1]$. This set is also not open. The reason for this is similar to the argument in b). No neighborhood of 1 will be entirely contained within C and thus the set is not open.

d). $D = \mathbb{R}$. The real numbers form an open set. This is simply because any neighborhood of any point $x \in \mathbb{R}$ is a subset of $\mathbb{R}$ and thus every point in the reals has a neighborhood contained within the reals trivially.

e). $E = \varnothing$. The empty set is an open set vacuously, as it is true that that empty set contains a neighborhood of x for all $x \in \varnothing$, because there is no x in $\varnothing$, thus the criteria of being open is met vacuously.

f). $F = \mathbb{Q}$. The rationals are not an open set[52] *due to the density of the rationals within the reals. Given $q \in \mathbb{Q}$, for any $\epsilon > 0$, the interval $(q-\epsilon, q+\epsilon)$ contains both rational and irrational numbers, and thus $(q-\epsilon, q+\epsilon) \not\subseteq \mathbb{Q}$. In fact as we will see in a later section, the rational numbers $\mathbb{Q}$ contain no interval. By a similar argument, the irrational numbers are also not an open set.*[53]

Similar to example a). for constants $a, b \in \mathbb{R}$ with $a < b$, the interval of the form (a, b) is an open set. We will soon see how this is connected to the general structure of open sets within $\mathbb{R}$ but before we do that let us first see a result that connects the notion of open sets with our set operations.

Theorem 58. *An arbitrary union of open sets is open, and a finite intersection of open sets is open.*

Proof. For some arbitrary index set Λ, let A_λ be an open set in $\mathbb{R}$ for each $\lambda \in \Lambda$. And take

$$A = \bigcup_{\lambda \in \Lambda} A_\lambda$$

For any $x \in A$ by the definition of an arbitrary union we have that $x \in A_\lambda$ for at least one $\lambda \in \Lambda$. As this A_λ is open there exists an $\epsilon > 0$ such that $(x-\epsilon, x+\epsilon) \subseteq A_\lambda$. And so we have

$$x \in (x-\epsilon, x+\epsilon) \subseteq A_\lambda \subseteq \bigcup_{\lambda \in \Lambda} A_\lambda = A$$

And this can be done for any $x \in A$, thus A contains a neighborhood of every $x \in A$ and thus A is open.

Now let us prove the second result about finite intersections. So, let $\{A_1, A_2, \ldots, A_N\}$ be a finite collection of open sets within $\mathbb{R}$ and define

$$B = \bigcap_{n=1}^{N} A_n$$

Now take $x \in B$. By definition of intersection we have that $x \in A_n$ for all $1 \leq n \leq N$. As each A_n is open, for each n there is an $\epsilon_n > 0$ such that $(x-\epsilon_n, x+\epsilon_n) \subseteq A_n$. Now take

$$\epsilon = \min(\epsilon_1, \epsilon_2, \ldots, \epsilon_N)$$

then by definition of minimum we have

$$(x-\epsilon, x+\epsilon) \subseteq (x-\epsilon_n, x+\epsilon_n) \subseteq A_n \quad \text{for each } 1 \leq n \leq N$$

and as this is contained in every A_n we have that

$$x \in (x-\epsilon, x+\epsilon) \subseteq \bigcap_{n=1}^{N} A_n = B$$

And as $x \in B$ was arbitrary, we have that B contains a neighborhood of x for every $x \in B$, thus B is open. □

[52] within $\mathbb{R}$

[53] once again, within $\mathbb{R}$

As the theorem states, one must be careful if you intersect an infinite number of open sets. For example

$$\bigcap_{n=1}^{\infty}\left(-\frac{1}{n},\frac{1}{n}\right)=\{0\}$$

each set $\left(-\frac{1}{n},\frac{1}{n}\right)$ is an open set, but the result $\{0\}$ is not open as clearly no neighborhood of 0 is contained in the singleton set $\{0\}$. Do note that this theorem simply states the conditions in which you are guaranteed to get an open set back when performing certain operations, but it does not preclude every infinite intersection of open sets from being open. For example

$$\bigcap_{n=1}^{\infty}\left(1+\frac{1}{n},\infty\right)=(2,\infty)$$

Our next theorem characterizes the structure of open sets within $\mathbb{R}$ and let's us assume that an open sets in the reals will always be of a certain form.

Lemma 59. *Let $\mathcal{U}=\{(a_\lambda,b_\lambda)\mid\lambda\in\Lambda\}$ be a collection of open intervals in $\mathbb{R}$ that all contain a common point y and let*

$$A=\{a_\lambda\mid\lambda\in\Lambda\},\qquad B=\{b_\lambda\mid\lambda\in\Lambda\}$$

be the collections of left and right endpoints of the intervals respectively, then

$$\bigcup_{\lambda\in\Lambda}(a_\lambda,b_\lambda)=(\inf A,\sup B)$$

where $\inf A$ *is replaced by* $-\infty$ *if A is not bounded below and similarly* $\sup B$ *is replaced by* ∞ *if B is not bounded above.*

Proof. The condition that y is contained in every interval (a_λ,b_λ) is what guarantees that the union of more than one interval can be written in terms of a single interval. Without this condition we could have situations like $(1,2)\cup(2,3)$ which can not be written as a single interval. In other words this guarantees that every a_λ is a lower bound for B and every b_λ is an upper bound for A.

For the start of the proof let us assume that A is bounded below and B is bounded above. With these assumptions $\inf A$ and $\sup B$ exist within $\mathbb{R}$ so call $\alpha=\inf A$ and $\beta=\sup B$. By definition of supremum and infimum we have that

$$(a_\lambda,b_\lambda)\subseteq(\alpha,\beta)\text{ for all }\lambda\in\Lambda$$

Thus it follows that

$$\bigcup_{\lambda\in\Lambda}(a_\lambda,b_\lambda)\subseteq(\alpha,\beta)$$

Now for an arbitrary $\epsilon>0$, by definition of an infimum, there is some a_λ with $\alpha<a_\lambda<\alpha+\epsilon$ and similarly there is some $b_{\lambda'}$ with $\beta-\epsilon<b_{\lambda'}<\beta$ thus

$$(\alpha+\epsilon,\beta-\epsilon)\subseteq(a_\lambda,b_{\lambda'})\subseteq\bigcup_{\lambda\in\Lambda}(a_\lambda,b_\lambda)$$

And since this holds for all $\epsilon>0$ we have that

$$(\alpha,\beta)=\bigcup_{\epsilon>0}(\alpha+\epsilon,\beta-\epsilon)\subseteq\bigcup_{\lambda\in\Lambda}(a_\lambda,b_\lambda)$$

and so as we have shown both directions of containment we have that

$$\bigcup_{\lambda\in\Lambda}(a_\lambda, b_\lambda) = (\alpha, \beta)$$

If A is not bounded below, then we replace α with $-\infty$ as the union can not be contained in (M, β) for $M < 0$, otherwise this would imply that M is a lower bound of A. Similarly, if B is not bounded above, then we must replace β with ∞. □

Theorem 60. *For $A \subseteq \mathbb{R}$ an open set, A is a finite or countable union of open intervals (of the form (a, b)).*

Proof. Fix $x \in A$. For this x define the following collection of open intervals

$$\Lambda_x = \{(a, b) \mid x \in (a, b) \text{ and } (a, b) \subseteq A\}$$

And now define

$$I_x = \bigcup_{(a,b)\in\Lambda_x}(a, b)$$

From our lemma above we have that $I_x = (c, d)$ for constants c, d (or $c = -\infty$, $d = \infty$). The interval, I_x, is the largest interval containing x, as any interval containing x would be of the form $x \in (e, f)$ and thus $(e, f) \in \Lambda_x$ and so

$$(e, f) \subseteq \bigcup_{(a,b)\in\Lambda_x}(a, b) = I_x$$

If y is another point contained within $I_x = (c, d)$, if we go through this process again, defining Λ_y as the collection of all intervals (a, b) containing y and defining I_y as the union of all intervals from Λ_y, then as $I_x \in \Lambda_y$ we have that

$$I_x \subseteq I_y$$

and as I_y is an interval containing x we have $I_y \in \Lambda_x$ thus

$$I_y \subseteq I_x$$

Thus we have that $I_x = I_y$ where y is any other point in I_x. This makes sense as I_x and I_y are maximally defined, it would be absurd if I_y were larger just because its basepoint was defined differently.

So we use this to define an equivalence class structure. We define $x \sim y$ if $I_x = I_y$. It is easy to check that this is indeed an equivalence relation, and from the definition of the relation we have that

$$[x] = I_x$$

i.e. the equivalence class of x is the maximally defined interval I_x. At this point, we will pick one representative from each class and call this collection of representatives R. Note that since we picked one representative from each class, for $x, y \in R$ if $x \neq y$ then it must be that $I_x \cap I_y = \varnothing$. Using this structure, we have

$$A = \bigcup_{x\in R} I_x$$

and thus our result will be proven if we can argue why there is a finite or countable number of total elements in R.

For I_x with $x \in R$, as I_x is of the form $I_x = (c, d)$, by the density of rationals within the reals, we know that I_x most definitely contains a rational number. Because of this we can choose a rational number $q_x \in I_x$ and in fact replace x with q_x and use q_x as the choice of representative in R. Now suppose that we have done this for every interval I_y and every representative $y \in R$. We then have $I_{q_x} \cap I_{q_y} = \varnothing$ if $q_x \neq q_y$, i.e. the map from intervals to its rational number representative is injective. Because of this the number of intervals must be less than the cardinality of the rational numbers, and so the union above is countable at most. □

From our section on subsequences, theorem 16 stated that $\{x_n\} \to L$ can be equivalently phrased as saying the sequence $\{x_n\}$ is eventually within the interval $(L - \epsilon, L + \epsilon)$. Because of this, we can phrase convergence in terms of neighborhoods. We say that $\{x_n\}$ converges to L if for every neighborhood U of L there is some $N \in \mathbb{N}$ such that $x_n \in U$ for all $n > N$. We mention this now as this will play a later role in a second way of phrasing continuity of a function f in section 7, but this generalization is also how sequence convergence is defined in more abstract topological settings.

We next come to a definition and an operation that will find the largest open set within a given set in $\mathbb{R}$.

Definition 39. *For a set A in the reals, a point $x \in A$ is called an* **interior point** *if there exists a neighborhood of x that is contained entirely in A. Equivalently, x is an interior point of A if there exists $\epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \subseteq A$.*

The collection of all interior points of A is called the **interior** *of A and is written A°. It is clear from the definition at the start of this section that if A is open, then $A^\circ = A$.*

We have the following proposition about the interior operation on subsets of $\mathbb{R}$. In this proposition we will also the converse of the last statement in the above definition. In particular, we will have that A is open if and only if $A = A^\circ$.

Proposition 61. *For a set A in the reals, we have the following.*

a). A° is an open set.

b). If B is an open set and $B \subseteq A$, then $B \subseteq A^\circ$.

c).

$$A^\circ = \bigcup_{\substack{B \text{ is open} \\ B \subseteq A}} B$$

d). The interior operation is idempotent in that $(A^\circ)^\circ = A^\circ$.

Do note that parts b). and c). state that A° is the largest open set that is contained within A.

Proof. Proof of a). Take $x \in A^\circ$. Thus as x is an interior point of A we have the existence of $\epsilon > 0$ such that $(x-\epsilon, x+\epsilon) \subseteq A$. Taking any $y \in (x-\epsilon, x+\epsilon)$ and defining $\delta = \min(|x-\epsilon-y|, |x+\epsilon-y|)$, we have

$$(y - \delta, y + \delta) \subseteq (x - \epsilon, x + \epsilon) \subseteq A$$

Thus we have shown that every $y \in (x-\epsilon, x+\epsilon)$ is an interior point of A, and thus

$$(x-\epsilon, x+\epsilon) \subseteq A^\circ$$

As x was an arbitrary element of A° and we have the existence of a neighborhood of x contained within A° we have that A° is open.

Proof of b). Take $x \in B$, as B is open we have that $(x-\epsilon, x+\epsilon) \subseteq B$ for some $\epsilon > 0$, and so

$$(x-\epsilon, x+\epsilon) \subseteq B \subseteq A$$

And thus x is an interior point of A. As x was an arbitrary element of B, this shows that $B \subseteq A^\circ$.

Proof of c). From part b). we have that $B \subseteq A^\circ$ for any set B that is open and $B \subseteq A$, so

$$\bigcup_{\substack{B \text{ is open} \\ B \subseteq A}} B \subseteq A^\circ$$

But from part a). A° is an open set with $A^\circ \subseteq A$, and so is an element in the set the union is taken over, so

$$A^\circ \subseteq \bigcup_{\substack{B \text{ is open} \\ B \subseteq A}} B$$

Proof of d). Take $x \in (A^\circ)^\circ$, thus x is an interior point of A°, so there exists some $\epsilon > 0$ such that

$$(x-\epsilon, x+\epsilon) \subseteq A^\circ \subseteq A$$

and thus x is an interior point of A, so $(A^\circ)^\circ \subseteq A^\circ$. Taking $y \in A^\circ$, there exists $\delta > 0$ such that

$$(y-\delta, y+\delta) \subseteq A$$

However following part a). we can show that each $x \in (y-\delta, y+\delta)$ has a neighborhood entirely contained within $(y-\delta, y+\delta)$ and thus within A. This means that

$$(y-\delta, y+\delta) \subset A^\circ$$

and so y is an interior point of A°, thus $A^\circ \subseteq (A^\circ)^\circ$. □

Example 25. *Let us find the interiors of the following sets.*

a). $(0,1]$. *As we saw at the examples are the start of this section, every point* $0 < x < 1$ *is an interior point of* $(0,1]$, *and* $x = 1$ *is not an interior point as every neighborhood of 1 leaves the set* $(0,1]$, *thus*

$$(0,1]^\circ = (0,1)$$

b). $\mathbb{Q}$. *As we saw in the earlier example, the rational numbers contain no interval due to the density of the irrational numbers within the reals. Thus, no* $q \in \mathbb{Q}$ *is an interior point, so*

$$\mathbb{Q}^\circ = \varnothing$$

Exercises for section 6.1:

1. Are the following sets open ? Prove or disprove.
   (a) $S = (0, 1) \cup \{2\}$
   (b) $S = (2, 3) \cup [3, \infty)$
   (c) $S = (-1, 2) \cup [3, \infty)$
   (d) $S = \cap_{n=1}^{\infty}(0, \frac{1}{n})$.
2. Find the interiors of the following sets.
   (a) $A = (0, 5]$.
   (b) $B = \mathbb{Z}$.
   (c) $C = [-4, 3) \cup (6, 17) \cup [128, 256] \cup \{312\}$.
   (d) $D = \left\{1 - \frac{1}{n} \mid n \in \mathbb{N}\right\}$
   (e) $E = \bigcup_{k=1}^{10}[k, k + 1)$
3. Let A an open set.
   (a) Show that if a finite number of points are removed from A, the remaining set is still open.
   (b) Is the same true if a countable number of points are removed ?
4. Show that any nonempty open set is uncountable.

Lecture 17 - 7/31/24

Midterm Today

Lecture 18 - 8/2/24

DNE

Lecture 19 - 8/5/24

6.2 Closed Sets

We begin this section with a defintion

Definition 40. *For a set E of real numbers, a point $p \in \mathbb{R}$ is called a* **limit point** *(or equivalently an* **accumulation point***) if every neighborhood of p meets E in some point q with $q \neq p$. Equivalently, for any $\epsilon > 0$, there is $q \in E$ with*

$$0 < |q - p| < \epsilon$$

Similar to a proof we saw in our section on sequences, a limit point, p, of a set, E, gets it name because it is a point that can be written as a limit of a sequence coming from E. Letting $\epsilon = 1$, the definition above implies the existence of $q_1 \in E$ with $q_1 \neq p$ but within a distance of 1 from p. For each $n \in \mathbb{N}$ we can let $\epsilon = \frac{1}{n}$ and find $q_n \in E$, $q_n \neq p$ that is within $\frac{1}{n}$ of p. And then clearly, $\{q_n\} \to p$. In fact, from our definition we see that even more is true, and this gives some rationale for how the equivalent name of accumulation point came into being.

Theorem 62. *If p is a limit point of E, then every neighborhood of p contains an infinite number of points from E.*

Proof. We proceed by contradiction. Let us assume that p is a limit point of E and that there is an $\epsilon > 0$ such that the cardinality of $(p-\epsilon, p+\epsilon) \cap E$ is finite. Thus there exists a $N \in \mathbb{N}$ such that we can label the points in this set,

$$(p-\epsilon, p+\epsilon) \cap E = \{x_1, x_2, \ldots, x_N\}$$

As there is a finite number of such points we can define

$$\epsilon' = \min(|p-x_1|, |p-x_2|, \ldots, |p-x_N|)$$

But then by this definition $\epsilon' > 0$ gives a neighborhood about p with

$$(p-\epsilon', p+\epsilon') \cap E = \varnothing$$

i.e. an interval about p that never meets E and this contradicts the definition of p being a limit point, and thus we have reached a contradiction. □

Remark 4. *Okay, there is a bit of subtlety here to watch out for. In our previous section on subsequences, we had the theorems 16 and 17 which stated that L was a limit of a sequence $\{x_n\}$ if $(L-\epsilon, L+\epsilon)$ contained some tail of the sequence for any $\epsilon > 0$, and respectively that L was a subsequential limit of $\{x_n\}$ if $(L-\epsilon, L+\epsilon)$ contained an infinite number of terms of $\{x_n\}$ for any $\epsilon > 0$. In either case L is the limit of $\{x_n\}$ or the limit of some subsequence of $\{x_n\}$ but there is no guarantee that for,*

$$A = \{x_n \mid n \in \mathbb{N}\}$$

i.e. the set of terms of the sequence, that L is a limit point of A.

The reason for this is that the definition of a limit point, p, of a set, A requires that every neighborhood of p meets A in some point besides p. A constant sequence $\{x_n\} = \{2\}$ has 2 as the limit of the sequence, but 2 is not a limit point of the set $\{2\}$. So, be careful, being the limit of a sequence or subsequence does not necessarily mean you are a limit point for the set of terms in the sequence.

Definition 41. *For a set E, a point p is called an* **isolated point** *if there exists a neighborhood of p that is disjoint from E, i.e. there exists $\epsilon > 0$ such that $(p-\epsilon, p+\epsilon) \cap E = \varnothing$. A set E in which every $e \in E$ is isolated is called* **discrete**.

Corollary 63. *Any finite set is discrete and any finite set has no limit points.*

Putting this corollary in the context of the previous remark, and similar to what we saw in Case 1 of the proof of the Bolzano-Weierstrass theorem is that a sequence $\{x_n\}$ whose terms come from a finite set can only converge if and only if it is eventually constant. However, do note that this result is not saying that discrete sets do not have limit points, in fact, in the homework you will see that a discrete set can have limit points but they can not be contained within the set itself.

Definition 42. *For a set E of real numbers, the* **derived set** *of E, given by E', is the set of all limit points of E. A set E is called* **closed** *if it contains all of its limit points, i.e. $E' \subseteq E$.*

Example 26. *Let us find the limit points of the following sets.*

a). $A = \left\{\frac{1}{n} \mid n \in \mathbb{N}\right\}$. *For any* $x \in A$, *we have that* $x = \frac{1}{n}$ *for some* $n \in \mathbb{N}$ *and* x *is in between*

$$\frac{1}{n+1} < x < \frac{1}{n-1}$$

The number x *is closer to* $\frac{1}{n+1}$ *with a distance of* $\frac{1}{n(n+1)}$ *between them. Thus taking* $\epsilon = \frac{1}{n(n+1)}$, *we see that* $(x-\epsilon, x+\epsilon) \cap A = \varnothing$. *Thus* x *is an isolated point of* A, *and in fact this shows that* A *is discrete, but this does not mean that* A *does not have a limit point in* $\mathbb{R}$. *For* 0, *and any* $\epsilon > 0$, *the archimedean property states that there exists a natural number* n *with* $\frac{1}{n} < \epsilon$, *but this means that*

$$\frac{1}{n} \in (0-\epsilon, 0+\epsilon) \cap A$$

and this shows that 0 *is a limit point of* A, *and in fact this is the only limit point of* A, *so* $A' = \{0\}$. *But as* A *does not contain its limit points,* A *is not closed.*

b). $B = \left\{\frac{1}{n} \mid n \in \mathbb{N}\right\} \cup \{0\}$. *Based of the work that was done above in part a). the set* B *is not discrete as 0 is not an isolated point, however now as* B *does contain its limit points we have that* B *is closed.*

c). $C = (0,1)$. *For* $0 < x < 1$ *and any* $\epsilon > 0$ *it is clear that* $(x-\epsilon, x+\epsilon)$ *meets* $(0,1)$ *as one of* $\frac{x}{2}, \frac{x+1}{2}, x + \frac{\epsilon}{2}, x - \frac{\epsilon}{2}$ *is in both sets depending on if* $\epsilon > 1$ *or* $\epsilon < 1$. *And as this can be done for any* $\epsilon > 0$ *we have that any* $x \in (0,1)$ *is a limit point of* C. *The number* $x = 0$ *is a limit point of* C *as for any* $\epsilon > 0$ *we have* $\frac{\epsilon}{2} \in (-\epsilon, \epsilon) \cap C$, *and for similar reasons* $x = 1$ *is a limit point of* C *as well. Thus* $C' = [0,1]$ *and* C *is not closed.*

d). $D = [0,1]$. *From the work we have seen in part c). we have that* D *is a closed set.*

e). $E = \mathbb{Q}$. *From the density of* $\mathbb{Q}$ *within itself we immediately have that the rational numbers are not a discrete set, in fact no* $x \in \mathbb{Q}$ *is isolated. Similarly, from the density of the rationals in the reals, every real number is a limit point of* E, *and thus* $E' = \mathbb{R}$. *Thus the rational numbers are not closed.*

f). $F = \mathbb{R}$. *From the work we did in part e). we see that the real numbers are a closed set.*

Hopefully the above examples give you some intuition about closed sets. They also showed that there are discrete sets that are not closed and closed sets that are not discrete. Example e). showed there are sets that are neither discrete nor closed, and we will see sets that are both discrete and closed momentarily. Next, we prove a theorem that shows that connection between closed and open sets.

Theorem 64. *A set* E *is closed if and only if* $E^c = \mathbb{R}\backslash E$ *is open.*

Proof. $\implies$ Assume that E is closed, then for any $x \in \mathbb{R}\backslash E$, x can not be a limit point of E as E is closed and contains all of its limit points. Thus there must exist an $\epsilon > 0$ such that

$$(x-\epsilon, x+\epsilon) \cap E = \varnothing$$

but this is another equivalent way of saying $x \in (x-\epsilon, x+\epsilon) \subseteq \mathbb{R}\backslash E$ and so $(x-\epsilon, x+\epsilon)$ is a neighborhood of x contained entirely within $\mathbb{R}\backslash E$. As this can be done for any choice of $x \in \mathbb{R}\backslash E$, we have shown that all points $x \in \mathbb{R}\backslash E$ are interior points and thus $\mathbb{R}\backslash E$ is open.

$\Longleftarrow$ Now assume that $\mathbb{R}\backslash E$ is open and take $x \in E'$. As x is a limit point, from the definition of limit point, every neighborhood of x meets E in some point $q \neq x$. This means for any $\epsilon > 0$ that the neighborhood $(x-\epsilon, x+\epsilon) \not\subseteq \mathbb{R}\backslash E$ and so x is not an interior point of $\mathbb{R}\backslash E$ and so it must be that $x \notin \mathbb{R}\backslash E$ since an open set contains only interior points. Thus $x \in E$. This shows that $E' \subseteq E$ and thus E contains all of its limit points and thus is closed. □

Example 27. *With this example we can see many sets are closed using prior results about open sets.*

a). From earlier $B = \left\{\frac{1}{n} \mid n \in \mathbb{N}\right\} \cup \{0\}$, *as*

$$B^c = (-\infty, 0) \cup \left[\bigcup_{n=1}^{\infty} \left(\frac{1}{n+1}, \frac{1}{n}\right)\right] \cup (1, \infty)$$

is a countable union of open sets and hence open, we have that B *is closed.*

b). The reals $\mathbb{R}$ *are closed as* $\mathbb{R}^c = \varnothing$ *is open. Similarly, the empty set* $\varnothing$ *is also closed.*

c). The integers $\mathbb{Z}$ *is closed as*

$$\mathbb{Z}^c = \bigcup_{k \in \mathbb{Z}} (k, k+1)$$

is a union of open sets and hence open. Also note that this set is discrete as well as each integer is isolated.

However one thing to be careful of is that closedness and openness are not a dichotomy unlike evenness and oddness for natural numbers. While it is true that if a natural number is not even then it must be odd, it is **not** true that if a set is not open then it is closed (and similarly not closed does not imply open) In particular

- Sets can be open and not closed - like (a, b).
- Sets can be closed and not open - like $[a, b]$.
- Sets can be both open and closed - like $\varnothing$ and $\mathbb{R}$.[54]
- Sets can be neither closed nor open - like $[a, b)$ and $(a, b]$.

Before our next theorem, recall the DeMorgan Laws for sets that connect union, intersection, and complementation for a common background universe set (in our case this is $\mathbb{R}$). In particular, we have for sets A_λ labeled from some index set Λ that

$$\left[\bigcup_{\lambda \in \Lambda} A_\lambda\right]^c = \bigcap_{\lambda \in \Lambda} A_\lambda^c, \qquad \left[\bigcap_{\lambda \in \Lambda} A_\lambda\right]^c = \bigcup_{\lambda \in \Lambda} A_\lambda^c$$

Because of this and what we have previously shown for open sets, we have the following theorem.

Theorem 65. *An arbitrary intersection of closed sets is closed and a finite union of closed sets is closed.*

[54] in fact these are the only two in the usual topology on $\mathbb{R}$, but with other topologies there can be more 'clopen' sets

Proof. Let us prove the first claim. Let A_λ be a closed set for every $\lambda \in \Lambda$. Then by our theorem from earlier we have that A_λ^c is an open subset of $\mathbb{R}$ for each $\lambda \in \Lambda$. As an arbitrary union of open sets is open we have that

$$\bigcup_{\lambda\in\Lambda} A_\lambda^c \text{ is open}$$

and thus it's complement is closed. So

$$\left[\bigcup_{\lambda\in\Lambda} A_\lambda^c\right]^c = \bigcup_{\lambda\in\Lambda} (A_\lambda^c)^c = \bigcap_{\lambda\in\Lambda} A_\lambda \text{ is closed}$$

For the second claim, assume that $A_1, A_2, \ldots, A_n$ is a finite collection of closed sets. Then $A_1^c, A_2^c, \ldots, A_n^c$ is a finite collection of open sets and from our previous result about open sets we have

$$\bigcap_{k=1}^{n} A_k^c \text{ is open}$$

and thus it's complement is closed, and so by the DeMorgan laws we have

$$\bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{n} (A_k^c)^c = \left[\bigcap_{k=1}^{n} A_k^c\right]^c$$

□

Similar to the result we saw for open sets, note that this guarantees that any finite union of closed sets is closed. The following

$$(0,1] = \bigcup_{n=1}^{\infty} \left[\frac{1}{n}, 1\right]$$

shows that generally a nonfinite union of closed sets can give something other than a closed set. However, do be careful, this does not mean that any nonfinite union of closed sets is not closed as we can see the following,

$$\mathbb{Z} = \bigcup_{k\in\mathbb{Z}} \{k\}$$

but the result simply says you are only guaranteed that a union of closed sets will be closed if the union is over a finite collection.

At the end of our section on open sets, we defined the interior operation. At this point we have a similar operation related to closed sets.

Definition 43. *For a set E in the real numbers we define its* **closure***, written $\overline{E}$, as the union of the set and its limit points, i.e.*

$$\overline{E} = E \cup E'$$

In particular we see that E is closed if and only if $E = \overline{E}$.

Proposition 66. *For a set B in the reals, we have the following.*

a). $\overline{B}$ is a closed set.

b). If A is a closed set and $B \subseteq A$, then $\overline{B} \subseteq A$.

c).

$$\overline{B} = \bigcap_{\substack{A \text{ is closed} \\ B \subseteq A}} A$$

d). The closure operation is idempotent in that $\overline{\overline{B}} = \overline{B}$.

Do note that parts b). and c). of the above say that $\overline{B}$ is the smallest closed set containing B.

Proof. Proof of a). To prove that $\overline{B}$ is a closed set, we need to prove why it contains all of its limit points, so let p be a limit point of $\overline{B}$. If $p \in B$, then $p \in \overline{B}$. Thus let us assume that $p \notin B$ and p is a limit point of $\overline{B}$. Let $\epsilon > 0$ and $(p - \epsilon, p + \epsilon)$ be a neighborhood about p. As p is a limit point of $\overline{B}$ there exists $q \in \overline{B}$ with $q \neq p$ that is contained in the neighborhood $(p - \epsilon, p + \epsilon)$. We now break into cases:

- *Case* 1: $q \in B$. In this case q is a point in the neighborhood of p contained in B with $p \neq q$.
- *Case* 2: $q \in B'$. Thus q is a limit point of B. Taking

  $$\epsilon' = \min(|p + \epsilon - q|, |p - \epsilon - q|)$$

  we then have that $(q - \epsilon', q + \epsilon') \subseteq (p - \epsilon, p + \epsilon)$ and as q is a limit point of B there exists $r \in (q - \epsilon', q + \epsilon')$ with $r \neq q$ and $r \in B$. But then $r \in (p - \epsilon, p + \epsilon)$.

Thus we have shown that for any $\epsilon > 0$, in either case, the existence of a point $r \in B$ with $r \in (p - \epsilon, p + \epsilon)$, and so it must be that p is a limit point of B and so $p \in B'$ and thus

$$p \in B \cup B' = \overline{B}$$

and so the closure of B is a closed set.

Proof of b). If $b \in B$ then $b \in A$ by the assumption that $B \subseteq A$. If $p \in B'$ then for any $\epsilon > 0$ we have that $(p-\epsilon, p+\epsilon) \cap B \neq \varnothing$. However, since A contains B we also have that $(p-\epsilon, p+\epsilon) \cap A \neq \varnothing$. Thus p is also a limit point of A, but as A is closed we have that $p \in A' \subseteq A$. Thus we have just shown that $B' \subseteq A$, and so

$$\overline{B} = B \cup B' \subseteq A$$

Proof of c). From part a). we have that $\overline{B}$ is a closed set containing B, and as any element contained in an arbitrary intersection must be an element of every set in the collection we have that

$$\bigcap_{\substack{A \text{ is closed} \\ B \subseteq A}} A \subseteq \overline{B}.$$

However from our proof above $\bigcap_{\substack{A \text{ is closed} \\ B \subseteq A}} A$ is a closed set as it is an arbitrary intersection of closed sets. And as each member of the intersection contains B we have that

$$B \subseteq \bigcap_{\substack{A \text{ is closed} \\ B \subseteq A}} A$$

But then part b) immediately implies that

$$\overline{B} \subseteq \bigcap_{\substack{A \text{ is closed} \\ B \subseteq A}} A$$

Proof of d). As we saw in part a). the set $\overline{B}$ is closed, thus it contains its limit points, i.e. $\overline{B}' \subseteq \overline{B}$. And so we have

$$\overline{\overline{B}} = \overline{B} \cup \overline{B}' = \overline{B}$$

□

Example 28. *Let us find the closures of the following sets.*

a). The set $A = \left\{ \frac{1}{n} \mid n \in \mathbb{N} \right\}$. *From the work we did in a prior example we have that*

$$\overline{A} = A \cup \{0\}$$

b). The set $\mathbb{Q}$, *we have that* $\overline{\mathbb{Q}} = \mathbb{R}$.

c). For the set $(0, 1)$ *we have that*

$$\overline{(0,1)} = [0, 1]$$

To close out this section, recall theorem 36 from section 4.1. This theorem said that for a bounded set A, if we let $\alpha = \sup A$ and $\beta = \inf A$, then that for any $\epsilon > 0$ there exists elements $a, b \in A$ with $\alpha - \epsilon < a < \alpha$ and $\beta < b < \beta + \epsilon$. In other words, α and β are limit points of A! Because the proof of this result is so similar to theorem 36 it will be omitted but we state the theorem anyways.

Theorem 67. *For a bounded set* B, *then* $\sup(B), \inf(B) \in \overline{B}$. *Put another way, a closed and bounded set contains its infimum and supremum and therefore has largest and smallest elements.*

Note that this is not true for open sets, as $(0, 1)$ has no largest or smallest element.

In the section on open sets we had a theorem that characterized the structure or form that open sets will generally take in $\mathbb{R}$. There is a similar result for closed sets but it requires a more general topic. We will return to this characterization result in the section on perfect sets.

Exercises for section 6.2:

1. Give an example for each of the following:
   a). An infinite set that is discrete.
   b). An infinite set that is not discrete.
2. Show that if a set A contains none of its limit points, then A is discrete.
3. Prove that x is a limit point of A if and only if there exists a sequence $x_1, x_2, \ldots$ of distinct points in A converging to x.
4. Find a set for which $\sup S$ is not a limit point of S.
5. Give an example of a set A that is not closed but such that every point in A is a limit point of A.

6. Are the following sets closed? Prove or disprove.
   (a) $S = [0,1] \cup \{2\}$
   (b) $S = (2,3) \cup [3,\infty)$
   (c) $S = (-\infty, 2] \cup [3,\infty)$
   (d) $S = \cap_{n=1}^{\infty}(0, \frac{1}{n})$.
7. Find the closures of the following sets.
   (a) $A = (0,5]$.
   (b) $B = \mathbb{Z}$.
   (c) $C = [-4,3) \cup (6,17) \cup [128,256] \cup \{312\}$.
   (d) $D = \left\{1 - \frac{1}{n} \mid n \in \mathbb{N}\right\}$
   (e) $E = \bigcup_{k=1}^{10}[k, k+1)$
8. Construct a bounded set of real numbers with exactly three limit points.
9. Construct a compact set of real numbers whose limit points form a countable set.
10. Given a set A, let $A^1 = A'$, $A^2 = (A^1)'$, and generally define $A^{n+1} = (A^n)'$. Show that for every natural number n there is a subset $A \subseteq \mathbb{R}$ such that $A^1, A^2, \ldots, A^{n-1}$ are non-empty but $A^n = \varnothing$.

6.3 Boundary & Density

In this short section let us quickly cover two topics that can be immediately defined with the work we have done do far.

Definition 44. *For a set A in the reals, the* **boundary** *of A, written ∂A, is defined to be the elements of the closure of A that are not in the interior of A, i.e.*

$$\partial A = \overline{A} \backslash A^{\circ}$$

From the definition we can see where the nomenclature 'boundary' comes from. For $x \in \partial A$ we have that x is a limit point of A as it is contained in $\overline{A}$, thus every neighborhood of x meets A in some point other than x. But $x \notin A^{\circ}$, so every neighborbood of x must meet $A^c = \mathbb{R}\backslash A$ as well or it would be an interior point. In other words what we see is that for any $\epsilon > 0$ that

$$x \in \partial A \iff [((x-\epsilon, x) \cup (x, x+\epsilon)) \cap A \neq \varnothing] \wedge [((x-\epsilon, x) \cup (x, x+\epsilon)) \cap \mathbb{R}\backslash A \neq \varnothing]$$

which explains what we mean by x being on the 'edge' or 'boundary' of A, all neighborhoods of x contain points both inside and outside of A.

Example 29. *Find the boundary of the following sets.*

a). $A = (0,1]$. As we have seen in the previous sections, $\overline{A} = [0,1]$ and $A^{\circ} = (0,1)$, thus

$$\partial A = \overline{A}\backslash A^{\circ} = [0,1]\backslash(0,1) = \{0,1\}$$

i.e. just the points 0 and 1, which lines up with our intuition we described above.

b). $B = \mathbb{Q}$. As we saw before $\overline{B} = \mathbb{R}$ and $B^\circ = \varnothing$ due to the density of the rationals in the reals. In particular, the rational numbers contain no interval and so have an empty interior. Because of this

$$\partial B = \overline{B} \backslash B^\circ = \mathbb{R}$$

So every real number is a boundary point of the rationals because of density.

Remark 5. *In fact, more generally from our definitions, in the prior sections we saw that $A = A^\circ$ for open sets and $B = \overline{B}$ for closed sets, so*

- *An open set contains none of its boundary points, i.e. an open set has no boundary.*
- *A closed set contains its boundary points, i.e. a closed set in $\mathbb{R}$ has boundaries above and below.*

This is similar and related to the comment at the end of the section on closed sets that closed sets has largest and smallest elements whilst open sets do not.

In previous sections we saw the density of $\mathbb{Q}$ within the real numbers, however now we would like to present the topic in more generality.

Definition 45. *Given two sets A and B in the reals, we say that A is* **dense** *inside of B if every point of B is a limit point of A. Equivalently, if for any $b \in B$ and any $\epsilon > 0$ there exists $a \in A$ with $a \in (b - \epsilon, b + \epsilon)$, or*

$$A \subseteq B \subseteq \overline{A}$$

Exercises for section 6.3:

1. Find the boundary of the following sets.
   (a) $A = (0, 5]$.
   (b) $B = \mathbb{Z}$.
   (c) $C = [-4, 3) \cup (6, 17) \cup [128, 256] \cup \{312\}$.
   (d) $D = \left\{1 - \frac{1}{n} \mid n \in \mathbb{N}\right\}$
   (e) $E = \bigcup_{k=1}^{10} [k, k + 1)$
   (f) $F = \bigcup_{k \in \mathbb{Z}} (k, k + 1)$.
2. From [FM]
   (a) Show that any number of the form $\sqrt{2}r$ with $r \in \mathbb{Q}$, $r \neq 0$, is irrational.
   (b) Use (a) to show that irrationals $\mathbb{R}\backslash\mathbb{Q}$ are dense in $\mathbb{R}$.
   (c) Deduce that $\mathbb{Q}$ contains no open interval.
3. From [FM]. This exercise constructs three disjoint sets, all of which are dense in $\mathbb{R}$, who all have the same boundary. First solve problem 2.
   (a) First prove that $\sqrt{\frac{2}{3}}$ is irrational, writing a proof similar to the irrationality of $\sqrt{2}$.
   (b) Use (a) to show that the sets $\sqrt{2}\mathbb{Q}\backslash\{0\}$ and $\sqrt{3}\mathbb{Q}\backslash\{0\}$ are disjoint.
   (c) Explain briefly why $\sqrt{2}\mathbb{Q}\backslash\{0\}$, $\sqrt{3}\mathbb{Q}\backslash\{0\}$ and $\mathbb{Q}$ are all pairwise disjoint.
   (d) Show that the boundary of either of the three sets is $\mathbb{R}$.
4. Continuing question 3, for $n \in \mathbb{N}$, $n > 3$, can n disjoint sets share the same boundary?

6.4 Compact Sets

Let's say you are asked to solve the following type of problem: For a given function $f : X \to X$ on a space X, find a value of x contained within the domain of $f(x)$ such that

$$x = f(x)$$

This is an example of a *fixed-point problem*. There are many methods of trying to find a solution to such a problem, but for now, let us fixate on a method involving sequences.

So let's start with a value $x_1 \in X$. If the range of f is contained within the domain of f, then we can take x_2 to be

$$x_2 = f(x_1)$$

And similarly, we continue with this process and define $x_{n+1} = f(x_n)$ for all $n \in \mathbb{N}$. At this point we have a sequence $\{x_n\}$ in X. If $\{x_n\}$ converges to a value l and f has a particular property[55], then we will have that

$$l = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n) = f(l)$$

and we have a solution to our original problem.

However, we will typically find that asking for $\{x_n\}$ to converge is usually asking too much. But not all is lost, as long as $\{x_n\}$ has a single convergent subsequence in X the problem will have a solution. Because of this, it can be very useful to know when a set X will have this property of guaranteeing the existence of convergent sequences and subsequences.

Example 30. *Let us look at some spaces X and determine if sequences on these spaces have or do not have convergent subsequences.*

a). $A = \{1, 2, \ldots, n\}$. As we saw in the section on the Bolzano-Weierstrass theorem, any sequence $\{x_n\}$ of terms coming from A must contain at least one convergent subsequence. This is because it is impossible for a sequence to take on an infinite distinct collection of values when restricted to a finite set.

b). $B = \mathbb{Z}$. The sequence $\{x_n\} = \{n\}$ made up of terms from B does not converge, and no subsequence of this sequence can converge either as $|x_m - x_n| \geq 1$ for all $n \neq m$.

c). $C = (0, 1)$. The sequence $\{x_n\} = \{\frac{1}{n}\}$ is made up of elements from C. As we have seen in previous sections, this sequence does converge in $\mathbb{R}$ and converges to the value 0, but this sequence does not converge in C as 0 does not exist within C. From our prior theorems we know that all subsequences of $\{x_n\}$ converge to 0, and from this, no subsequence of $\{x_n\}$ converges in C.

d). $D = [0, 1]$. Similar to part c). if we look at the sequence given by $\{x_n\} = \{\frac{1}{n}\}$ we have that this sequence converges to 0. In this case, as $0 \in D$ we have that $\{x_n\}$ converges in D. And every subsequence of $\{x_n\}$ converges in D as well because of this.

From what we see in our example, it seems like a). and d). have this property we are looking for. In example b). the fact that $\mathbb{Z}$ goes on forever means there is enough 'space' in the set for a sequence to take on distinct values without clustering anywhere. And with example c). we see that the sequence clusters, but as the set in question does not contain its limit points, the limit does not exist in this case. This likely gives you an idea of what properties a set needs to guarantee at least one subsequence from any given sequence converges. First, let us see some formal definitions, and in the next subsection we will see how many of these properties are actually equivalent.

[55] continuity on X

Definition 46. *For a set A in the reals, the set is called* **sequentially compact** *if every sequence $\{x_n\}$ of terms coming from A, $x_n \in A$ for all $n \in \mathbb{N}$, has a convergent subsequence with its limit in A.*

Definition 47. *For a set A in the reals, a collection $\mathcal{U}$ of open sets*

$$\mathcal{U} = \{U_\lambda \text{ open in } \mathbb{R} \mid \lambda \in \Lambda\}$$

is called an **open cover** *of A if it has the property that*

$$A \subseteq \bigcup_{\lambda \in \Lambda} U_\lambda$$

The set A is called **compact** *if any open cover $\mathcal{U}$ of A can be refined into a* **finite subcover**, *i.e. $\mathcal{U}' = \{U_1, U_2, \ldots, U_n\}$ a finite collection of open sets in $\mathbb{R}$ with*

$$A \subseteq \bigcup_{k=1}^{n} U_k$$

Ok, there are are few things to mention. First of all, why the two definitions, because often in math it is useful to have multiple ways to check if a set has a particular property, especially as some conditions may be easier to check than others in certain scenarios. However, my main reason for doing this in this instance, is that while we see in $\mathbb{R}$ that these two definitions coincide in the next section, in more abstract topological settings these definitions are actually distinct. As such, I wanted to point out the two definitions to avoid possible future confusion in studying math.

As a second quick remark, in the definition of compactness it is required to look at all possible open covers $\mathcal{U}$ of a given set A, where the elements $U \in \mathcal{U}$ are all open sets in $\mathbb{R}$. And such open covers could contain an absurdly large numbers of open sets, i.e. there was no restriction on the cardinality of the index set Λ. As it turns out, this is not possible. In $\mathbb{R}$ if we only look at open covers $\mathcal{U}$ containing a countable number of open sets, no generality is lost.[56]

From what we saw in our example above, we have that b). and c). are most definitely not sequentially compact, but we have the feeling that a). and d). are. Before we prove a collection of theorems that will verify this, let us see these examples again but from the perspective of open covers.

Example 31. *a). $A = \{1, 2, \ldots, n\}$. If $\mathcal{U}$ is an open cover of A, we have that by definition $j \in U$ for some $U \in \mathcal{U}$ for all $1 \leq j \leq n$. Let us label U_j as a member from $\mathcal{U}$ that contains j.*[57]*. Then $\{U_1, U_2, \ldots, U_n\}$ is a finite subcover of A inside of $\mathcal{U}$.*

b). $B = \mathbb{Z}$. Let $\mathcal{U} = \{(-n, n) \mid n \in \mathbb{N}\}$ be the collection of intervals of the form $(-n, n)$. We have that $\mathcal{U}$ is an open cover as

$$\mathbb{Z} \subseteq \bigcup_{n=1}^{\infty} (-n, n).$$

However, no finite subcover from $\mathcal{U}$ can cover $\mathbb{Z}$.

[56] For those that are curious, this ability to refine from an arbitrary open cover in $\mathbb{R}$ to a countable open cover is a property that comes from $\mathbb{R}$ being *Lindelöf* using more general topology language

[57] we are technically invoking choice here as there many be many members of $\mathcal{U}$ that contain j but we only need one

c). $C = (0,1)$. Let $\mathcal{U} = \{(\frac{1}{n}, 2) \mid n \in \mathbb{N}\}$. As

$$(0,1) \subseteq \bigcup_{n=1}^{\infty} \left(\frac{1}{n}, 2\right) = (0,2)$$

we have that $\mathcal{U}$ is an open cover of C. But any finite subcollection from $\mathbb{U}$ is of the form

$$\mathcal{U}' = \left\{ \left(\frac{1}{n_1}, 2\right), \left(\frac{1}{n_2}, 2\right), \ldots, \left(\frac{1}{n_m}, 2\right) \right\}$$

for some fixed $m \in \mathbb{N}$. If we take $N = \max(n_1, n_2, \ldots, n_m)$, then the union of all sets in this finite collection

$$\bigcup_{\mathcal{U}'} \left(\frac{1}{n_m}, 2\right) = \left(\frac{1}{N}, 2\right)$$

which will never cover $(0,1)$. Thus $\mathcal{U}$ is a cover of C that has no finite subcover.

d). $D = [0,1]$. The collection $\mathcal{U} = \{(\frac{1}{n}, 2) \mid n \in \mathbb{N}\}$ from example c). is not a cover of D as it never contains 0. If we fix this fact by appending the open interval $V = (-\frac{1}{2}, \frac{1}{2})$ to $\mathcal{U}$, i.e. making the new collection $\mathcal{U}' = \mathcal{U} \cup V$, then this is an open cover of D.

Furthermore, using the language from c). any finite subcollection from $\mathcal{U}'$ that contains V and has $N > 2$ will be a finite subcover of D.

Hopefully at this point, the similarity in results that we have seen in the prior two examples makes you believe that sequential compactness and compactness are identical in $\mathbb{R}$. We will soon see this is true (with a little more as well), but first let us prove two results that show some initial connections between closed sets and compact sets.

Theorem 68. *A compact subset, K, of $\mathbb{R}$ is closed.*

Proof. We will prove that the complement of K in $\mathbb{R}$ is open and appeal to a result in the previous section on closed sets. Thus let $p \notin K$. For any $q \in K$, define $\epsilon = \frac{|p-q|}{2}$ and the sets $U_p = (p-\epsilon, p+\epsilon)$ and $V_q = (q-\epsilon, q+\epsilon)$. We then immediately have that U_p, V_q are two open sets in $\mathbb{R}$ and from the choice of ϵ that $V_q \cap U_p = \varnothing$.

Now there is nothing special about q in this case, it is simply some random point in K. So while keeping $p \in K^c$ fixed, let us perform this process of creating U_p and V_q for every $q \in K$ and call

$$\mathcal{V} = \{V_q \mid q \in K\}$$

the collection of all the V_q. Clearly, by construction, $\mathcal{V}$ is an open cover of K, and thus by the assumption that K is compact we know there is a finite subcover $\mathcal{V}'$

$$\mathcal{V}' = \{V_{q_1}, V_{q_2}, \ldots, V_{q_N}\}$$

Now define the following

$$U = \bigcap_{n=1}^{N} U_{p_n}.$$

We have that $p \in U$ as $p \in U_{p_n}$ for all $1 \leq n \leq N$, and we also have that U is an open set as it is a finite intersection of open sets. Even more, as $U_{p_n} \cap V_{q_n} = \varnothing$ for all $1 \leq n \leq N$ we have that

$$U \cap \left[\bigcup_{n=1}^{N} V_{q_n}\right] = \varnothing$$

But $\mathcal{V}'$ is a cover of K, and as such

$$K \subseteq \bigcup_{n=1}^{N} V_{q_n}$$

so it must be that $U \cap K = \varnothing$. And so U is an open set with $p \in U \subseteq K^c$. As this can be done with any $p \in K^c$ we have that K^c is open and thus K is a closed set. □

So the theorem above says that a compact set in $\mathbb{R}$ must be closed. Does the converse hold: are all closed sets compact? No. As we saw with our earlier examples, $\mathbb{Z}$ is closed but most definitely not compact. We will soon see the property required on top of closedness to guarantee compactness.

Theorem 69. *If A is a closed set that is a subset of a compact set K, then A is compact.*

Proof. Assume that $\mathcal{U}$ is an arbitrary open cover of A. Our goal is to show why this has a finite subcover of A. As we know that A is closed in $\mathbb{R}$, we have that A^c is an open set. Because of this $\mathcal{U} \cup A^c$ is an open cover of $\mathbb{R}$ and thus an open cover of K.

Since K is compact, the cover $\mathcal{U} \cup A^c$ must contain a finite subcover $\mathcal{U}'$ of K.

- If $A^c \notin \mathcal{U}'$, then the collection $\mathcal{U}'$ is a finite subcover of A.
- If $A^c \in \mathcal{U}'$, then the collection $\mathcal{U}'' = \mathcal{U}' \backslash \{A^c\}$ is a finite subcover of A.

and in either case we see that $\mathcal{U}$ has a refinement to a finite subcover of A, and thus A is compact. □

Lecture 21 - 8/9/24

Ok, let us now prove our result that will help us characterize compact subsets of $\mathbb{R}$.

Theorem 70. *Let A be a subset of the reals, $\mathbb{R}$, then the following are equivalent.*

1). A is closed and bounded.

2). A is compact.

3). A is sequentially compact.

Remark 6. *The equivalence of 1). and 2). in the above theorem is often called the* **Heine-Borel** *theorem or property.*

Proof. **1). $\implies$ 2).** As A is bounded, there exists some $N \in \mathbb{N}$ such that

$$A \subseteq [-N, N]$$

The result will immediately follow from theorem 69 if we can show that $[-N, N]$ is compact. This will follow from a lemma whose proof is very reminiscent of the Bolzano-Weierstrass theorem.

Lemma 71. *For a fixed $n \in \mathbb{N}$, the interval $[-n, n]$ is compact.*

Proof. Call $I = [-n, n]$ and suppose to the contrary that there is an open cover $\mathcal{U}$ such that no finite subcollection of $\mathcal{U}$ will cover I. If we split I in half into $[-n, 0]$ and $[0, n]$, then it must be the case that at least one of these intervals can not be covered by a finite subcollection of $\mathcal{U}$. Pick one, if there is a choice and call this I_1.

Now split I_1 in half, and by assumption at least one of the halves can not be covered by any finite subcollection of $\mathcal{U}$. If there is a choice, pick one of these sets and call this I_2. Continuing on this process, we can create a sequence of intervals I_m with

$$I_1 \supseteq I_2 \supseteq I_3 \cdots$$

in which the length of interval I_m is $\frac{n}{2^{m-1}}$ and I_m can not be covered by any finite subcollection of $\mathcal{U}$. As n is fixed, we see that the lengths of these intervals go to 0 as $m \to \infty$. And by construction, each interval I_m is of the form $[a_m, b_m]$. From lemma 42, we have that

$$\bigcap_{m=1}^{\infty} I_m = \{p\}$$

As $\mathcal{U}$ is an open cover of $[-n, n]$ there must exist some $U \in \mathcal{U}$ such that $p \in U$. As U is open, there exists some $\epsilon > 0$ such that $(p - \epsilon, p + \epsilon) \subseteq U$. And as the lengths of I_m go to zero, there exists some $N \in \mathbb{N}$ such that $\frac{n}{2^{N-1}} < \epsilon$. But this implies that

$$p \in I_N \subseteq (p - \epsilon, p + \epsilon) \subseteq U$$

which shows that I_N can be covered by one element from $\mathcal{U}$ which is in direct contradiction of the fact that I_N was chosen in the selection process because it could not be covered by any finite subcollection of $\mathcal{U}$. This is a contradiction, thus it must be the case that any open cover of $[-n, n]$ has a finite subcover, and thus $[-n, n]$ is compact. □

Thus from our lemma, as A is a closed subset of a compact set $[-N, N]$, we have that A is compact.

2). $\implies$ 3). Assume that A is compact. Let $\{x_n\}$ be an arbitrary sequence coming from A. Define the set of terms from the sequence $\{x_n\}$.

$$B = \{x_n \mid n \in \mathbb{N}\}$$

We have two cases.

Case 1: B is a finite set in A. In this case

$$B = \{y_1, y_2, \ldots, y_N\}$$

and any sequence $\{x_n\}$ made up of terms from B must have that an infinite number of terms from $\{x_n\}$ equals y_j for at least one $1 \leq j \leq N$. Otherwise, the sequence $\{x_n\}$ would equal each y_j at most a finite number of times, and as there are a finite number of terms $y_1, y_2, \ldots, y_N$, this would mean a sequence only has a finite number of terms, which is a clear contradiction. Thus, $\{x_n\}$ contains at least one convergent subsequence with a limit in A, namely a constant sequence equal to $\{y_j\}$ for some $1 \leq j \leq N$.

Case 2: B is an infinite set, i.e. B contains an infinite number of distinct terms. We claim that A contains at least one limit point of B. By way of contradiction, assume this is not true. Thus every point $a \in A$ is not a limit point of B. But this means that for each $a \in A$, there is an $\epsilon > 0$ such that $(a - \epsilon, a + \epsilon) \cap B = \{a\}$. In other words, each $a \in A$ has a neighborhood V_a that meets the set B in the point a alone and nowhere else. But then

$$\mathcal{V} = \{V_a \mid a \in A\}$$

is an open cover of A, and by assumption that A is compact, this open cover has a finite subcover $\mathcal{V}' = \{V_{a_1}, V_{a_2}, \dots V_{a_m}\}$. But this would imply that

$$B = B \cap A \subseteq B \cap \left[\bigcup_{k=1}^{m} V_{a_k}\right] = \left[\bigcup_{k=1}^{m} B \cap V_{a_k}\right] = \{a_1, a_2, \dots, a_m\}$$

or that an infinite set is contained within a finite set, a clear contradiction. Thus, A contains at least one limit point of B. If we call $l \in A$ the limit point of B that exists in A, then by theorem 62 we have that $(l - \epsilon, l + \epsilon)$ contains an infinite number of points from B for any $\epsilon > 0$. Thus by theorem 17 l is a subsequential limit of $\{x_n\}$.

In either case we see that for a sequence of terms from A, $\{x_n\}$, the existence of a convergent subsequence with limit in A, and thus A is sequentially compact.

3). $\implies$ 1). Now suppose that A is sequentially compact. If A is not bounded, then A is not contained in the interval $[-n, n]$ for any $n \in \mathbb{N}$. As A is not contained in $[-1, 1]$ there is an element $x_1 \in A$ with $|x_1| > 1$. Similarly, as A is not contained in $[-2, 2]$ there is an element $x_2 \in A$ with $|x_2| > 2$. Continuing on, we can create a sequence $\{x_n\}$ of terms within A in which $|x_n| > n$ for each $n \in \mathbb{N}$. This sequence and any subsequence of this sequence diverges because of this property $|x_n| > n$, but this is in direct contradiction with the sequential compactness of A that states that $\{x_n\}$ should contain at least one convergent subsequence. Thus it must be that A is bounded.

Now take $p \in A'$. As p is a limit point of A, for every $\epsilon > 0$, the interval $(p - \epsilon, p + \epsilon)$ meets A in some point other than p. For $\epsilon = 1$, as

$$(p - 1, p + 1) \cap A \neq \varnothing$$

take x_1 to be an element of A in this intersection. Thus $x_1 \in A$ and $|p - x_1| < 1$. Similarly, for $\epsilon = \frac{1}{2}$, as

$$\left(p - \frac{1}{2}, p + \frac{1}{2}\right) \cap A \neq \varnothing$$

take x_2 to be an element in this intersection, and thus $x_2 \in A$ with $|p - x_2| < \frac{1}{2}$. Continuing on in this process, we can see that because p is a limit point of A that we can find a $x_n \in A$ with $|p - x_n| < \frac{1}{n}$ for all $n \in \mathbb{N}$.

Thus $\{x_n\}$ is a sequence of terms from A. By construction, we have that $\{x_n\} \to p$. However, by the sequential compactness of A, we know that there is a subsequence $\{x_{n_k}\}$ that converges to an element l in A. But from theorem 15, all subsequences of a convergent sequence must converge to the same limit, and so $p = l \in A$. As p was an arbitrary element of the derived set of A we have that $A' \subseteq A$ and thus A is closed. □

To pay a debt or keep a promise from an earlier section, we state the general or second version of the Bolzano-Weierstrass theorem.

Bolzano-Weierstrass Theorem: A bounded infinite set in $\mathbb{R}$ has at least one limit point.

The proof of this is actually contained within the proof above (case 2). $\implies$ 3).) with B being the infinite set in question and A being a compact set of the form $[-n, n]$ that contains B as it is bounded.

Now as I mentioned in the section on the Bolzano-Weierstrass theorem for sequences, the notion of compactness generalizes finite sets. In that section we saw how compactness extended

the pidgeonhole principle for infinite but bounded sets (bounded sequences). We will see this again when we look at images of compact sets under continuous functions, but for now we can phrase this in terms of covers. Any open cover of a finite set $A = \{a, b, \ldots, z\}$ clearly has a finite subcover as at maximum you only need to choose one element of the cover for every element of the set and you are done. Compact sets are 'almost' finite in this sense, as even though they may contain an infinite number of elements, they can always be covered by a finite subcollection of any initial covering collection.

To close out this section, let us see two last theorems about compact sets that can be very useful. The first states that a collection of compact sets that have the 'finite intersection property' will have nonempty intersection overall, and the second states that compact subsets of $\mathbb{R}$ are complete in there own right.

Theorem 72. *Given an arbitrary collection $\mathcal{K} = \{K_\lambda\}$ of compact sets in $\mathbb{R}$, if every finite subcollection $\mathcal{K}' = \{K_1, K_2, \ldots, K_n\}$ has non-empty intersection $K_1 \cap K_2 \cap \cdots K_n \neq \varnothing$, then*

$$\bigcap_\lambda K_\lambda \neq \varnothing$$

Proof. By way of contradiction, assume that every finite subcollection has non-empty intersection but the arbitrary intersection of all compact sets in $\mathcal{K}$ is empty. Define the collection $\mathcal{U} = \{K_\lambda^c\}$, then as compact sets within $\mathbb{R}$ are closed, we have that $\mathcal{U}$ is an arbitrary collection of open sets.

Fix K_1 for the moment. We claim that $\mathcal{U}$ is an open cover of K_1. If this were not the case then there would exist $x \in K_1$ with

$$x \notin \bigcup_\lambda K_\lambda^c \iff x \in \bigcap_\lambda K_\lambda = \varnothing$$

and as we see above by the DeMorgan Laws and our initial assumption, this is impossible. So it must be that $\mathcal{U}$ is an open cover of K_1. As K_1 is compact, there is a finite subcover $\mathcal{U}' = \{K_{\lambda_1}^c, K_{\lambda_2}^c, \ldots, K_{\lambda_n}^c\}$ of K_1,

$$K_1 \subseteq \bigcup_{k=1}^n K_{\lambda_k}^c \iff \bigcap_{k=1}^n K_{\lambda_k} = \left[\bigcup_{k=1}^n K_{\lambda_k}^c\right]^c \subseteq K_1^c$$

and this would imply that

$$K_1 \cap K_{\lambda_1} \cap K_{\lambda_2} \cap \cdots K_{\lambda_n} = \varnothing$$

and this is in direct contradiction of the fact that the intersection of any finite subcollection from $\mathcal{K}$ is non-empty. □

Theorem 73. *A compact subset K of $\mathbb{R}$ is complete.*

Proof. We only need to show that K is Cauchy complete as supremums and infimums of bounded sets within K can be approximated by sequences within K. So, let $\{x_n\}$ be a Cauchy sequence in K. As K is sequentially compact, we have that $\{x_n\}$ contains a convergent subsequence, i.e. $\{x_{n_k}\} \to l$ with $l \in K$. From theorem 20 we have that $\{x_n\}$ converges to l, and so K is Cauchy complete. □

Exercises for section 6.4:

1. (a) Show that an arbitrary intersection of compact sets is compact.
   (b) Show that a finite union of compact sets is compact.
2. For two non-empty sets A, B, define
$$A + B := \{a + b : \ a \in A, \ b \in B\}.$$
   (a) Show that if A is open, then $A + B$ is open.
   (b) Show that if A and B are compact, then $A + B$ is compact.
   (c) Given an example where A and B are closed but $A + B$ is not.
3. Give an example of a non-compact set A such that both $\sup A$ and $\inf A$ belong to A.
4. (a) Show that if the terms of a sequence $\{x_n\}$ are seen as a set B, and this set has two distinct limit points, then $\{x_n\}$ diverges.
   (b) Show that if $\{x_n\}$ is a sequence inside of a compact set A and $\{x_n\}$ has a single limit point, then $\{x_n\}$ converges.

6.5 Perfect Sets (Optional)

Perfect Sets

Closed set where every point is a limit point of E.

Theorem P nonempty and perfect then P is uncountable.

Cantor Set example.

Def of a condensation point, difference from accumulation point.

Baby cantor bendixson theorem.

Exercises for section 6.5:

Lecture 22 - 8/12/24

6.6 Connected Sets

There are many definitions of what it means for a set to be connected. At its core, the main ideas of the definition is that a 'connected' set should not be able to be broken into two pieces without boundary. As we have seen that open sets do not contain their boundary, they will play a role in one form of the definition.

Definition 48. *The following are four equivalent definitions of what it means for a set to be* **connected**.

1. A space X is called **disconnected** *if there exists non-empty open sets U, V that are1 disjoint $U \cap V = \varnothing$ that make up the whole space, i.e. $X = U \cup V$. A space X is called* **connected** *if it is not disconnected.*

2. Two sets A and B are called **separated** *if*
$$A \cap \overline{B} = \overline{A} \cap B = \varnothing$$

and X is **connected** *if it is not a union of two non-empty separated sets.*

3. *X is* **connected** *if the only clopen (closed and open sets) in X are X and $\varnothing$.*
4. *X is* **connected** *if the only sets with empty boundary are X and $\varnothing$.*

At the end of this section there will be a proposition showing that these definitions are indeed equivalent. For now, let us start with a theorem that shows the character of connected sets in $\mathbb{R}$.

Theorem 74. *A subset $E \subseteq \mathbb{R}$ is connected if and only if the following statement holds: If $x, y \in E$, then $x < z < y$ implies that $z \in E$.*

Proof. $\implies$ We argue by contrapositive. Suppose that we have $x, y \in \mathbb{R}$ with $x < y$, and suppose there is a z with $x < z < y$ but $z \notin E$. Now define A and B in the following way,

$$A = (-\infty, z) \cap E, \qquad B = (z, \infty) \cap E$$

then we clearly have that $E = A \cup B$. Now, we come to a lemma

Lemma 75. *For two sets A, B in the reals, we have that*

$$\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$$

Proof. As $\overline{A}$ and $\overline{B}$ are closed sets, their intersection $\overline{A} \cap \overline{B}$ is closed as the intersection of two closed sets is closed. From Proposition 66, as

$$A \cap B \subseteq \overline{A} \cap \overline{B}$$

we automatically have that $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$. □

From our lemma we have that

$$\overline{A} = \overline{(-\infty, z) \cap E} \subseteq \overline{(-\infty, z)} \cap \overline{E} = (-\infty, z] \cap \overline{E}$$

and thus we have that

$$\overline{A} \cap B \subseteq (-\infty, z] \cap \overline{E} \cap (z, \infty) \cap E = \varnothing$$

thus $\overline{A} \cap B = \varnothing$ and similarly one can see that $A \cap \overline{B} = \varnothing$. As $x \in A$ and $y \in B$ we have that A and B are non-empty sets. Thus A and B are separated sets that make up E and so E is not connected, and one direction of our proof is complete.

$\impliedby$ Now let us assume that E is not connected. Thus assume that $E = A \cup B$ where A and B are separated sets. As A and B are assumed to be non-empty, take $x \in A$ and $y \in B$, and without loss of generality assume that $x < y$ (otherwise we could just relabel the sets). Define the value z by

$$z = \sup(A \cap [x, y])$$

which exists as this set is nonempty and bounded above. From theorem 67, we have that supremums of sets are contained within closed sets. Thus it must be that $z \in \overline{A}$, and by the assumption that A and B are separated we have that $z \notin B$. Thus $z \neq y$ and so we have that

$$x \leq z < y$$

From here we have two possible cases:

- *Case* 1: If $z \notin A$, then as $E = A \cup B$, we have that $z \notin E$ and as $z \neq x$

$$x < z < y.$$

- *Case* 2: If $z \in A$, then as A and B are seperated, we have that $z \notin \overline{B}$. Thus $z \in (\overline{B})^c$, which is an open set as compliments of closed sets are open. Thus, this contains a neighborhood of z, i.e. there exists an $\epsilon > 0$ such that $(z - \epsilon, z + \epsilon) \subseteq (\overline{B})^c$. Take $z_1 \in (z, z + \epsilon) \subseteq \mathbb{R}$. Then $z_1 \notin \overline{B}$ so $z_1 \notin B$. By possibly choosing a smaller ϵ we can assume that $z_1 < y$, and $z_1 \notin A$ otherwise this would contradict the definition of z as a supremum, thus $z_1 \notin E$ and

$$x < z_1 < y$$

In either case we have shown the existence of an element between x and y that is not contained in E, so this direction is proven by contrapositive. □

This theorem immediately tells us that both closed and open intervals $[a, b]$ and (a, b) respectively are connected sets within $\mathbb{R}$. From here one can deduce that $\mathbb{R}$ itself is connected as $\mathbb{R} = (-\infty, \infty)$. However, $\pm\infty$ are technically not real numbers as we avioided infinites within $\mathbb{R}$ by having the Archimedean property, thus the theorem above does not apply to the reals. For those of you that are sticklers about detail[58], we will present a second theorem confirming the connectedness of $\mathbb{R}$. This can be proven in another way that will be left as a homework exercise.

Theorem 76. $\mathbb{R}$ *is connected.*

Proof. By way of contradiction, assume that $\mathbb{R}$ is not connected. Thus there exists open sets A, B that are not empty, with $A \cap B = \varnothing$ and $\mathbb{R} = A \cup B$.

As A and B are not empty, assume that $a \in A$ and $b \in B$. Without loss of generality, assume that $a < b$ otherwise, simply rename the sets and elements. Now define the the following set

$$X = \{x \in [a, b] \mid [a, x] \subseteq A\}$$

Now X is a nonempty set as $a \in X$. For any $y \in X$ we must have $y < b$ as $y \geq b$ implies that

$$[a, b] \subseteq [a, y] \subseteq X$$

and this is a clear contradiction as $b \notin A$. Thus b is an upper bound of X and thus X is bounded above. Therefore the supremum of X exists within $\mathbb{R}$, call it α.

If $\alpha \in A$, then as A is open there exists some $\epsilon > 0$ such that $(\alpha - \epsilon, \alpha + \epsilon) \subseteq A$, but then we would have

$$\left[a, \alpha + \frac{\epsilon}{2}\right] \subseteq A$$

and thus $\alpha + \frac{\epsilon}{2} \in X$ and this contradicts the definition of $\alpha = \sup X$. Thus it must be that $\alpha \notin A$.

Well as $A \cap B = \varnothing$ and $\mathbb{R} = A \cup B$, it must be that $\alpha \in B$. As B is open, there exists some $\epsilon > 0$ such that $(\alpha - \epsilon, \alpha + \epsilon) \subseteq B$. As $\alpha = \sup X$, for this ϵ there exists $y \in X$ with

$$\alpha - \epsilon < y < \alpha$$

[58] which who am I kidding, I am one

and thus $[a, y] \subseteq A$. Take $z = \frac{\alpha - \epsilon + y}{2}$, then

$$\alpha - \epsilon < z < y < \alpha < \alpha + \epsilon$$

Then $z \in [a, y] \subseteq A$ and $z \in (\alpha - \epsilon, \alpha + \epsilon) \subseteq B$ and thus

$$z \in A \cap B = \varnothing$$

and this is a contradiction. Thus we must have that $\alpha \notin B$ as well, so

$$\alpha \in A^c \cap B^c \iff \alpha \in (A \cup B)^c = (\mathbb{R})^c = \varnothing$$

And this is a clear contradiction, thus it must be that $\mathbb{R}$ is connected. □

Before we prove our propositions showing that the different definitions of connectedness are actually equivalent, we have one last definition. The following is not necessary for the proof, but it is a topic that is related to connected sets. In $\mathbb{R}$ the defintion is fairly boring, but this definition will come up in a more general context in the study of sets in $\mathbb{R}^n$.

Definition 49. *A subset E of $\mathbb{R}$ is called* **convex** *if for any two points $a, b \in E$ we have*

$$(1 - t)a + tb \in E \quad \forall t \in [0, 1]$$

Equivalently, E is convex if for $a, b \in E$ with $a < b$ we have that $c \in E$ for any $a < c < b$.

From theorem 74 we immediately see that a set is convex in $\mathbb{R}$ if and only if it is connected. This is only true in dimension one.[59] Even further, theorem 74 shows that the only convex sets in $\mathbb{R}$ are singleton sets $\{a\}$ and intervals (both open and closed). We lastly present a promised proof of the proposition from the start of the section.

Proposition 77. *The definitions of connectedness as the start of the section are equivalent.*

Proof. We will prove this in a cycle.

1) $\implies$ 2) Let us assume that X is not connected and show the existence of two non-empty separated sets. As X is not connected there are nonempty open sets U and V such that $U \cap V = \varnothing$ and $X = U \cup V$.

As $U \cup V = X$ and $U \cap V = \varnothing$, we have that $U^c = V$, and thus V is also a closed set as it is the complement of an open set, and similarly U is a closed set as well. Thus $U = \overline{U}$ and $V = \overline{V}$, thus

$$U \cap \overline{V} = \overline{U} \cap V = U \cap V = \varnothing$$

Thus U and V are two nonempty separated sets whose union is X.

2) $\implies$ 3) Assume definition 2 and we will show that X is disconnected if there is a clopen set A with $A \neq X, \varnothing$. Define $U = A$ and $V = A^c$, then

$$X = U \cup V$$

[59] think of the annulus in $\mathbb{R}^2$

and U is not empty and from the assumption that $A \neq X$, we have that V is not empty. We have that U is both closed and open as A is, and similarly V is both closed and open as compliments of open sets are closed and vice-versa. Thus $\overline{U} = U$ and $\overline{V} = V$, and we have

$$\overline{U} \cap V = U \cap \overline{V} = U \cap V = A \cap A^c = \varnothing$$

and so U and V are non-empty separated sets. Thus X is not connected. Thus by the contrapositive, X will be connected if X has no clopen sets besides X and $\varnothing$.

3) $\implies$ 4) For any set E in $\mathbb{R}$ we have that

$$E^\circ \subseteq E \subseteq \overline{E}$$

as the interior of a set is the largest open set contained in E and the closure of E is the smallest closed set containing E. If the boundary of E is empty then

$$\varnothing = \partial E = \overline{E} \backslash E^{circ}$$

and this can only be if $E^\circ \supseteq \overline{E}$ and that would imply

$$E = E^\circ = \overline{E}$$

and thus E would be both closed and open. Thus by definition 3, the only sets with this property are X and $\varnothing$.

4) $\implies$ 1) Suppose that $X = A \cup B$ with $A \cap B = \varnothing$ and A and B are open. Then from this we immediately have that $A^c = B$ and $B^c = A$ and thus A and B are also both closed sets. Thus $A = A^\circ = \overline{A}$ and similarly for B. But then we have

$$\partial A = \partial B = \varnothing$$

If X is connected, then from definition 4 it must be that $A = X, \varnothing$ and $B = \varnothing, X$ respectively (as $B = A^c$). Without loss of generality $A = X$ and $B = \varnothing$. As this holds for any two disjoint open sets with $A \cup B = X$ we have that it is impossible to find two non-empty open sets with these properties. Thus X is connected. □

Just as one last comment. We saw that in $\mathbb{R}$ that the connected (convex) sets were singleton sets $\{a\}$ and intervals. There are examples of spaces that are the 'least' connected in a sense that the only connected subsets they have are the singleton sets, or put another way, that these spaces contain no interval.

Definition 50. *A space X is called* **totally disconnected** *if the only connected subsets of this space are the singleton sets $\{a\}$ for $a \in X$.*

We have seen a few examples of totally disconnected subsets of $\mathbb{R}$ in this class, in particular the rational numbers $\mathbb{Q}$, the irrational numbers, the algebraic numbers $\mathbb{A}$ (minus i), and the Cantor set $\mathfrak{C}$.

Exercises for section 6.6:

1. The goal of this problem is to show that $\mathbb{R}$ is connected in a different way.

(a) For $A, B \subseteq \mathbb{R}$ two connected subsets. Show that if $A \cap B \neq \varnothing$ then $A \cup B$ is connected.

(b) Proceed by induction. Prove that if $\{A_1, A_2, \dots\}$ is a collection of connected subsets in $\mathbb{R}$ and that if $A_n \cap A_{n+1} \neq \varnothing$, then

$$\bigcup_{n=1}^{\infty} A_n \text{ is connected}$$

(c) Show why $\mathbb{R}$ is connected using the above.

2. Give a proof as to why $\mathbb{Q}$ and $\mathbb{A}\backslash\{i\}$ are totally disconnected subsets of $\mathbb{R}$.

Lecture 23 - 8/14/24

7 Continuity

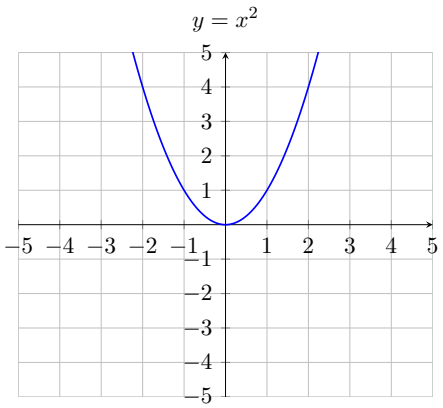
7.1 Limits of Functions

Before we define a limit for a function at a particular point, let us recall the following definitions about functions. We often write a function as $f : A \to B$ with a rule $f(x)$. The set A is the domain of f, sometimes called $\text{dom}(f)$. The set B is the codomain of f and is where the values of f map into. The rule $f(x)$ tells us how f acts on an element x, in particular $x \in A$ and $f(x) \in B$. The range of a function f, written $\text{ran}(f)$ is given by

$$\text{ran}(f) = \{y \in B \mid \exists x \in A \text{ with } y = f(x)\} = f(A)$$

where the last equality is simply how one can write the range of a function as the image of its domain.[60]

Example 32. *The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ has $dom(f) = codom(f) = \mathbb{R}$, and $ran(f) = [0, \infty)$. This can be seen in the following graph.*



As seen in the above example, when a function is defined $f : A \to B$ where A, B are both subsets of $\mathbb{R}$, then we typically view the graph of f as a subset of $\mathbb{R}^2$.

$$\text{Gr}(f) = \{(x, f(x)) \mid x \in A\}$$

This will be useful in this section as we will be focusing on functions from the reals to itself.

Often for a real–valued function $f : \mathbb{R} \to \mathbb{R}$, the domain of f will not be the entirety of the real numbers, i.e. you will need to 'pare down' $\mathbb{R}$ into a collection of values that can actually be input into f. Generally the two rules that need to be followed are:

1. If x is such that the denominator of f equals 0 when x is input, then $x \notin \text{dom}(f)$.
2. If the rule for f contains an even radical (square root, fourth root, etc), then if inputting x makes the expression under such a radical negative, then $x \notin \text{dom}(f)$.[61]

[60] when extending f to its definition on power sets, i.e. $f : P(A) \to P(B)$

[61] this is to ensure that f is real–valued

And this can help you determine $\text{dom}(f)$, which put another way, is the largest set of allowable inputs into f.

Definition 51. *Given two functions f, g, what we mean when we say f* **equals g as functions**, *written $f = g$, is that $dom(f) = dom(g)$ and $f(x) = g(x)$ for all $x \in dom(f)$.*

Example 33. *Given the functions $f(x) = x - 5$ and $g(x) = \frac{x^2-25}{x+5}$. It is true for $x \neq -5$ that*

$$g(x) = \frac{x^2 - 25}{x + 5} = \frac{(x-5)(x+5)}{x+5} = x - 5 = f(x)$$

but $f \neq g$ as functions as $dom(f) = \mathbb{R}$ but $dom(g) = (-\infty, -5) \cup (-5, \infty)$.

Lastly, we have the following pointwise operations on functions f and g that can help us define new functions.

- The sum $f + g$ is defined by
$$(f + g)(x) = f(x) + g(x)$$
- The difference $f - g$ is defined by
$$(f - g)(x) = f(x) - g(x)$$
- The product $f \cdot g$ is defined by
$$(f \cdot g)(x) = f(x) \cdot g(x)$$
- The quotient $\frac{f}{g}$ is defined by
$$\frac{f}{g}(x) = \frac{f(x)}{g(x)}$$

and we have that

$$\text{dom}(f + g) = \text{dom}(f - g) = \text{dom}(f \cdot g) = \text{dom}(f) \cap \text{dom}(g)$$
$$\text{dom}\left(\frac{f}{g}\right) = \text{dom}(f) \cap \text{dom}(g) \cap \{x \mid g(x) \neq 0\}$$

The last operation we have between two functions $f : A \to B$ and $g : B \to C$ is *function composition* defined by

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

which can be thought of performing multiple functions in a sequence or building a larger machine from two smaller machines. Is important to note that it is imperative in the definitin that the codomain of f is a subset of the domain of g. In particular, with the example given here $g \circ f$ is defined but $f \circ g$ would only be defined if $C \subseteq A$.

To make a note about where we have been and where we are going: the value of a function f at a point x, $f(x)$ is often what we in the math world called a *pointwise operation* as its value is only dependent on the point x. The notion of the limit of a function is different. A limit of f at a point a is a *local operation* as the value of the limit of f at a (if it exists) depends on the behavior of f on intervals containing a.[62]

[62] and not the value of f at a itself

Definition 52. *For an interval I, a function $f : I \to \mathbb{R}$ has a* **limit at** *$a \in I^\circ$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for all $x \in I$ satisfying the property $0 < |x - a| < \delta$ implies that $|f(x) - l| < \epsilon$. In this case we write this in the compact form*

$$\lim_{x \to a} f(x) = l$$

Remark 7. *The definition I've given above is actually not as general as it could be. This is simply a form that will be useful for us in this course as most of our examples will be functions with connected domains within $\mathbb{R}$.*

The general definition of the limit of a function f at a point a only requires that a is a limit point of the space it is contained within (usually the domain of f). In this sense, I lied to you earlier when I said Calculus was not possible on sets with gaps like $\mathbb{Q}$. But while it is generally true that limits can be defined in these cases, they will usually be the exception of what we consider and not the rule.

Example 34. *Let us compute the following limits.*

a). $\lim_{x \to 1} 2x + 1$. *We conjecture that the limit is 3, so as a bit of scratchwork we first compute*

$$|2x + 1 - 3| = |2x - 2| = 2|x - 1|$$

and so we can see that if $|x - 1| < \frac{\epsilon}{2}$ then we can guarantee the expression above is less than ϵ. Thus for $\epsilon > 0$, take $\delta = \frac{\epsilon}{2}$ and we see that

$$|x - 1| < \delta \implies |2x + 1 - 3| < \epsilon$$

and as this can be done for every $\epsilon > 0$ we have that

$$\lim_{x \to 1} 2x + 1 = 3$$

b). $\lim_{x \to 3} x^2$. *We conjecture that the limit is 9, so let us look at the following*

$$|x^2 - 9| = |(x - 3)(x + 3)| = |x - 3||x + 3|$$

If we can guarantee that $\delta < 1$, then we would have

$$|x - 3| < \delta < 1 \implies -1 < x - 3 < 1 \implies 5 < x + 3 < 7$$

and so $|x + 3| < 7$. And so let us take $\delta = \min\left(1, \frac{\epsilon}{7}\right)$. Then as $\delta < 1$ we have

$$|x^2 - 9| = |(x - 3)(x + 3)| = |x - 3||x + 3| < 7|x - 3|$$

and so if $|x - 3| < \delta$, then we have

$$|x^2 - 9| < 7|x - 3| < 7\delta < 7\left(\frac{\epsilon}{7}\right) = \epsilon$$

As this δ can be constructed for any $\epsilon > 0$ we have that

$$\lim_{x \to 3} x^2 = 9$$

c). $\lim_{x \to 0} \sqrt{x}$. *We conjecture that the result is 0, from the following*

$$|\sqrt{x} - 0|^2 = (\sqrt{x})^2 = x = x - 0$$

we see that $|\sqrt{x}-0| = \sqrt{|x-0|}$. *If we take* $\delta = \epsilon^2$, *then* $|x-0| < \delta$ *imples that*

$$|\sqrt{x}-0| = \sqrt{|x-0|} < \sqrt{d} = \epsilon$$

and as this can be done for any $\epsilon > 0$ *we have that*

$$\lim_{x\to 0} \sqrt{x} = 0$$

d). $\lim_{x\to 0} x^2$. *We conjecture that this limit goes to 0, so we look at*

$$|x^2 - 0| = |x^2| = |x|^2$$

And so if we take $\delta = \sqrt{\epsilon}$, *then we see that*

$$0 < |x-0| < \sqrt{\epsilon} \implies |x^2-0| = |x|^2 \leq (\sqrt{\epsilon})^2 = \epsilon$$

and as this can be done for any $\epsilon > 0$ *we have that*

$$\lim_{x\to 0} x^2 = 0$$

e). $\lim_{x\to 1} f(x)$ *for the function given by*

$$f(x) = \begin{cases} x^2 - 1, & x \neq 1 \\ 1, & x = 1 \end{cases}$$

We conjecture that the value of this limit is 0. *For any* $x \neq 1$, *we have*

$$|x^2 - 1 - 0| = |x^2 - 1| = |x-1||x+1|$$

If we can guarantee that $\delta < 2$, *then* $|x-1| < 2$ *gives the following*

$$|x-1| < 2 \implies -2 < x - 1 < 2 \implies 0 < x + 1 < 4$$

and thus we would have $|x+1| < 4$. *Thus taking* $\delta = \min\left(2, \frac{\epsilon}{4}\right)$, *then the assumption* $|x-1| < \delta$ *implies that* $|x+1| < 4$, *thus*

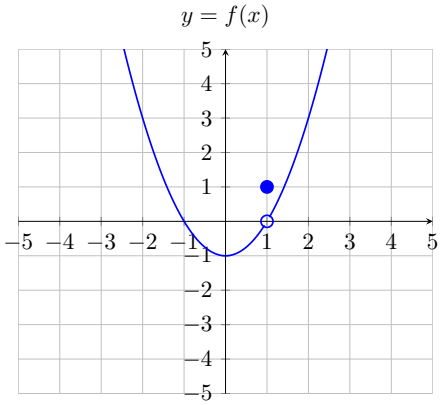
$$|x^2 - 1 - 0| = |x^2 - 1| = |x-1||x+1| < 4|x-1| < 4\delta < 4\left(\frac{\epsilon}{4}\right) = \epsilon$$

As we are avoiding $x = 1$ *we have just shown that*

$$0 < |x-1| < \delta \implies |f(x) - 0| < \epsilon$$

and as this holds for any $\epsilon > 0$ *we have that*

$$\lim_{x\to 1} f(x) = 0$$



In examples c). and d). one can see how the character of the function f may effect the convergence rates of the limit. In particular in example c). we saw that $\delta = \epsilon^2$ and thus we required a tighter interval about 0 in the x-axis (input space) to guarantee that $f(x) = \sqrt{x}$ was within ϵ of 0 in the y-axis (output space). Because of this we say $\sqrt{x}$ converges to 0 as $x \to 0$ *slower than linearly.* In example d). the choice of $\delta = \sqrt{\epsilon}$ showed that we could take a 'wider' interval in the input space to guarantee $f(x) = x^2$ was within ϵ in the output space. In this sense we say x^2 converges to 0 as $x \to 0$ *faster than linearly.*

Example e). clarifies what is meant by saying that a limit is a local operation. For example, the value of $f(x)$ at $x = 1$ has no bearing on the answer to $\lim_{x\to 1} f(x)$. If we redefined $f(x) = 507$ at $x = 1$ but left the rest of the definition the same, we would see that the value of the limit remains the same. Thus the value of the limit depended on the values of f around $x = 1$ but not at $x = 1$.

Theorem 78. *For a function $f : I \to \mathbb{R}$ with I an interval and $a \in I^\circ$ we have that $\lim_{x\to a} f(x) = l$ if and only if for every sequence $\{x_n\} \in I\backslash\{a\}$ with $\{x_n\} \to a$ implies that $\{f(x_n)\} \to l$.*

Proof. $\implies$ Assume that $\lim_{x\to a} f(x) = l$, and let $\epsilon > 0$. By the definition of the limit of a function, for this ϵ there exists a δ such that

$$0 < |x - a| < \delta, \implies |f(x) - l| < \epsilon$$

As $\{x_n\} \to a$, there exists $N \in \mathbb{N}$ such that $|x_n - a| < \delta$ for all $n > N$. Thus we also have that

$$|f(x_n) - l| < \epsilon, \quad \text{for all } n > N$$

As this can be done for every $\epsilon > 0$ we have that $\{f(x_n)\} \to l$.

$\impliedby$ We proceed by contrapositive, thus assume that $\lim_{x\to a} f(x) \neq l$. Thus by negating quantifiers, we have that

$$\exists \epsilon > 0,\ \forall \delta > 0,\ \exists x, \text{ such that } 0 < |x - a| < \delta \text{ and } |f(x) - l| \geq \epsilon$$

For this specific ϵ we will construct a sequence by making use of $\delta = \frac{1}{n}$ for $n \in \mathbb{N}$. For $\delta = 1$, the above gives the existence of an x_1 with

$$0 < |x_1 - a| < 1 \ \text{ and } \ |f(x_1) - l| \geq \epsilon$$

Similarly for each $n \in \mathbb{N}$ the limit of f as $x \to a$ not converging to l gives the existence of an x_n such that

$$0 < |x_n - a| < \frac{1}{n} \ \text{ and } \ |f(x_n) - l| \geq \epsilon$$

And thus we have created a sequence $\{x_n\} \in I\backslash\{a\}$ with $\{x_n\} \to a$ but $\{f(x_n)\}$ does not converge to l. Thus our result is proven by contrapositive. □

Another use of this theorem is that similar to our result about distinct subsequential limits implying a sequence diverges, if one wishes to show that a function $f(x)$ does not have a limit at a, you only need to find two sequences $\{x_n\}$, $\{y_n\}$ with $\{x_n\} \to a$ such that $\{y_n\} \to a$ and $\{f(x_n)\}$ and $\{f(y_n)\}$ converge to different values.

Example 35. *Consider the following*

a). Let us look at the characteristic function of the rationals $\chi_{\mathbb{Q}}(x)$ given by

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \\ 0, & \text{if } x \in \mathbb{R}\backslash\mathbb{Q} \end{cases}$$

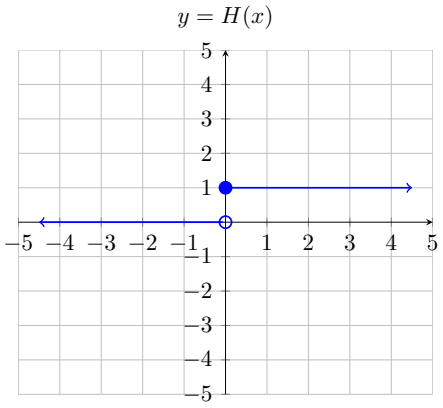
and look at the limit as $x \to 0$ of $\chi_{\mathbb{Q}}(x)$.

If we take $\{x_n\}$ to be a sequence of rational numbers tending to 0, we have that $\{f(x_n)\} = \{1\}$ is the constant sequence 1, and thus converges to 1. And if $\{y_n\}$ is a sequence of irrational numbers tending to 0, then $\{f(y_n\} = \{0\}$ is the constant sequence 0, and thus converges to 0. Because of this we have that the limit of $\chi_{\mathbb{Q}}(x)$ as $x \to 0$ does not exist.

b). Let us look at the Heaviside step function $H(x)$ given by

$$H(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

given by



If we take $\{x_n\} = \left\{\frac{1}{n}\right\}$, then $\{x_n\} \to 0$ and $H(x_n) = 1$ for all $n \in \mathbb{N}$, so $\{H(x_n)\} \to 1$. If we take $\{y_n\} = \left\{-\frac{1}{n}\right\}$, then $\{y_n\} \to 0$ and $H(y_n) = 0$ for all $n \in \mathbb{N}$, so $\{H(y_n)\} \to 0$. And because of this we see that the limit of $H(x)$ as $x \to 0$ does not exist.

In example a). the limit did not exist at 0 as any interval containing 0 contains an infinite number of rational and irrational numbers due to density and thus $\chi_{\mathbb{Q}}(x)$ is alternating between 0 and 1 for an infinite number of values. But in example b). all values to the right of $x = 0$ have $H(x) = 1$ and all values to the left have $H(x) = 0$. The limit does not exist in this case as any neighborhood of 0 contains values to the right and left of 0 but this does suggest the following definition(s).

Definition 53. *For a function $f : I \to \mathbb{R}$ and $a \in I$ we have the following:*

- *We say f has a* **right-handed limit at** *$a \in I$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that $0 < x - a < \delta$ implies that $|f(x) - l| < \epsilon$. In this case we write this in the compact form*

$$f(a+) = \lim_{x \to a^+} f(x) = l$$

- *We say f has a* **left-handed limit at** *$a \in I$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that $0 < a - x < \delta$ implies that $|f(x) - l| < \epsilon$. In this case we write this in the compact form*

$$f(a-) = \lim_{x \to a^-} f(x) = l$$

Do note that for $a \in I$ with $a \in \partial I$ that only of these limits may exist, i.e. at a point of the boundary of I, the interval I may only contain values to the left or right of a.

Now that we have this definition we come to a theorem that is likely not surprising for any student who remembers their first semester Calculus.

Theorem 79. *A function $f : I \to \mathbb{R}$ has a limit at $a \in I^\circ$ if and only if both left and right-handed limits exist at a and $f(a+) = f(a-)$.*

Proof. $\implies$ If the limit of f exists at a, then both the left and right-handed limits of f exist as $0 < x - a < \delta$ and $0 < a - x < \delta$ are both implies by the condition $0 < |x - a| < \delta$. It is also easy to see why $f(a+) = f(a-)$ in this case as well.

$\impliedby$ If we have that both the left and right-handed limits of f exist at a and $f(a+) = f(a-) = l$, then for $\epsilon > 0$ we know there exists a $\delta_1 > 0$ such that

$$0 < x - a < \delta_1 \implies |f(x) - l| < \epsilon$$

and a $\delta_2 > 0$ such that

$$0 < a - x < \delta_2 \implies |f(x) - l| < \epsilon$$

Thus if we take $\delta = \min(\delta_1, \delta_2)$, then we have that

$$0 < |x - a| < \delta \implies |f(x) - l| < \epsilon$$

and as this can be done for any $\epsilon > 0$ we have that $\lim_{x \to a} f(x) = l$ and thus the limit exists. □

Due to theorem 78 which gives us another way to characterize limits at a point, a, we immediately receive the following from our pointwise definition of sums, differences, products, and quotients of functions.

Theorem 80. *For a functions $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ and a point $a \in I^\circ$, if we have*

$$\lim_{x \to a} f(x) = L, \qquad \lim_{x \to a} g(x) = M$$

then

$$\lim_{x \to a} (f \pm g)(x) = L \pm M$$
$$\lim_{x \to a} (f \cdot g)(x) = L \cdot M$$

and provided that $M \neq 0$, we have

$$\lim_{x \to a} \left(\frac{f}{g} \right)(x) = \frac{L}{M}$$

Proof. The proof of this result falls out quickly when we pair theorem 78 in this section with theorem 14 from our section on sequences. By theorem 78, as we know the limit of f and g both exist at a, we have that for any sequence $\{x_n\}$ contained in $I\backslash\{a\}$ with $\{x_n\} \to a$ that

$$\{f(x_n)\} \to L, \qquad \{g(x_n)\} \to M$$

and thus we have

$$\{(f+g)(x_n)\} = \{f(x_n) + g(x_n)\} \to L + M$$

by the algebraic limit rules. As the sequence $\{x_n\}$ was arbitrary, we have that this result holds for any sequence with the same properties, thus by theorem 78 we have that

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

The proof for $f - g$, $f \cdot g$ and $\frac{f}{g}$ follows via a similar argument. □

Exercises for section 7.1:

1. For a function $g : I \to \mathbb{R}$ and $a \in I^\circ$ and

   $$\lim_{x\to a} g(x) = M$$

   If $M \neq 0$, show why there is some neighborhood about $x = a$ in which $g(x) \neq 0$ for all x in this neighborhood.

2. Prove using the $\varepsilon - \delta$ definition of limit,

   (a) that $\lim_{x\to 4} 4x + 7 = 23$.

   (b) for every $a > 0$, $\lim_{x\to a} \frac{1}{x} = \frac{1}{a}$.

   (c) for every $a > 0$, $\lim_{x\to a} x^3 = a^3$.

3. Define the function $f(x) = x\chi_{\mathbb{Q}}(x)$ (that is, $f(x) = x$ if $x \in \mathbb{Q}$ and $f(x) = 0$ if $x \in \mathbb{R}\backslash\mathbb{Q}$). Show that

   (a) $\lim_{x\to 0} f(x) = 0$.

   (b) For any $a \neq 0$, $\lim_{x\to a} f(x)$ does not exist[63].

4. Define the function $f(x) = \sin(x^{-1})$ for $x \in (0, \infty)$. Show that the limit $\lim_{x\to 0^+} f(x)$ does not exist. [Hint: use that $\sin(n\pi) = 0$ and $\sin((2n + 1/2)\pi) = 1$ for every $n \in \mathbb{N}$]

Lecture 24 - 8/16/24

[63] This is an example of a function which is continuous at one point and one point only.

7.2 Continuity

Definition 54. *We saw that a function $f : I \to \mathbb{R}$ is* **continuous at** *$a \in I^\circ$ if for every $\epsilon > 0$, there exists a $\delta > 0$ such that*

$$|x - a| < \delta, \implies |f(x) - f(a)| < \epsilon$$

Put equivalently, this is saying that

$$\lim_{x \to a} f(x) = f(a)$$

And we say that f is **continuous on** *I if f is continuous at every point of I.*

Remark 8. *For points $a \in I$ that are not interior points and thus in the boundary of I, we still talk about f being continuous at a but in these cases we use a one-handed limit in the definition. For example f being continuous on $[a, b]$ means f is continuous in the usual sense on (a, b) and at $x = a$ we use right handed limits and at $x = b$ left handed limits.*

Do note the slight difference in this definition and our definition of limits in general. Here we no longer require that $|x - a| > 0$, as in the case that $x = a$ the difference between $f(x)$ and $f(a)$ vanishes. From theorem 78 we have that $\lim_{x \to a} f(x) = l$ if and only if for every sequence $\{x_n\} \to a$ we had $\{f(x_n)\} \to l$. In terms of this definition, we see that the continuity of f at a means

$$\lim_{n \to \infty} f(x_n) = f\left(\lim_{n \to \infty} x_n\right)$$

i.e. one can interchange taking the limit as $n \to \infty$ and applying f without issue.

Theorem 81. *Given functions $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ that are continuous on I, we have that $f + g, f - g, f \cdot g$ are continuous on I, and $\frac{f}{g}$ is continuous on $I \backslash \{x \in I \mid g(x) = 0\}$.*

Proof. Take $a \in I$, as f and g are continuous at a, we have that

$$\lim_{x \to a} f(x) = f(a), \qquad \lim_{x \to a} g(x) = g(a)$$

From theorem 80 we have that

$$\lim_{x \to a} (f \pm g)(x) = f(a) \pm g(a) = (f \pm g)(a)$$
$$\lim_{x \to a} (f \cdot g)(x) = f(a) \cdot g(a) = (f \cdot g)(a)$$

and for $g(a) \neq 0$, we have

$$\lim_{x \to a} \left(\frac{f}{g}\right)(x) = \frac{f(a)}{g(a)} = \left(\frac{f}{g}\right)(a)$$

Thus we see that $f \pm g, f \cdot g, \frac{f}{g}$ are all continuous at a and as a is arbitrary the result holds. □

Similar to our result on the algebraic limit rules in section 2.2, specifically exercise 7, the theorem above immediately implies that any polynomial function, $p(x)$, is continuous on the real line, and any rational function $R(x)$ is continuous on its domain of definition.

Now with the added assumption of continuity, we can see how limits interact with function composition.

Theorem 82. *For continuous functions $f : I \to \mathbb{R}$ and $g : J \to \mathbb{R}$ with $ran(f) \subseteq dom(g)$, the function $g \circ f$ is continuous.*

Proof. Let $a \in \text{dom}(f)$ and $\{x_n\}$ be an arbitrary sequence contained in the domain of f with $\{x_n\} \to a$, then as f is continuous at a we have that $\{f(x_n)\} \to f(a)$. As $\text{ran}(f) \subseteq \text{dom}(g)$ and g is continuous at $f(a)$ we have that $\{g(f(x_n))\} \to g(f(a))$. Putting this another way, we have that $\{(g \circ f)(x_n)\} \to (g \circ f)(a)$. As this can be done with any sequence $\{x_n\}$ converging to a we have that

$$\lim_{x \to a} (g \circ f)(x) = (g \circ f)(a)$$

and thus $g \circ f$ is continuous at $x = a$. As a was an arbitrary element of I we have that $g \circ f$ is continuous on I. □

Example 36. *Let us consider the function $F : \mathbb{R} \to \mathbb{R}$ given by*

$$F(x) = \frac{x}{\sqrt[4]{x^4 + 71}}$$

If we define the following functions

$$\begin{aligned} f &: \mathbb{R} \to \mathbb{R} & f(x) &= x \\ g &: (-\infty, 0) \cup (0, \infty) \to \mathbb{R} & g(x) &= \frac{1}{x} \\ h &: \mathbb{R} \to [0, \infty) & h(x) &= \sqrt[4]{x} \\ j &: \mathbb{R} \to \mathbb{R} & j(x) &= x^4 + 71 \end{aligned}$$

then we have that

$$F(x) = (f \cdot (g \circ h \circ j))(x)$$

and thus as $f(x), g(x), h(x)$, and $j(x)$ are all continuous on their domain we have that $F(x)$ is continuous on $\mathbb{R}$.

The notion of continuity can also be phrased in terms of topics we have learned in our section on topology. In particular in abstract topological spaces, X, that do not have a notion of distance[64] this is typically taken as the definition of a continuous function. Before we state the theorem, recall that for a function $f : X \to Y$ that the **image** of $A \subseteq X$ is written

$$f(A) = \{f(x) \mid x \in A\}$$

and the **pre–image** of a set $B \subseteq Y$ is written

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}.$$

Theorem 83. *A function $f : \mathbb{R} \to \mathbb{R}$ is continuous on $\mathbb{R}$ if $f^{-1}(V)$ is an open set for every open set V.*

Proof. $\implies$ Assume that f is continuous on $\mathbb{R}$ and let V be an open set within $\mathbb{R}$. If $V \cap \text{ran}(f) = \varnothing$, then $f^{-1}(V) = \varnothing$ and thus is open. So, assume that $V \cap \text{ran}(f) \neq \varnothing$ and take $y \in V$ with $f(x) = y$,

[64] a space with a distance function on it that determines its topology is called a metric space

thus $x \in f^{-1}(V)$. As V is open, there exists an $\epsilon > 0$ such that $(f(x) - \epsilon, f(x) + \epsilon) \subseteq V$. And note that

$$L \in (f(x) - \epsilon, f(x) + \epsilon) \iff |f(x) - L| < \epsilon$$

Now as f is continuous and thus continuous at x we have that for this $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|y - x| < \delta, \implies |f(y) - f(x)| < \epsilon$$

And for all $y \in (x - \delta, x + \delta)$, which is equivalent to $|y - x| < \delta$, implies that $|f(y) - f(x)| < \epsilon$, or equivalently, $f(y) \in (f(x) - \epsilon, f(x) + \epsilon)$. So

$$f((x - \delta, x + \delta)) \subseteq (f(x) - \epsilon, f(x) + \epsilon) \subseteq V$$

and thus $(x - \delta, x + \delta) \subseteq f^{-1}(V)$. Thus we have just shown that for an arbitrary $x \in f^{-1}(V)$ that a neighborhood of x is contained within $f^{-1}(V)$ and thus this set is open.

$\Longleftarrow$ Now assume that $f^{-1}(V)$ is open for every open set V in $\mathbb{R}$. Take $x \in \mathbb{R}$ and let $\epsilon > 0$. The set $(f(x) - \epsilon, f(x) + \epsilon)$ is open, and thus by assumption we have that $f^{-1}((f(x) - \epsilon, f(x) + \epsilon))$ is an open set. As $x \in f^{-1}((f(x) - \epsilon, f(x) + \epsilon))$, as this set is open, there is some $\delta > 0$ such that $(x - \delta, x + \delta) \subseteq f^{-1}((f(x) - \epsilon, f(x) + \epsilon))$. Thus we have that $f((x - \delta, x + \delta)) \subseteq (f(x) - \epsilon, f(x) + \epsilon)$. This is equivalent to saying

$$|x - a| < \delta, \implies |f(x) - f(a)| < \epsilon$$

And as this can be done for any choice of $\epsilon > 0$ we have that f is continuous at x. And as x was arbitrary, we have that this holds for all $x \in \mathbb{R}$ and thus f is continuous on the real line. □

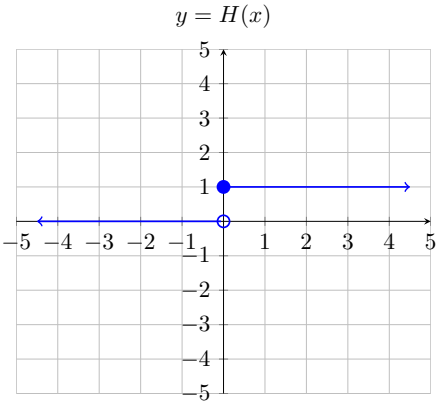
Example 37. *Consider the Heaviside step function again*

$$H(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

The interval $(\frac{1}{2}, \frac{3}{2})$ *is an open set but*

$$H^{-1}\left(\frac{1}{2}, \frac{3}{2}\right) = [0, \infty)$$

Thus we have that the pre-image of H *for an open set is not an open set, i.e.* $H(x)$ *is not continuous on* $\mathbb{R}$ *by the previous theorem, and this can be seen from its jump discontinuity at* $x = 0$.



Exercises for section 7.2:

1. In this problem, use the $\varepsilon - \delta$ definition of continuity.
   (a) Show that $x \mapsto \sqrt{x}$ is continuous on $(0, \infty)$.
   (b) Show that $x \mapsto x^{\frac{1}{3}}$ is continuous on $(0, \infty)$.
   (c) Generalize the previous argument to $x \mapsto x^{1/p}$ for any $p \in \mathbb{N}$.
2. Let $f_1, \ldots, f_n$, n functions which are continuous on the real line. Show that the set
$$A = \{x \in \mathbb{R} : f_j(x) > 0, \ \forall\, j = 1, \ldots, n\},$$
   is open.
3. Show that for a continuous function $f : \mathbb{R} \to \mathbb{R}$ that the pre-image of any closed set is a closed set.
4. Cantor Devil Stair problem (fix this)

7.3 Continuous Functions on Compact Sets

In this section we will explore what happens when we combine the ideas of continuity and compactness, in particular we will focus on continuous functions f, defined on compact sets, K, i.e. $f : K \to \mathbb{R}$.

There is no guarantee that a continuous function will necessarily map open sets to open sets, or closed sets to closed sets, for example

- The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$, we have
$$f((-1, 1)) = [0, 1)$$
  and thus even though f is continuous, it does not necessarily map open sets to open sets.
- The function $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = e^{-x}$ maps
$$g([0, \infty)) = (0, 1]$$
  and thus even though g is continuous, it does not necessarily map closed sets to closed sets.

But it will turn out that continuity of a function is sufficient to map compact sets to compact sets.

Theorem 84. *Given a function $f : \mathbb{R} \to \mathbb{R}$ that is continuous and K a compact set, the image $f(K)$ is compact.*

For this theorem we will give two proofs: one involving the definition of compactness and the other using sequential compactness. Either will work as we have shown that these are equivalent in section 6.4.

Proof. (*Argument via Compactness*) Assume that $f : \mathbb{R} \to \mathbb{R}$ is continuous and let K be a compact set. Take $f(K)$ and let $\mathcal{U} = \{U_i \mid i \in \Lambda\}$ be an arbitrary open cover of $f(K)$. As f is continuous, we know that the pre-image $f^{-1}(V)$ is open for any open set V. Thus define the following

$$f^{-1}(\mathcal{U}) = \{f^{-1}(U_i) \mid i \in \Lambda\}$$

By what we mentioned above involving the continuity of f, $f^{-1}(\mathcal{U})$ is a collection of open sets. We now come to a lemma

Lemma 85. *For a function $f : X \to Y$, we generally have the following*

- $f(f^{-1}(B)) \subseteq B$
- $f^{-1}(f(A)) \supseteq A$

Proof. For the first claim, take $y \in f(f^{-1}(B))$, then $y = f(x)$ for some $x \in f^{-1}(B)$ and so by definition we have that $y = f(x) \in B$. As y is arbitrary we have that $f(f^{-1}(B)) \subseteq B$.

For the second claim, take $x \in A$, then $y = f(x) \in f(A)$. And so by definition $x \in f^{-1}(f(A))$, and thus as x is arbitrary we have that $A \subseteq f^{-1}(f(A))$. □

From this lemma we immediately have

$$f(K) \subseteq \bigcup_{i\in\Lambda} U_i \implies K \subseteq f^{-1}(f(K)) \subseteq \bigcup_{i\in\Lambda} f^{-1}(U_i)$$

and thus $f^{-1}(\mathcal{U})$ is an open cover of K. As K is compact, this open cover has a finite subcover, thus we have

$$K \subseteq \bigcup_{i=1}^{n} f^{-1}(U_i)$$

for a finite collection of indices. And thus by our lemma we have

$$f(K) \subseteq f\left(\bigcup_{i=1}^{n} f^{-1}(U_i)\right) = \bigcup_{i=1}^{n} f(f^{-1}(U_i)) \subseteq \bigcup_{i=1}^{n} U_i$$

And as we have seen that the open cover of $f(K)$ can be refined into a finite subcover and the original cover was chosen arbitrarily, we have shown that $f(K)$ is compact. □

Proof. (*Argument via Sequential Compactness*) Assume that f is continuous and let K be a compact set. Let $\{y_n\}$ be a sequence contained within $f(K)$. For each n as $y_n \in f(K)$ we have that $y_n = f(x_n)$ for some $x_n \in K$.[65]

Now the sequence $\{x_n\}$ is contained within K. As K is compact and thus sequentially compact, there exists a subsequence $\{x_{n_k}\}$ that is convergent with $\{x_{n_k}\} \to l$ and $l \in K$. From the continuity of f, we have that $\{x_{n_k}\} \to l$ implies that $\{y_{n_k}\} = \{f(x_{n_k})\} \to f(l)$, and so $\{y_n\}$ contains a convergent subsequence $\{y_{n_k}\}$ that converges to $f(l) \in f(K)$. And so $f(K)$ is sequentially compact and thus compact. □

[65] technically we are using the axiom of choice here as we are choosing an element from $f^{-1}(\{y_n\})$.

As we saw earlier, in $\mathbb{R}$ compact sets are closed and bounded, thus $f(K)$ for a continuous function f and a compact set K is closed and bounded. Similar to the definition for sequences, we say that $f : I \to \mathbb{R}$ is **bounded above** if there exists a M such that

$$f(x) \leq M, \qquad \forall x \in I$$

and similarly we say f is **bounded below** if there exists L such that

$$L \leq f(x), \qquad \forall x \in I$$

We call f **bounded** if we have

$$|f(x)| \leq M, \qquad \forall x \in I$$

Thus the theorem we just proved assures us that the image of continuous functions is bounded over compact sets.

Example 38. *Consider the following examples:*

a). $f : (0, 1] \to \mathbb{R}$ *given by* $f(x) = \frac{1}{x}$.

This function is continuous on the interval $(0, 1]$ *but it is not bounded as* $f(x) \to \infty$ *as* $x \to 0$ *from the right. This agrees with our theorem as* $(0, 1]$ *is not a compact set. This function becomes unbounded because the domain arbitrarily gets close to a singularity of* f.

b). $g : [0, \infty) \to \mathbb{R}$ *given by* $g(x) = x^2$.

This function is continuous on $[0, \infty)$ *but is not bounded as* $g(x) \to \infty$ *as* $x \to \infty$. *Once again, this agrees with our theorem as* $[0, \infty)$ *is not compact. In this case, the function* g *becomes unbounded because the domain is unbounded.*

The examples above show how continuous functions can be bounded over non-compact domains. There are functions that are bounded over non-compact domains, but these have a different kind of defect.

Example 39. *Consider the following:*

a). $F : \mathbb{R} \to \mathbb{R}$ *given by* $F(x) = \frac{1}{x^2+1}$.

This function is continuous over its domain and is bounded by 1. While this function does achieve its global maximum of 1 at $x = 0$*, it has no minimum value.*

b). $G : (0, 1) \to \mathbb{R}$ *given by* $G(x) = 2 - x$.

This function is continuous over its domain and is bounded by 1, but this function has no maximum or minimum value over its domain.

c). $H : [0, 1] \to \mathbb{R}$ *given by* $H(x) = e^{-x}$.

This function is continuous over its domain and is bounded by 1 and it does achieve its global maximum and minimum values on its domain, 1 at $x = 0$ *and* $\frac{1}{e}$ *at* $x = 1$ *respectively.*

In particular, and hopefully what the example elucidated, is that we can not guarantee that a continuous function achieves its global maximum or minimum over its domain, but we can say something when the domain of the function is compact.[66] This leads us to the following

[66] continuous functions over non-compact domains can achieve their global maxima and minima but usually care must be taken and the argument needs to be made case by case or by having some decay conditions at $\pm\infty$.

Theorem 86. (Extreme Value Theorem - EVT) *For a continuous function $f : K \to \mathbb{R}$, if K is compact then the following values*

$$M = \sup_{x \in K} f(x), \qquad m = \inf_{x \in K} f(x)$$

are defined and these values are attained by f, i.e. there exist $x, y \in K$ such that $f(x) = M$ and $f(y) = m$.

Proof. If K is compact, then by our prior theorem we have that $f(K)$ is compact as f is continuous. Thus $f(K)$ is closed and bounded. As $f(K)$ is bounded above and below and non-empty, we have that M and m exist within $\mathbb{R}$. And by definition

$$M = \sup f(K), \qquad m = \inf f(K)$$

From theorem 67 we have that a closed set contains its supremum and infimum, thus $M, m \in f(K)$. Thus there exist $x, y \in K$ such that $f(x) = M$ and $f(y) = m$. □

Lecture 25 - 8/19/24

Now the notion of compactness or really compact domains for continuous functions gives us another incredibly useful property, however let us see some motivation first.

Example 40. *Consider the following:*

a). The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$.

Let $a > 0$ be a random positive input (the negative case is similar) and let us say we want a 'fixed' error window, i.e. we will fix $\epsilon > 0$ and we will ask ourselves what must our input window, $(a - \delta, a + \delta)$ look like to guarantee that the image is within the output window, $(f(a) - \epsilon, f(a) + \epsilon)$.

Thus our output window looks like $(a^2 - \epsilon, a^2 + \epsilon)$. To find the inputs that correspond to this output window for $f(x) = x^2$ we can make use of the inverse function, $\sqrt{x}$. Thus $(\sqrt{a^2 - \epsilon}, \sqrt{a^2 + \epsilon})$ is our input window, i.e. for any $y \in (\sqrt{a^2 - \epsilon}, \sqrt{a^2 + \epsilon})$ we will have $f(y) \in (a^2 - \epsilon, a^2 + \epsilon)$.

Jumping ahead momentarily to section 8, we have that the linear approximation to $\sqrt{x}$ at a^2 is given by

$$L(x) = \sqrt{a^2} + \frac{1}{2\sqrt{a^2}}(x - a^2)$$

and we have that $L(a^2 - \epsilon) = a - \frac{\epsilon}{2a}$ and $L(a^2 + \epsilon) = a + \frac{\epsilon}{2a}$, and thus our approximate input window is $(a - \frac{\epsilon}{2a}, a + \frac{\epsilon}{2a})$.[67] *But we see that*

$$y \in (a - \frac{\epsilon}{2a}, a + \frac{\epsilon}{2a}), \implies f(y) \in (a^2 - \epsilon, a^2 + \epsilon)$$

The point is this. If we are keeping ϵ fixed because we want the same error window around each output, we see that our input window has a width of roughly $\frac{\epsilon}{a}$. Because of this, the further a is from 0 the input window will shrink by the same factor. This makes sense as x^2 grows faster and faster the further from 0 you are. But the point is that the width of the input interval δ is definitely dependent on ϵ but also where the input a is in the domain.

b). The function $g : (-\infty, 0) \cup (0, \infty) \to \mathbb{R}$ given by $g(x) = \frac{1}{x}$.

[67] the right endpoint is a little larger than it should be due to the linear approximation to $\sqrt{x}$.

Let $a > 0$ be a random positive input (the negative case is similar) and let us say we want a 'fixed' error window, i.e. we will fix $\epsilon > 0$ and we will ask ourselves what must our input window, $(a-\delta, a+\delta)$ look like to guarantee that the image is within the output window, $(f(a)-\epsilon, f(a)+\epsilon)$.

Thus our output window looks like $(\frac{1}{a}-\epsilon, \frac{1}{a}+\epsilon)$. To find the inputs that correspond to this output window for $g(x) = \frac{1}{x}$ we can make use of the inverse function which is $g(x)$ itself. Thus $(\frac{1}{\frac{1}{a}-\epsilon}, \frac{1}{\frac{1}{a}+\epsilon})$ is our input window, i.e. for any $y \in (\frac{1}{\frac{1}{a}-\epsilon}, \frac{1}{\frac{1}{a}+\epsilon})$ we will have $f(y) \in (\frac{1}{a}-\epsilon, \frac{1}{a}+\epsilon)$.

Simplifying the input window we see that it is $\left(\frac{a}{1-a\epsilon}, \frac{a}{1+a\epsilon}\right)$. And even with ϵ fixed, we can see that if a is approaching 0 that the input window shrinks, and once again this makes sense as $\frac{1}{x}$ grows infinitely as we approach the singularity at 0, thus the input window must shrink if we expect the output window to stay constant in width. But once again, we see that the input window is not just dependent on ϵ but also the location of the input a.

c). The function $h : \mathbb{R} \to \mathbb{R}$ given by $h(x) = mx + b$ for constants m, b.

Let $a > 0$ be a random positive input (the negative case is similar) and let us say we want a 'fixed' error window, i.e. we will fix $\epsilon > 0$ and we will ask ourselves what must our input window, $(a-\delta, a+\delta)$ look like to guarantee that the image is within the output window, $(f(a)-\epsilon, f(a)+\epsilon)$.

Thus our input window looks like $(ma+b-\epsilon, ma+b+\epsilon)$. Similar to the examples above, to find the corresponding input window we will apply the inverse function $h^{-1}(y) = \frac{y-b}{m}$, and we find the associated input window is $\left(\frac{ma+b-\epsilon-b}{m}, \frac{ma+b+\epsilon-b}{m}\right)$. After simplifying we see the input window is $\left(a-\frac{\epsilon}{m}, a+\frac{\epsilon}{m}\right)$, i.e.

$$y \in \left(a-\frac{\epsilon}{m}, a+\frac{\epsilon}{m}\right), \implies f(y) \in (ma+b-\epsilon, ma+b+\epsilon)$$

In this case, we see that the width of the input window is only dependent upon ϵ, i.e. it will not shrink or grow as the value of a changes. This is because a linear function has a constant growth rate, so the input window does not need to grow or shrink to account for singularities or locations where the growth rate of the function changes.

Another way to phrase f being continuous over an entire interval I is that

$$\forall x \in I,\ \forall \epsilon > 0,\ \exists \delta > 0,\ \forall y \in I,\ |x-y| < \delta \implies |f(x)-f(y)| < \epsilon$$

and in particular what we saw with our examples in that δ is dependent upon ϵ and x, i.e. that the size of the neighborhood about x in input space depends on the allowable amount of error in output space, $\epsilon > 0$, and where we are in the domain of f, x. This is because a functions behavior can easily change at different points (approaching a singularity, the function increases at different rates at different places).

But there are some examples we saw (linear functions) in which δ is only dependent upon ϵ, i.e. the location of where we are taking a limit has no effect on the size of the input neighborhood. As it turns out, this is true of any continuous function over a compact domain.

Definition 55. *We say that $f : I \to \mathbb{R}$ is* **uniformly continuous** *on I if*

$$\forall \epsilon > 0,\ \exists \delta > 0,\ \forall x, y \in I,\ |x-y| < \delta \implies |f(x)-f(y)| < \epsilon$$

Thus f is uniformly continuous on I if δ is only dependent on ϵ.

If f is uniformly continuous, then after choosing an allowable error, ϵ, you can take an input neighborhood $(x-\delta, x+\delta)$ and slide it around, i.e. change x, and you will still have $|f(x)-f(y)| < \epsilon$ for all $y \in (x-\delta, x+\delta)$. In this sense, after fixing $\epsilon > 0$, the input neighborhoods will have a fixed 'width' for all $x \in I$.

Example 41. *Consider the following:*

a). The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$. This function is not uniformly continuous by what we saw in the prior example. In particular, for a given $\epsilon > 0$, the $\delta > 0$ for the input window is dependent upon ϵ as well as the location of the input x.

b). The function $g : (-\infty, 0) \cup (0, \infty) \to \mathbb{R}$ given by $g(x) = \frac{1}{x}$. This function is not uniformly continuous by what we saw in the prior example. In particular, for a given $\epsilon > 0$, the $\delta > 0$ for the input window is dependent upon ϵ as well as the location of the input x.

c). As we saw in the prior example, any linear function is uniformly continuous on $\mathbb{R}$.

d). The function $F : [0, 10] \to \mathbb{R}$ given by $F(x) = x^2$. This function is uniformly continuous. Even though we saw that the size of input windows is dependent on choice of ϵ and location of input, in this case as the domain is bounded we can simply take δ at its worst/tiniest when the input is 10 and use this δ for all other inputs.

Example d). suggests the following theorem. The reason x^2 is not uniformly continuous on $\mathbb{R}$ but is over any finite interval $[a, b]$ is because we saw the width of the input interval shrinks as the distance of the input from 0 grows. On $\mathbb{R}$, this means our choice of δ would go to zero, but δ is supposed to be positive.

Theorem 87. *If $f : K \to \mathbb{R}$ is continuous and K is compact, then f is uniformly continuous on K.*

For this proof, again two arguments will be given.

Proof. (*Argument via Compactness*): Fix $\epsilon > 0$. As f is continuous on K, for each $x \in K$ for this $\epsilon > 0$ there is a $\delta_x > 0$ such that

$$|y - x| < \delta_x, \implies |f(y) - f(x)| < \frac{\epsilon}{2}$$

Next define $U_x = \left(x - \frac{\delta_x}{2}, x + \frac{\delta_x}{2}\right)$, i.e. the open set that is the neighborhood of x that is centered about x and of width δ_x. As this can be done for each $x \in K$, define

$$\mathcal{U} = \{U_x \mid x \in K\}.$$

We then immediately have that $\mathcal{U}$ is an open cover of K. As K is compact there is some finite subcover of $\mathcal{U}$, thus

$$K \subseteq U_{x_1} \cup U_{x_2} \cup \cdots U_{x_N}$$

And now define $\delta = \frac{1}{2}\min(\delta_{x_1}, \delta_{x_2}, \ldots, \delta_{x_N})$.

If we have two points $x, y \in K$ with $|x - y| < \delta$, then we have that $x \in U_{x_j}$ for some $1 \leq j \leq N$ from the finite subcover condition, thus $|x - x_j| < \frac{\delta_{x_j}}{2} < \delta_{x_j}$ and

$$|y - x_j| = |y - x + x - x_j| \leq |y - x| + |x - x_j| < \delta + \frac{\delta_{x_j}}{2} < \delta_{x_j}$$

and so we have

$$|f(x)-f(y)| = |f(x)-f(x_j)+f(x_j)-f(y)| \leq |f(x)-f(x_j)| + |f(y)-f(x_j)| < \frac{\epsilon}{2}+\frac{\epsilon}{2} = \epsilon$$

Thus we have just shown that for any $x, y \in K$ with $|x-y| < \delta$ that $|f(x)-f(y)| < \epsilon$. As this δ works for arbitrary x, y we have that f is uniformly continuous on K. □

Proof. (*Argument via Sequential Compactness*): Let us assume by contradiction that f is not uniformly continuous, i.e.

$$\exists \epsilon > 0,\ \forall \delta > 0,\ \exists x, y \in K,\ |x-y| < \delta \text{ and } |f(x)-f(y)| \geq \epsilon$$

Thus for this fixed $\epsilon > 0$, we take $\delta = 1$, and the condition above gives us the existence of $x_1, y_1 \in K$ with $|x_1 - y_1| < 1$ and $|f(x_1)-f(y_1)| \geq \epsilon$. And similarly for each $n \in \mathbb{N}$ and taking $\delta = \frac{1}{n}$, the condition above gives us $x_n, y_n \in K$ with $|x_n - y_n| < \frac{1}{n}$ and $|f(x_n)-f(y_n)| \geq \epsilon$.

Thus we have two sequences $\{x_n\}$ and $\{y_n\}$ in K. As K is compact and therefore sequentially compact, we have convergent subsequences of these sequences, i.e. there exist $l, m \in K$ such that $\{x_{n_k}\} \to m$ and $\{y_{n_k}\} \to l$. As we have $|x_n - y_n| < \frac{1}{n}$, we have that $\{x_n - y_n\} \to 0$ and thus $l = m$. As f is continuous on K we have that $\{f(x_{n_k})\} \to f(m)$ and $\{f(y_{n_k})\} \to f(l)$. But then we have

$$|f(x_{n_k}) - f(y_{n_k})| \geq \epsilon, \quad \text{and} \quad \lim_{k\to\infty} |f(x_{n_k}) - f(y_{n_k})| = 0$$

as $f(m) = f(l)$. From our result about how sequences interact with orders, this implies $0 \geq \epsilon$ and this is a clear contradiction. □

To close out this section, we know that for a general function $f : X \to Y$ that if f is injective, then f^{-1} exists as a function $f^{-1} : \text{ran}(f) \to Y$. If f is also surjective then we have $f^{-1} : Y \to X$.[68] So, of course, a natural question that arises is that if f is a continuous injective function, when can we say that f^{-1} is continuous? The answer to this will be subtle as there are two general cases we will consider. We will encounter the first in this section and save the later for a section involving monotonic functions.

Theorem 88. *Suppose $f : K \to \mathbb{R}$ is a continuous one-to-one function and K is compact, then the inverse of f is a continuous function.*

Proof. As f is one-to-one we know that f^{-1} exists as a function $f^{-1} : \text{ran}(f) \subseteq \mathbb{R} \to \mathbb{R}$. We will show f^{-1} is continuous by showing that for V an open set of $\mathbb{R}$ that the pre-image of f^{-1} of V is open. Very quickly, we have that this pre-image is by definition

$$(f^{-1})^{-1}(V) = \{x \in \text{ran}(f) \mid f^{-1}(x) \in V\}$$

As f is bijective onto its range, for each x in this set there exists $p \in K$ such that $f(p) = x$ and thus

$$\begin{aligned}(f^{-1})^{-1}(V) &= \{x \in \text{ran}(f) \mid f^{-1}(x) \in V\} = \{f(p) \in \text{ran}(f) \mid f^{-1}(f(p)) \in V\} \\ &= \{p \in V \mid f(p)\} = f(V)\end{aligned}$$

[68] this is not necessary but most people like to say a function is invertible if and only if the function is bijective. The surjectivity is not required, but you can also always says that an injective function $f : X \to Y$ is bijective as $f : X \to \text{ran}(f)$.

Thus we can prove this result by showing that $f(V)$ is open for every set V that is open in K.

So take V in K open. Thus V^c is a closed set in K[69], and we saw that closed subsets of compact spaces are compact, thus V^c is compact. As we know that the continuous image of a compact set is a compact set we have that $f(V^c)$ is compact. And thus $f(V^c)$ is closed as compact sets are closed. And so $f(V^c)^c$ is open.

Lemma 89. *For $f : X \to Y$ an injective (one-to-one) function, for any $A, B \subset X$ we have*

$$f(A)\backslash f(B) = f(A\backslash B)$$

Proof. Let $y \in f(A)\backslash f(B)$, thus $y = f(x)$ for some $x \in A$ and $y \neq f(b)$ for any $b \in B$, thus it must be that $x \notin B$, thus $x \in A\backslash B$ and $y \in f(A\backslash B)$, and so

$$f(A)\backslash f(B) \subseteq f(A\backslash B)$$

It should be noted that this direction is generally true as we did not invoke the injectivity of f anywhere.

Now take $y \in f(A\backslash B)$, thus $y = f(x)$ for $x \in A$ and $x \notin B$. It is clear that $y \in f(A)$. If $y \in f(B)$, then there exists $p \in B$ with $y = f(p)$, but the injectivity of f would imply

$$f(x) = y = f(p) \implies p = x$$

and this would mean $x \in B$ and $x \notin B$, which is a contradiction. Thus $y \notin f(B)$, so $y \in f(A)\backslash f(B)$, and so

$$f(A)\backslash f(B) \supseteq f(A\backslash B)$$

□

This lemma implies that $\text{ran}(f)\backslash f(A) = f(X\backslash A)$, i.e. that $f(A^c) = f(A)^c$ in the relative spaces each inhabit. Thus as $f(V^c)^c = f(V)$, we have that $f(V)$ is open, and so f^{-1} is continuous. □

Exercises for section 7.3:

1. We say that f is **Lipschitz-continuous** with constant M on I if

$$\forall x, y \in I,\ |f(x) - f(y)| \leq M|x - y|.$$

   Suppose that f and g are both Lipschitz continuous on $\mathbb{R}$.

   (a) Show that $f + g$ is Lipschitz continuous on $\mathbb{R}$.

   (b) Show that if in addition f and g are bounded, then $f \cdot g$ is Lipschitz continuous.

   (c) Give a counterexample to show that it is necessary to assume boundedness in (b).

2. On Lipschitz-continuity and uniform continuity.

   (a) Show that if $f : I \to \mathbb{R}$ is Lipschitz-continuous on an interval I, then f is uniformly continuous on I.

[69] i.e. V^c is closed in the relative topology that K inherits from $\mathbb{R}$

(b) Show that the function $f(x) = \sqrt{x}$ is uniformly continuous on $[0,1]$ but not Lipschitz-continuous on $[0,1]$. (for the latter, you need to show that ratios of the form $\frac{f(x)-f(y)}{x-y}$ with $x, y \in [0,1]$ and $x \neq y$ are unbounded)

3. Suppose that f and g are both uniformly continuous on $\mathbb{R}$.

   (a) Show that $f+g$ is uniformly continuous on $\mathbb{R}$.

   (b) Show that if in addition f and g are bounded, then $f \cdot g$ is uniformly continuous.

   (c) Give a counterexample to show that it is necessary to assume boundedness in (b).

4. Show that if f is uniformly continuous on $\mathbb{R}$, there exists positive constants M_1, M_2 such that $|f(x)| \leq M_1 + |x|M_2$ for all $x \in \mathbb{R}$.

5. Let A be a non-compact set in $\mathbb{R}$:

   (a) Give an example of a continuous function on A that is not bounded.

   (b) Give an example of a continuous and bounded function on A that has no maximum or minimum.

   (c) Assume that A is bounded, show there is a continuous function on A that is not uniformly continuous.

7.4 Continuous Functions on Connected Sets

In this section we will explore the connections between continuous functions and connected sets. In particular we will focus on continuous functions defined on connected sets. Recall that in $\mathbb{R}$ connected sets are either open, closed, or half-open intervals and singleton sets. We will primarily focus on the case of our domains being intervals due to the following.

Theorem 90. *Any function $f : D \to \mathbb{R}$ with D a discrete set is automatically continuous.*

Proof. Take $x \in D$ and let $\epsilon > 0$. As D is a discrete set, there is a neighborhood about x that only meets D in the point x, i.e. there is a $\delta > 0$ such that $(x-\delta, x+\delta) \cap D = \{x\}$. In this case for $y \in D$ with $|x-y| < \delta$ implies that $y = x$, and so $|f(x) - f(y)| = 0$. Thus we have found a δ for this ϵ such that

$$|x-y| < \delta, \implies |f(x) - f(y)| < \epsilon$$

As this can be done for any ϵ we have that f is continuous at x and as x was arbitrary we have that f is continuous on D. □

In our prior section we saw that continuous functions mapped compact sets to compact sets, and we would like to show a similar result for connected sets. Because of the theorem above, we will restrict ourselves to intervals as singletons are clearly mapped to singletons.[70]

Theorem 91. *For $f : E \to \mathbb{R}$ a continuous function, if E is connected, then its image, $f(E)$ is also connected.*

[70] and for discrete sets the continuity of any abstract function is an artifact of discrete sets having no limit points

Proof. We will prove this result by contrapositive. Thus assume that $f(E)$ is disconnected, i.e. let A and B be two separated sets that are non-empty with $f(E) = A \cup B$. Now define the sets

$$S = E \cap f^{-1}(A), \qquad T = E \cap f^{-1}(B).$$

We have that $E = S \cup T$ and that S, T are both non-empty as A, B are non-empty.

As f is a continuous function, the pre-image of any open set is open and the pre-image of any closed set is closed. Thus $f^{-1}(\overline{A})$ is a closed set that contains S. As closures are the smallest closed set containing a given set we have

$$\overline{S} \subseteq f^{-1}(\overline{A}) \iff f(\overline{S}) \subseteq \overline{A}$$

and similarly we have $f(\overline{T}) \subseteq \overline{B}$.

Recall that for general functions we have $f(C \cap D) \subseteq f(C) \cap f(D)$, and thus if $x \in \overline{S} \cap T$, then

$$f(x) \in f(\overline{S} \cap T) \subseteq f(\overline{S}) \cap f(T) \subseteq \overline{A} \cap B = \varnothing$$

as A and B are separated. And so this results in a contradiction, thus it must be the case that $\overline{S} \cap T = \varnothing$ and similarly one can show that $S \cap \overline{T} = \varnothing$. Thus S and T are non-empty separated sets making up E and so E is disconnected. Thus we have proven the result by contrapositive. □

As an immediate result of this theorem we have the following classical intermediate value theorem that says a continuous function over an interval assumes all intermediate values.

Theorem 92. (Intermediate Value Theorem - IVT) *For $f : [a, b] \to \mathbb{R}$ a continuous function with $f(a) < f(b)$ (resp. $f(a) > f(b)$), then for any y with $f(a) < y < f(b)$ (resp. $f(b) < y < f(a)$) there exists $x \in [a, b]$ such that $y = f(x)$.*

Proof. As $[a, b]$ is a connected set, by the theorem we just proved, $f([a, b])$ is a connected set as f is continuous. As $f(a), f(b) \in f([a, b])$, then the definition of connectedness means that if $f(a) < y < f(b)$ then $y \in f([a, b])$. Thus there exists $x \in [a, b]$ such that $y = f(x)$. □

Example 42. *One classical use of the intermediate value theorem is to show the existence of at least one root to the polynomial equation $p(x) = 0$ when*

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

with n odd.

For the moment assume that $a_n > 0$, as x^n grows faster than any other power within the polnomial we have

$$\lim_{x \to \infty} p(x) = \lim_{x \to \infty} x^n = \infty, \qquad \lim_{x \to -\infty} p(x) = \lim_{x \to -\infty} x^n = -\infty$$

Thus we can assume for large enough values of x that $p(x) > 0$ and that for large enough negative values of x that $p(x) < 0$. As polynomial functions are continuous the intermediate value theorem implies the existence of a c such that $p(c) = 0$.

Definition 56. *For a function $f : I \to \mathbb{R}$ we have the following:*

- *f is called* **monotonically increasing** *(resp.* **strictly increasing***) if $x \leq y$ implies that $f(x) \leq f(y)$ (resp. if $x < y$ implies that $f(x) < f(y)$)*

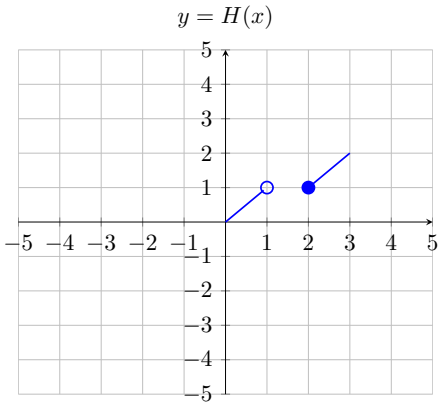
- *f is called* **monotonically decreasing** *(resp.* **strictly decreasing***) if $x \leq y$ implies that $f(x) \geq f(y)$ (resp. if $x < y$ implies that $f(x) > f(y)$)*

Note that monotonically increasing functions are order preserving while monotonically decreasing functions are order reversing.

Any function that is strictly increasing or decreasing is injective and thus f^{-1} exists for any strictly increasing or decreasing function.

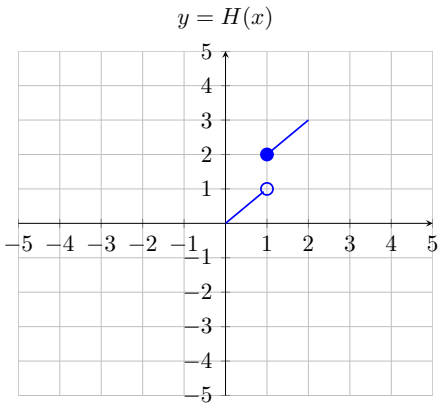
Example 43. *Consider the following function $f : [0, 1) \cup [2, 3] \to \mathbb{R}$ given by*

$$f(x) = \begin{cases} x, & x \in [0, 1) \\ x - 1, & x \in [2, 3] \end{cases}$$



And so f is injective and onto its range, $ran(f) = [0, 2]$. Thus f has an inverse function given by $f^{-1} : [0, 2] \to \mathbb{R}$

$$f^{-1}(x) = \begin{cases} x, & x \in [0, 1) \\ x + 1, & x \in [1, 2] \end{cases}$$



Now f is a continuous function as it is continuous on each connected component (see exercise 6), but f^{-1} is not a continuous function.

The point of this example is that it gives an example of a continuous function (over disconnected components) whose image is one connected set in which f^{-1} is no longer continuous. This may make you think that the only condition preventing f^{-1} from being continuous was the disconnected domain and you would be correct.

Theorem 93. *For $f : I \to \mathbb{R}$ a continuous function defined on an interval I that is either strictly increasing or decreasing, f^{-1} exists and is a continuous function.*

Proof. Without loss of generality, assume that f is strictly increasing and that I is of the form (a, b). As we know that the continuous image of a connected set is connected, we have that $\text{ran}(f)$ is an interval as well. Let $c = f(a)$ and $d = f(b)$, by the intermediate value theorem we can say that $\text{ran}(f) = (c, d)$. As strictly increasing implies that f is injective we have that f^{-1} exists as a function $f^{-1} : (c, d) \to (a, b)$.

For $x, y \in (c, d)$ if it were the case that $f^{-1}(x) \geq f^{-1}(y)$ for $x < y$, then as f is strictly increasing and order preserving this would imply that

$$x = f(f^{-1}(x)) \geq f(f^{-1}(y)) = y$$

which is a clear contradiction. Thus we have that $x < y$ implies that $f^{-1}(x) < f^{-1}(y)$ and so f^{-1} is also a strictly increasing function.

To show that f^{-1} is a continuous function, let $\epsilon > 0$. Let us look at the condition we want to prove, we want to show the existence of $\delta > 0$ such that

$$|p - q| < \delta, \implies |f^{-1}(p) - f^{-1}(q)| < \epsilon.$$

Call $x = f^{-1}(p)$, then we see

$$\begin{aligned} |f^{-1}(p) - f^{-1}(q)| &< \epsilon \\ -\epsilon < x - f^{-1}(q) &< \epsilon \\ -x - \epsilon < -f^{-1}(q) &< \epsilon - x \\ x - \epsilon < f^{-1}(q) &< x + \epsilon \\ f(x - \epsilon) < q &< f(x + \epsilon) \end{aligned}$$

and similarly we have

$$|p - q| < \delta, \implies p - \delta < q < p + \delta$$

Thus we would like to make a choice of δ that makes $p + \delta < f(x + \epsilon)$ and $p - \delta > f(x - \epsilon)$. Thus take $\delta = \min(f(x+\epsilon) - p, p - f(x-\epsilon))$. As f is strictly increasing, this choice of δ is positive. And this choice gives

$$|p - q| < \delta, \implies p - \delta < q < p + \delta, \implies f(x - \epsilon) < q < f(x + \epsilon)$$

which as we saw above is equivalent to $|f^{-1}(p) - f^{-1}(q)| < \epsilon$. As this can be done for any $\epsilon > 0$ we have that f^{-1} is continuous. □

Exercises for section 7.4:

only needs to be continuous on components.

1. (*Tricky!*) Give an example of a function on $\mathbb{R}$ that has the intermediate value property for every interval (it takes on all values between $f(a)$ and $f(b)$ on $a \leq x \leq b$) but fails to be continuous at a point. Can such a function have jump discontinuities
2. Show that the graphs of $f(x) = \frac{1}{1+x^2}$ and $g(x) = 5x^5 + 4x^4$ must intersect.

3. Another proof of Intermediate Value Theorem goes as follows. Define $S := \{x \in [a,b],\ f(x) \leq y\}$.
   (a) Show that $c = \sup S$ exists.
   (b) Show that, by using continuity of f at c, the options $f(c) < y$ and $f(c) > y$ both contradict that $c = \sup S$. (and thus we must have $f(c) = y$).
4. Given $f : I \to \mathbb{R}$ a continuous function over an interval, I, and assume that f is injective. Show why f is strictly increasing or decreasing for the entirety of I. Do you need the domain of f to be connected for this result to remain true?
5. Let $y > 0$, use the intermediate value theorem to prove why the nth root of y exists as a real number.
6. Let $f : X \to \mathbb{R}$ be a function. Assume $X = A \cup B$ and that $f(x) = a$ on A and $f(x) = b$. What conditions on a, b, A, B will assure that f is a continuous function? There are multiple answers.

7.5 Discontinuities (Optional)

discontinuities and types, first kind or simple, second kind, Ex 4.27 in rudin.

monotonic functions - definition, all left and right hand limits exist. and have no discontinuities of second kind. Set of discontinuities is countable, maybe G delta and oscillations. Remark 4.31 and devil stair.

f from an interval to r that is strictly increasing or decreasing and continuous has a continuous inverse that is also strictly increasing or decreasing, cantor set exampe?

zoology of discontinuities, once again, mention oscillation.

Exercises for section 7.5:

7.6 Limits at infinity and singularities (Optional)

infinite limits and limits at infinity. neighborhoods of infinity and negative infinity.

infinite limits, neighborhoods of infinity, things we can say converge, and things that are indeterminate forms. do this for sequences first. then infinite limits for functions and functions at singularities, result for rational functions and use this as a possible lead in to L'hopital.

Exercises for section 7.6:

Lecture 26 - 8/21/24

8 Differentiation

8.1 Definitions & Properties

Definition 57. *Let $f : I \to \mathbb{R}$ be a function and let I be an open interval and let $a \in I$. For $x \neq a$ define $q(x) = \frac{f(x)-f(a)}{x-a}$. If $q(x)$ has a finite limit as $x \to a$ then we say that f is* **differentiable** *at a and write*

$$f'(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$$

Equivalently, we say that f is **differentiable** *at a if there exists $y \in \mathbb{R}$ such that*

$$\forall \epsilon > 0,\ \exists \delta > 0,\ \forall x \in I,\ 0 < |x-a| < \delta \implies \left| \frac{f(x)-f(a)}{x-a} - y \right| < \epsilon$$

If f is differentiable at all points of I we say that f is differentiable on I.

Once you know $f'(a)$ exists at a point, the following ϵ–δ form can be very useful.

$$0 < |x-a| < \delta, \implies \left| \frac{f(x)-f(a)-f'(a)(x-a)}{x-a} \right| < \epsilon$$

Example 44. *Let us consider the following:*

a). The function $f(x) = x^3$ at $a = 2$.

For $x \neq 2$ we have that the quotient is

$$q(x) = \frac{f(x)-f(2)}{x-2} = \frac{x^3-8}{x-2} = x^2+2x+4$$

and thus as polynomials are continuous we have that $\lim_{x\to 2} q(x) = 12$ and so $f'(2) = 12$.

b). The function $g(x) = |x|$ at $a = 0$.

For $x \neq 0$ the quotient is

$$q(x) = \frac{g(x)-g(0)}{x-0} = \frac{|x|}{x} = \begin{cases} 1, & x > 0 \\ -1, & x < 0 \end{cases}$$

and from this we see that $q(0+) = \lim_{x\to 0^+} q(x) = 1$ and $q(0-) = \lim_{x\to 0^-} q(x) = -1$. Thus the limit of $q(x)$ does not exist at 0, so $g(x)$ is not differentiable at 0.

c). The function $h : [0,\infty) \to \mathbb{R}$ given by $h(x) = \sqrt{x}$ at $a = 0$.

For $x > 0$ the quotient is given by

$$q(x) = \frac{h(x)-h(0)}{x-0} = \frac{\sqrt{x}}{x} = \frac{1}{\sqrt{x}}$$

As $\lim_{x\to 0^+} q(x) = \infty$, we have that the limit of the quotient does not exist at 0, thus $h(x)$ is not differentiable at $a = 0$.

d). The function $F(x) = x^n$ for $n \in \mathbb{N}$ at a, i.e. a basic monic polynomial. For this example we will use a slightly different definition of the quotient.

$$q(h) = \frac{F(a+h) - F(a)}{h}$$

By the binomial theorem we have that

$$F(a+h) = \sum_{k=0}^{n} \binom{n}{k} h^{n-k} a^k$$

thus

$$F(a+h) - F(a) = \sum_{k=0}^{n} \binom{n}{k} h^{n-k} a^k - a^n = \sum_{k=0}^{n-1} \binom{n}{k} h^{n-k} a^k$$

and as $n - k > 0$ for $0 \leq k \leq n-1$, we have that every term in this sum has a factor of h, thus

$$q(h) = \frac{F(a+h) - F(a)}{h} = \sum_{k=0}^{n-1} \binom{n}{k} h^{n-k-1} a^k$$

As $h \to 0$ the only term that will not vanish is the term with $k = n-1$, thus

$$F'(a) = \lim_{h \to 0} q(h) = \binom{n}{n-1} a^{n-1} = \frac{n!}{(n-1)!1!} a^{n-1} = n a^{n-1}$$

It is often useful to frame the differentiability of f at a point a in terms of a decay condition on a remainder from a linear approximation to f at a.

Theorem 94. *A function $f : I \to \mathbb{R}$ with I an open interval is differentiable at $a \in I$ with $f'(a) = l$ if and only if there exists a $\delta > 0$ and a function $\eta : (a-\delta, a+\delta) \to \mathbb{R}$ such that for all $x \in (a-\delta, a+\delta)$ we have*

$$f(x) = f(a) + l(x-a) + (x-a)\eta(x)$$

and $\lim_{x \to a} \eta(x) = 0$.

Proof. $\implies$ If f is differentiable at a and $f'(a) = l$, then for $\epsilon > 0$ there is a $\delta > 0$ and we can define $\eta : (a-\delta, a+\delta) \to \mathbb{R}$ by

$$\eta(x) = \begin{cases} 0, & x = a \\ \frac{f(x)-f(a)}{x-a} - f'(a) & x \neq a \end{cases}$$

then we have

$$f(x) = f(a) + l(x-a) + (x-a)\eta(x)$$

and $\lim_{x \to a} \eta(x) = 0$.

$\impliedby$ If we assume there exists a $\delta > 0$ and a function $\eta : (a-\delta, a+\delta) \to \mathbb{R}$ such that for all $x \in (a-\delta, a+\delta)$ we have

$$f(x) = f(a) + l(x-a) + (x-a)\eta(x)$$

and $\lim_{x \to a} \eta(x) = 0$. Then we have that

$$\eta(x) = \frac{f(x) - f(a)}{x-a} - l$$

and so

$$0 = \lim_{x \to a} \eta(x) = \lim_{x \to a} \frac{f(x) - f(a)}{x-a} - l$$

and this says that f is differentiable at a and $f'(a) = l$. □

Often times the result above is written in 'little-o' Landau notation.

Definition 58. *For two functions f and g and a a point common to the domain of both functions, we write $f(x) = o(g(x))$ as $x \to a$ if*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = 0$$

i.e. f approaches 0 faster than g as $x \to a$, or f becomes negligible compared to g as $x \to a$.

The result above shows that when f is differentiable at a we can write

$$f(x) = f(a) + f'(a)(x-a) + o(x-a)$$

for values of x sufficiently close to a. And this way of viewing the derivative gives us some immediate classical results.

Theorem 95. *If f is differentiable at a, then f is continuous at a.*

Proof. As f is differentiable at a, we have that

$$f(x) = f(a) + f'(a)(x-a) + o(x-a)$$

for some neighborhood of a, i.e. $(a-\delta, a+\delta)$. But then we have that

$$\lim_{x\to a}(f(x) - f(a)) = \lim_{x\to a} f^a(x-a) + \lim_{x\to a} o(x-a) = 0 + 0 = 0$$

and this shows that $\lim_{x\to a} f(x) = f(a)$ and thus f is continuous at a. □

Do note however that the converse of this result is not true as we saw in example b). above.

In our section on continuity we saw that the sum, difference, product, quotient, and composition of continuous functions are continuous, and in fact the same will be true for differentiability as well. Let us collect these results now.

Theorem 96. *For functions $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ that are differentiable on I, we have the following:*

i). The sum/difference of f and g is differentiable on I and

$$(f \pm g)'(a) = f'(a) \pm g'(a)$$

ii). The product of f, g is differentiable on I and

$$(f \cdot g)' = f'(a)g(a) + f(a)g'(a)$$

iii). The quotient of f, g is differentiable on I except at points where $g(x) = 0$ and

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{(g(a))^2}$$

Proof. Let $a \in I$ be a point in the interval and assume that

$$f(x) = f(a) + f'(a)(x-a) + (x-a)\eta(x)$$
$$g(x) = g(a) + g'(a)(x-a) + (x-a)\chi(x)$$

where $\eta(x)$ and $\chi(x)$ are defined on $(a-\delta, a+\delta)$ and

$$\lim_{x\to a} \eta(x) = \lim_{x\to a} \chi(x) = 0$$

And now let us begin with the proof in earnest.

Proof of i). If we look at the following quotient, we have

$$\frac{(f+g)(x) - (f+g)(a)}{x-a} = \frac{f(x) + g(x) - f(a) - g(a)}{x-a} = \frac{f(x) - f(a) + g(x) - g(a)}{x-a}$$

If we replace $f(x) - f(a)$ and $g(x) - g(a)$ with what we have found above, we have that

$$\frac{(f+g)(x) - (f+g)(a)}{x-a} = \frac{f'(x-a) + (x-a)\eta(x) + g'(a)(x-a) + (x-a)\chi(x)}{x-a}$$

thus

$$\frac{(f+g)(x) - (f+g)(a)}{x-a} = f'(a) + g'(a) + \eta(x) + \chi(x)$$

Taking the limit of both sides as $x \to a$ shows that $(f+g)(x)$ is differentiable at a and its derivative is $f'(a) + g'(a)$. The argument for differences of functions is similar.

Proof of ii). Let us look at the quotient,

$$\begin{aligned}
\frac{(fg)(x) - (fg)(a)}{x-a} &= \frac{f(x)g(x) - f(a)g(a)}{x-a} = \frac{f(x)g(x) - f(a)g(x) + f(a)g(x) - f(a)g(a)}{x-a} \\
&= \frac{g(x)(f(x) - f(a)) + f(a)(g(x) - g(a))}{x-a} \\
&= \frac{g(x)[f'(a)(x-a) + (x-a)\eta(x)] + f(a)[g'(a)(x-a) + (x-a)\chi(x)]}{x-a} \\
&= g(x)f'(a) + f(a)g'(a) + \eta(x) + \chi(x)
\end{aligned}$$

By taking the limits of both sides as $x \to a$ and using that as g is differentiable at a it is therefore continuous at a, thus $\lim_{x\to a} g(x) = g(a)$, we have that $(fg)(x)$ is differentiable at a and

$$(fg)'(a) = f'(a)g(a) + f(a)g'(a)$$

Proof of iii). We will first prove a simpler result, let us look at $\frac{1}{g(x)}$ at a where $g(a) \neq 0$. As g is differentiable and thus continuous at a, we have that $g(x) \neq 0$ for an entire neighborhood containing a. Let us assume we are only looking within this neighborhood. Now we look at the quotient given by

$$\begin{aligned}
\frac{(\frac{1}{g})(x) - (\frac{1}{g})(a)}{x-a} &= \frac{\frac{1}{g(x)} - \frac{1}{g(a)}}{x-a} = \frac{g(a) - g(x)}{(x-a)g(x)g(a)} = -\frac{g(x) - g(a)}{(x-a)g(x)g(a)} \\
&= -\frac{g'(a)(x-a) + (x-a)\chi(x)}{(x-a)g(x)g(a)} = -\frac{g'(a)}{g(x)g(a)} - \frac{\chi(x)}{g(x)g(a)}
\end{aligned}$$

By taking the limit of both sides as $x \to a$ and once again using the continuity of g at a we have that $\frac{1}{g}$ is differentiable at a and

$$\left(\frac{1}{g}\right)' = -\frac{g'(a)}{(g(a))^2}$$

And so now using part ii). we have that $\frac{f}{g}$ is differentiable and

$$\left(\frac{f}{g}\right)'(a) = \left(f \cdot \frac{1}{g}\right)'(a) = f'(a)\left(\frac{1}{g(a)}\right) - f(a)\left(\frac{g'(a)}{(g(a))^2}\right) = \frac{f'(a)g(a) - f(a)g'(a)}{(g(a))^2}$$

□

Due to this result and induction on degrees of polynomials, we have that every polynomial is differentiable at every point of the real numbers. The quotient rule also tells us that any rational function is also differentiable at any point in its domain (implicitly exluding points in which the denominator is zero).

A quick use of the definition of the derivative shows that for any constant function $f(x) = c$,

$$f'(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h} = \lim_{h\to 0} \frac{c-c}{h} = 0$$

is 0 as we likely expected from out prior Calculus knowledge. It will turn out that the converse of this result is also true, i.e. if f has a derivative of 0 at every point, then $f(x) = c$ is a constant function, but we will leave this to a later section as this argument requires the mean value theorem.

Theorem 97. (Chain Rule) *Suppose that f is continuous on $[a, b]$, and $f'(x)$ exists at some $c \in (a, b)$, g is defined on some interval containing the range of f and g is differentiable at the point $f(c)$, then $(g \circ f)(x)$ is differentiable at $x = c$ and*

$$(g \circ f)'(c) = g'(f(c))f'(c)$$

Proof. Let $d = f(c)$ and $y = f(x)$ and $h(x) = g(f(x))$. As f is differentiable at c we have that

$$f(x) = f(c) + f'(c)(x - c) + (x - c)\eta(x)$$

on some small interval about c with $\lim_{x\to c} \eta(x) = 0$. Similarly as g is differenitable at d we have that

$$g(y) = g(d) + g'(d)(y - d) + (y - d)\nu(y)$$

on some small interval about d and $\lim_{y\to d} \nu(y) = 0$. Now let us look at the quotient

$$\begin{aligned}\frac{h(x) - h(c)}{x - c} &= \frac{g(f(x)) - g(f(c))}{x - c} = \frac{g(y) - g(d)}{x - c} = \frac{g(y) - g(d)}{y - d} \cdot \frac{y - d}{x - c} \\ &= \frac{g(y) - g(d)}{y - d} \cdot \frac{f(x) - f(c)}{x - c} = \frac{g'(d)(y - d) + (y - d)\nu(y)}{y - d} \frac{f'(c)(x - c) + (x - c)\eta(x)}{x - c} \\ &= (g'(d) + \nu(y))(f'(c) + \eta(x))\end{aligned}$$

As $x \to c$ we have that $y \to d$ as f is continuous and thus $\lim_{x\to c} f(x) = f(c)$, and so taking the limit of both sides we have

$$\lim_{x\to c} \frac{h(x) - h(c)}{x - c} = (g'(d) + 0)(f'(c) + 0)$$

and so $h(x)$ is differentiable at $x = c$ and it's derivative is given by $g'(d)f'(c)$, thus

$$(g \circ f)'(c) = g'(f(c))f'(c)$$

□

Example 45. *Consider the following:*

a). The function $h(x) = \sin x$.[71] *Let us find the derivative of this function using the definition*

$$\begin{aligned} h'(a) &= \lim_{h\to 0} \frac{\sin(a+h) - \sin a}{h} = \lim_{h\to 0} \frac{\sin a \cos h + \sin h \cos a - \sin a}{h} \\ &= \sin a \lim_{h\to 0} \frac{\cos h - 1}{h} + \cos a \lim_{h\to 0} \frac{\sin h}{h} = \sin a(0) + \cos a(1) = \cos a \end{aligned}$$

Thus we see the derivative of $\sin x$ *is* $\cos x$.

b). The function given by

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

From what we remarked earlier we know that this function is differentiable for $x \neq 0$ *and for* $x \neq 0$ *we have*

$$f'(x) = \sin\left(\frac{1}{x}\right) - \frac{1}{x}\cos\left(\frac{1}{x}\right)$$

For $x = 0$, *we have*

$$f'(0) = \lim_{x\to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x\to 0} \frac{x \sin\left(\frac{1}{x}\right)}{x} = \lim_{x\to 0} \sin\left(\frac{1}{x}\right)$$

as this limit does not exist as $x \to 0$ *we have that* $f'(0)$ *does not exist, thus* f *is not differentiable at* $x = 0$.

c). The function given by

$$g(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & x = 0 \end{cases}$$

As we have remarked earlier this function is differentiable for $x \neq 0$ *and its derivative is given by*

$$g'(x) = 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right)$$

At $x = 0$ *we have*

$$g'(0) = \lim_{x\to 0} \frac{g(x) - g(0)}{x - 0} = \lim_{x\to 0} \frac{x^2 \sin\left(\frac{1}{x}\right) - 0}{x - 0} = \lim_{x\to 0} x \sin\left(\frac{1}{x}\right) = 0$$

and so g *is differentiable at* $x = 0$ *and* $g'(0) = 0$ *by the squeeze theorem. But, due to the* $\frac{1}{x}$ *term, we have that* $\lim_{x\to 0} g'(x)$ *does not exist, thus* $g'(x)$ *is not continuous at* $x = 0$.

d). The function $F(x) = x^2$. *As have seen earlier,* $F'(x) = 2x$, *and as polynomials are continuous we have that* $F'(x)$ *exists at every point and is continuous. Similarly, the second derivative of* F *exists, and we write* $F''(x) = 2$ *and this function is continuous as well.*

[71] we have not proven the addition and subtraction formulas yet, but let us assume them for this example

The examples above lead into the following definition.

Definition 59. *For a function $f : I \to \mathbb{R}$ where I is an open interval, we call f a $C^1(I)$* **function** *if f is differentiable on I and f' is a continuous function on I. Similarly, we define f to be a $C^k(I)$* **function** *if f is k-times differentiable function on I and $f^{(k)}$ is continuous on I.*

In terms of the examples above, b). was not differentiable on all of $\mathbb{R}$, c). was differentiable on $\mathbb{R}$ but not C^1 and d). was C^1 on $\mathbb{R}$.

to finish

how tangent line is best linear approximation, mention examples at the start again, put in pictures.

Exercises for section 8.1:

1. Using the pythagorean identity
$$\cos^2 x + \sin^2 x = 1$$
and the derivative of $\sin x$ that was found in the section, find what the derivative of $\cos x$ is.

2. Use the derivative rules from this section and what was found in the previous problem to find the derivatives of $\tan x, \cot x, \sec x$, and $\csc x$ where they are defined.

3. Assume that $a \in \mathbb{N}$, what condition is required of a to guarantee that
$$f(x) = \begin{cases} x^a \sin\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & x = 0 \end{cases}$$
is C^1 on $\mathbb{R}$? Please provide proof of this result.

4. What is the derivative to $f(x) = \sqrt{x}$ at $a \in (0, \infty)$.

5. Using the prior question, find the best linear approximation to $g(x) = 3 + 4x^2 + 2\sqrt{x}$ at $a = 1$.

8.2 Local Extrema & Mean Value Theorem

maybe some intro material

Theorem 98. *Let $f : I \to \mathbb{R}$ be a function on an open interval and assume that f is differentiable at $a \in I$, then the following is true:*

- *If f is monotonically increasing on an interval (c, d) containing a, then $f'(a) \geq 0$. .*
- *If f is monotonically decreasing on an interval (c, d) containing a, then $f'(a) \leq 0$. .*
- *If f is constant on an interval (c, d) containing a, then $f'(a) = 0$.*

Proof. For the first result, if we let $x > a$, the quotient
$$q(x) = \frac{f(x) - f(a)}{x - a} \geq 0$$

due to the monotonic increasing nature of f. And we a similar fact for $y < a$,

$$q(y) = \frac{f(a) - f(y)}{a - y} \geq 0$$

As we know that f is differentiable at a, we know the limit of the quotient exists as $x \to a$ or $y \to a$, and thus $f'(a) \geq 0$.

The proof of the claim for monotonically decreasing functions is the same as above with the inequalities being switched.

If f is constant on an interval containing a, then f is both monotonically increasing and decreasing, thus $f'(a) = 0$. □

This result does not really need to exist. It is placed here as we will be seeing the converse of the result by the end of this section. The reality is that we already know that a constant function is differentiable at every point of its domain and its derivative is zero by definition. It is also probably obvious that due to the constant or increasing nature of a monotonically increasing function that its derivative is nonnegative when it exists (similarly nonpositive for monotonically decreasing). We will see that many results in this section depend upon the assumption of differentiability of a function for the claims to hold. At the end we will make a small comment on what we know about differentiability for monotonic functions.

Definition 60. *Let $f : I \to \mathbb{R}$ be a function. We call say f has a* **local maximum** *at a if there is a neighborhood of a, i.e. $(a - \delta, a + \delta)$ such that $f(x) \leq f(a)$ for all x in this neighborhood. Similarly, we say f has a* **local minimum** *at a if there is a neighborhood of a, i.e. $(a - \delta, a + \delta)$ such that $f(x) \geq f(a)$ for all x in this neighborhood.*

Theorem 99. *Let $f : I \to \mathbb{R}$ be defined on an open interval. If f has a local maximum or minimum at $a \in I$ and f is differentiable at a, then $f'(a) = 0$.*

Proof. Without loss of generality assume that f has a local maximum at a. Thus there is a $\delta > 0$ such that for $x \in (a - \delta, a + \delta)$ we have $f(x) \leq f(a)$.

For $x \in (a, a + \delta)$ we have that

$$q(x) = \frac{f(x) - f(a)}{x - a} \leq 0$$

as the numerator is nonpositive and the denominator is positive. As we know that f is differentiable at a, taking the limit as $x \to a$ gives us that $f'(a) \leq 0$.

For $y \in (a - \delta, a)$ we have that

$$q(y) = \frac{f(y) - f(a)}{y - a} \geq 0$$

as the numerator is nonpositive and the denominator is negative. As we know f is differentiable at a, taking the limit as $y \to a$ gives us that $f'(a) \geq 0$. Thus it must be that $f'(a) = 0$. □

put in some words

Theorem 100. (Mean Value Theorem - MVT)*: Let $f : [a, b] \to \mathbb{R}$ be a continuous function that is differentiable on (a, b), then there is a point $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Proof. We will first prove a simpler case, that of $f(a) = f(b)$ and showing there exists a point $c \in (a, b)$ with $f'(c) = 0$.[72].

If $f(a) = f(b)$, then there are on of two possibilities. Either f is constant on the interval $[a, b]$ or f is not constant on the interval $[a, b]$. If f is constant, then $f'(c) = 0$ for all $c \in (a, b)$ and thus the result holds. If f is non-constant, then as f is a continuous function over the compact domain of $[a, b]$ the extreme values of f are attained on $[a, b]$. Let c be where f attains a global minimum and d be where f attains a global maximum. It must be that at least one of c, d are not equal to a or b, otherwise, for example if $c = a$ and $d = b$ then we would have

$$f(b) = f(a) = f(c) \leq f(x) \leq f(d) = f(b)$$

and this contrdicts f being non-constant on $[a, b]$. Thus at least one of c or d is in (a, b), and by the previous theorem we have $f'(c) = 0$ or $f'(d) = 0$ at this point as global extrema are local extrema. Thus, the proof of our simpler case is proven.

In the general case, the secant line connecting the points $(a, f(a))$ and $(b, f(b))$ on the graph of $f(x)$ is given by

$$y = f(a) + \left[\frac{f(b) - f(a)}{b - a}\right](x - a).$$

If we define a new function $g(x)$ that is the difference between $f(x)$ and this secant line,

$$g(x) = f(x) - \left(f(a) + \left[\frac{f(b) - f(a)}{b - a}\right](x - a)\right)$$

then $g(x)$ is continuous on $[a, b]$ and differentiable on (a, b) and we have that $g(a) = g(b) = 0$. Thus there exists $c \in (a, b)$ with $g'(c) = 0$. Thus

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}$$

and we see the result follows. □

words, converse

Theorem 101. *Suppose f is differentiable in (a, b), then:*

- *If $f'(x) \geq 0$ for all (a, b), then f is monotonically increasing.*
- *If $f'(x) = 0$ for all (a, b), then f is constant.*
- *If $f'(x) \leq 0$ for all (a, b), then f is monotonically decreasing.*

Proof. All of these results follow from the mean value theorem. For $A, B \in (a, b)$ with $A < B$ there exists $c \in (A, B)$ such that

$$f(B) - f(A) = f'(c)(B - A)$$

we see that

[72] this is often called Rolle's Theorem

- If $f'(c) \geq 0$, then $f(A) \leq f(B)$.
- If $f'(c) = 0$, then $f(A) = f(B)$.
- If $f'(c) \leq 0$, then $f(A) \geq f(B)$.

and so the result holds □

derivative gives local information. in terms of neighborhoods.

we had an example where a function has a derivative at every point but the derivative is not continuous. Not every function is the derivative of some function (find example here)

Theorem 102. *Suppose f is differentiable on $[a, b]$ and suppose $f'(a) < c < f'(b)$ (resp. $f'(b) < c < f'(a)$). Then there is a point $x \in (a, b)$ such that $f'(x) = c$.*

Proof. Assume without loss of generality that $f'(a) < c < f'(b)$. Define $g(x) = f(x) - cx$. Thus g is differentiable on $[a, b]$ and $g'(a) < 0$. Thus there is some $t_1 \in (a, a + \delta)$ such that $g(t_1) < g(a)$. Similarly, $g'(b) > 0$, thus there is some $t_2 \in (b - \delta, b)$ such that $g(t_2) < g(b)$. As g is continuous on $[a, b]$, by the extreme value theorem g attains its global minimum on $[a, b]$, but the above shows that it must attain the global minimum on the interior, (a, b). Thus there is some $d \in (a, b)$ where g is a global minimum, thus $g'(d) = 0$. Thus

$$0 = g'(d) = f'(d) - c$$

and the result is proven. □

Derivative does not have discontinuities of first kind or simple discontinuities, but can have second kind as we saw earlier.

we saw monotonic functions have countable number of discontinuties and only of first kind, lebesgue showed derivative of monotonic functions exist at everywhere except for a set of measure zero.

local max and min doesn't mean differentiable, abs value

strict increasing and differentiable doesn't mean positive derivative always, x cubed

function can be increasing a one point but not anywhere else x times 1 plus sin of 1 over x.

bounded derivative means lipschitz continuous

this may be a good place to put Liouville numbers as an appendix, and have an argument why e is transcendental, pi being transcendental may be more difficult.

Exercises for section 8.2:

8.3 Inverse Function Theorem

inverse function theorem, general version and local version after mvt. Result from chain rule if you know inverse is differentiable already, exponential function and log, algebraic and trnscendental functions, ln x and inverse trig, algebraic functions don't necessarily have algebraic antiderivatives or primatives

---

Exercises for section 8.3:

8.4 Second Derivatives & Convexity

second derivatives and class C2, second derivative test, exceptions, convexity and concavity, how sign of second derivative gives convexity

---

Exercises for section 8.4:

8.5 Higher Derivatives & Taylor's Theorem

higher derivatives, class cn and taylors theorem (polynomial not series) examples

---

Exercises for section 8.5:

8.6 L'Hopitals Rule

expansions of rational functions, vanishing to a specific order, l hopital , examples, going back to division problem after lhopital. derivatives of quotient in terms of expansions if terms vanish to same order.

Exercises for section 8.6:

francois page 101, proof of b, typo contradition

9 Integration

Integration

Def 1: Let $[a, b]$ be a given interval of $\mathbb{R}$. By a partition P of $[a, b]$ we mean a finite set of points $x_0, x_1, \ldots, x_n$ where

$$a = x_0 \leq x_1 \leq \cdots x_{n-1} \leq x_n = b.$$

We write $\Delta x_i = x_i - x_{i-1}$ for $i \in \{1, \ldots, n\}$.

For what follows, we will only consider functions f that are bounded over the the interval $[a, b]$, i.e. functions for which there exist real numbers $m, M \in \mathbb{R}$ such that $m \leq f(x) \leq M$ for all $x \in [a, b]$. We will consider integrals of unbounded functions and integrals of functions over unbounded intervals at a later time.

For a partition P we define the following

$$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x)$$
$$m_i = \inf_{x \in [x_{i-1}, x_i]} f(x)$$

and further define the upper and lower sums of f for a partition P by

$$U(f, P) = \sum_{k=1}^{n} M_k \Delta x_k$$
$$L(f, P) = \sum_{k=1}^{n} m_k \Delta x_k$$

Lastly, define the upper and lower Riemann integrals of f on $[a, b]$ by

$$\overline{\int_a^b} f dx = \inf U(f, P), \qquad \underline{\int_a^b} f dx = \sup L(f, P).$$

where the infimum and supremum are taken over all partitions of the interval $[a, b]$.

Note: The upper and lower Riemann integrals always exist for a bounded function. As for any partition P,

$$m(b - a) \leq L(f, P) \leq U(f, P) \leq M(b - a).$$

Thus the collection of all upper sums $\{U(f, P) \mid P \text{ a partition}\}$ is a set of real numbers bounded below by $m(b - a)$, and as such, has an infimum. The argument is similar as to why the collection of all lower sums has a supremum.

Def 2: For a bounded function f, if the upper and lower Riemann integrals of f on $[a, b]$ are equal, then we say that f is Riemann integrable, and denote the integral of f on $[a, b]$ by

$$\int_a^b f dx.$$

For shorthand, we let R denote the set of Riemann integrable functions, i.e. $f \in R$ means that f is Riemann integrable.

This is all well and nice, but we will develop a more general theory for this class.

Def 3: Let α be a monotonically increasing function on $[a, b]$. (Note that $\alpha(a), \alpha(b)$ are finite values) Corresponding to each partition P of $[a, b]$ we denote

$$\Delta\alpha_k = \alpha(x_k) - \alpha(x_{k-1}).$$

Note that as α is monotone increasing, $\Delta\alpha_k \geq 0$ for all $k \in \{1, \ldots, n\}$.

We now define analoguous upper and lower sums for a function f and a given partition P with respect to α.

$$U(f, P, \alpha) = \sum_{k=1}^{n} M_k \Delta\alpha_k$$
$$L(f, P, \alpha) = \sum_{k=1}^{n} m_k \Delta\alpha_k.$$

and use these to define the upper and lower Riemann integrals with respect to α

$$\overline{\int_a^b} f d\alpha = \inf U(f, P, \alpha), \qquad \underline{\int_a^b} f d\alpha = \sup L(f, P, \alpha),$$

where the supremum and infimum are taken over all partitions of $[a, b]$. Once again, this is possible as for any partition P, if $m \leq f(x) \leq M$ for all $x \in [a, b]$, then

$$m[\alpha(b) - \alpha(a)] \leq L(f, P, \alpha) \leq U(f, P, \alpha) \leq M[\alpha(b) - \alpha(a)].$$

We can now give the definition of the Riemann–Stieltjes integral of f with respect to α.

Def 4: For a bounded function f, if the upper and lower Riemann integrals of f with respect to α are equal, then we say that f is Riemann integrable with respect to α, and denote the Riemann–Stieltjes integral of f on the interval $[a, b]$ by

$$\int_a^b f d\alpha.$$

For shorthand, we let $R(\alpha)$ denote the set of Riemann integrable functions with respect to α, i.e. $f \in R(\alpha)$ means f is Riemann integrable with respect to α.

Note: We have assumed very little of α. In particular we have assumed no continuity or smoothness (continuity of the derivative) of α. Later we will see how the differentiability or continuity of α can simplify computing the Riemann–Stieltjes integral.

Note: When $\alpha(x) = x$, the identity function, then the Riemann–Stieltjes integral just reduces to the Riemann integral.

A large part of the theory we will develop here is to discern the character of the sets R, $R(\alpha)$. Phrased another way, we are simply asking what properties or characteristics are required of a bounded function for it to be integrable?

Def 5: We say a partition P^* is a refinement of P if $P^* \supset P$. Given two distinct partitions P_1 and P_2, we define their common refinement to be $P^* = P_1 \cup P_2$.

Theorem 1: If P^* refines P, then

$$L(f, P, \alpha) \leq L(f, P^*, \alpha), \quad U(f, P^*, \alpha) \leq U(f, P, \alpha).$$

Proof. We will only prove the result for lower sums. The proof for upper sums will be similar. It also suffices to prove the result when P^* refines P by only one point, as the case of P^* refining P by an arbitrary finite number of points will follow by the transitivity of $\leq$.

Thus assume that $P^* \backslash P = \{x\}$. If $P = \{x_0, x_1, \ldots, x_n\}$, then there exists $k \in \{1, 2, \ldots, n\}$ such that $x_{k-1} < x < x_k$. (So $P^* = \{x_0, x_1, \ldots, x_{k-1}, x, x_k, \ldots, x_n\}$.) Now denote the following

$$n_1 = \inf_{y \in [x_{k-1}, x]} f(y)$$
$$n_2 = \inf_{y \in [x, x_k]} f(y)$$

Thus we can write the lower sum of P^* as

$$L(f, P^*, \alpha) = \sum_{\substack{j \in \{1,\ldots,n\} \\ j \neq k}} m_j \Delta\alpha_j + n_1[\alpha(x) - \alpha(x_{k-1})] + n_2[\alpha(x_k) - \alpha(x)].$$

Thus

$$\begin{aligned} &L(f, P^*, \alpha) - L(f, P, \alpha) \\ &\qquad = n_1[\alpha(x) - \alpha(x_{k-1})] + n_2[\alpha(x_k) - \alpha(x)] - m_k[\alpha(x_k) - \alpha(x_{k-1})] \\ &\qquad = (n_1 - m_k)[\alpha(x) - \alpha(x_{k-1})] + (n_2 - m_k)[\alpha(x_k) - \alpha(x)]. \end{aligned}$$

As $[x_{k-1}, x] \subset [x_{k-1}, x_k]$, we have that $n_1 \geq m_k$, and similarly, $n_2 \geq m_k$. Thus, we have that $L(f, P^*, \alpha) - L(f, P, \alpha) \geq 0$. □

This result leads to another that is very intuitive. To be specific it simply states a natural ordering relationship between the upper and lower Riemann integrals of f with respect to α.

Theorem 2: For a bounded function f on an interval $[a, b]$,

$$\underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha.$$

Proof. Let P_1, P_2 be arbitrary partitions of $[a, b]$, and let P^* be the common refinement of these two partitions. The previous theorem immediately gives us that

$$L(f, P_1, \alpha) \leq L(f, P^*, \alpha) \leq U(f, P^*, \alpha) \leq U(f, P_2, \alpha).$$

If we fix P_2, the above statement gives that $U(f, P_2, \alpha)$ is an upper bound for the set $\{L(f, P, \alpha) \mid P \text{ a partition}\}$, and as such, must be greater than or equal to its supremum, thus

$$\underline{\int_a^b} f d\alpha = \sup L(f, P, \alpha) \leq U(f, P_2, \alpha).$$

Similarly, $\underline{\int_a^b} f d\alpha$ is a lower bound for the set $\{U(f, P, \alpha) \mid P \text{ a partition}\}$, and thus must be less than or equal to this sets infimum, so

$$\underline{\int_a^b} f d\alpha \leq \inf U(f, P, \alpha) = \overline{\int_a^b} f d\alpha.$$

□

Note: The importance of this theorem is it's use in showing the integrability of a function f. Because of this result it is sufficent to show

$$\underline{\int_a^b} f d\alpha \geq \overline{\int_a^b} f d\alpha.$$

to prove that f is integrable, or it is sufficent to assume

$$\underline{\int_a^b} f d\alpha < \overline{\int_a^b} f d\alpha.$$

and derive a contradiction to show that f is integrable.

Integration: Day 2

We begin today with a theorem concerning the integrability of a function.

Theorem 1: For a bounded function f on the interval $[a, b]$, $f \in R(\alpha)$ if and only if $\forall \epsilon > 0$ there exists a partition P of $[a, b]$ such that

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

Proof. $\Leftarrow$ Assume that $\forall \epsilon > 0$ there exists a partition P of $[a, b]$ such that

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

Thus we may choose an arbitrary $\epsilon > 0$. And corresponding to this ϵ is a partition P with the property above. The following string of inequalities

$$L(f, P, \alpha) \leq \underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha \leq U(f, P, \alpha),$$

follow from the definitions of the upper and lower Riemann integral as well as Theorem 1.2 (Lecture 1, theorem 2). This implies that

$$\begin{aligned} \overline{\int_a^b} f d\alpha &\leq U(f, P, \alpha) \\ -\underline{\int_a^b} f d\alpha &\leq -L(f, P, \alpha) \end{aligned}$$

Thus

$$0 \leq \overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha \leq U(f, P, \alpha) - L(f, P, \alpha) < \epsilon.$$

Now, note that we could follow this argument for any given ϵ, i.e. there was nothing special of the ϵ that we choose. Thus, as this argument holds for all $\epsilon > 0$, it must the case that

$$\underline{\int_a^b} f d\alpha = \overline{\int_a^b} f d\alpha.$$

So $f \in R(\alpha)$.

$\Rightarrow$ Assume that $f \in R(\alpha)$, i.e. the upper and lower Riemann integrals of f are equal. Now, recall the definition of the upper and lower Riemann integrals

$$\overline{\int_a^b} f d\alpha = \inf U(f, P, \alpha), \qquad \underline{\int_a^b} f d\alpha = \sup L(f, P, \alpha),$$

with the infimum and supremum taken over all partitions P of $[a, b]$. Thus, for $\epsilon > 0$, there exists a partition P_1 such that

$$\overline{\int_a^b} f d\alpha \leq U(f, P_1, \alpha) \leq \overline{\int_a^b} f d\alpha + \frac{\epsilon}{2}.$$

and similarly there is a partition P_2 such that

$$\underline{\int_a^b} f d\alpha - \frac{\epsilon}{2} \leq L(f, P_2, \alpha) \leq \underline{\int_a^b} f d\alpha.$$

Thus taking the common refinement $P^* = P_1 \cup P_2$ of the two partitions, we have

$$U(f, P^*, \alpha) \leq U(f, P_1, \alpha) \leq \overline{\int_a^b} f d\alpha + \frac{\epsilon}{2}$$
$$-L(f, P^*, \alpha) \leq -L(f, P_2, \alpha) \leq -\underline{\int_a^b} f d\alpha + \frac{\epsilon}{2}$$

which implies that

$$U(f, P^*, \alpha) - L(f, P^*, \alpha) < \epsilon,$$

as the upper and lower Riemann integrals of f are equal. □

Okay, all this theory is fine and good and all, but we need to be able to do something with it. So, just for the sake of believability, let us do an example.

Example 1: Let us find the integral of $f(x) = x^2$ with respect to $\alpha(x) = x$ on the interval $[0, 1]$. Thus, there are two things to be shown

- That f is integrable on $[0, 1]$
- and, computing the integral of f itself.

In regards to the first point, let n be a natural number, and take the partition of $[0, 1]$ given by

$$P_n = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}.$$

Then for each subinterval $[\frac{k-1}{n}, \frac{k}{n}]$ of the partition we have the following

$$M_k = \left(\frac{k}{n}\right)^2, \quad m_k = \left(\frac{k-1}{n}\right)^2, \quad \Delta\alpha_k = \Delta x_k = \frac{1}{n}.$$

(Note as $f(x) = x^2$ is an increasing function on $[0,1]$ its largest value on any subinterval $[c,d]$ will be the functions value at the right endpoint, $f(d)$. Similarly, the smallest value will be at the left endpoint.) We now have the following computation,

$$\begin{aligned} U(x^2, P_n) - L(x^2, P_n) &= \sum_{k=1}^{n} (M_k - m_k)\Delta x_k = \sum_{k=1}^{n} \left[\left(\frac{k}{n}\right)^2 - \left(\frac{k-1}{n}\right)^2 \right] \frac{1}{n} \\ &= \sum_{k=1}^{n} \left(\frac{k}{n} - \frac{k-1}{n}\right)\left(\frac{k}{n} + \frac{k-1}{n}\right)\frac{1}{n} = \frac{1}{n^3}\sum_{k=1}^{n} 2k - 1 \\ &= \frac{1}{n^3}\left[2\left(\frac{n(n+1)}{2}\right) - n\right] = \frac{1}{n}. \end{aligned}$$

The formula $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ was used above. Note there was nothing special about the natural number n chosen. Thus for $\epsilon > 0$, by the archimedean property, there exists and $N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$, thus, pairing this with the above, we have

$$U(x^2, P_N) - L(x^2, P_N) = \frac{1}{N} < \epsilon.$$

And so, from the previous theorem we have that $f(x) = x^2$ is integrable. Now, to compute the integral.

Let us look at the following. We see that

$$P_{2n} = \{0, \frac{1}{2n}, \frac{1}{n}, \dots 1\}$$

refines P_n. And

$$P_{4n} = \{0, \frac{1}{4n}, \frac{1}{2n}, \frac{3}{4n}, \dots 1\},$$

refines P_{2n}. Thus, we see a sequence

$$P_n \subset P_{2n} \subset P_{4n} \subset \cdots P_{2^k n} \subset P_{2^{k+1} n} \subset \cdots$$

of refinements of partitions. As we now know that $f(x) = x^2$ is integrable, i.e. that the integral exists, we are allowed to compute the integral as

$$\int_0^1 x^2 dx = \lim_{k\to\infty} U(f, P_{2^k n}).$$

(Note that the choice of $n > 1$ really doesn't matter in what follows). By direct computation,

$$\begin{aligned} U(f, P_{2^k n}) &= \sum_{j=1}^{2^k n} M_j \Delta x_j = \sum_{j=1}^{2^k n} \frac{j^2}{2^{2k} n^2} \cdot \frac{1}{2^k n} \\ &= \frac{1}{2^{3k} n^3} \sum_{j=1}^{2^k n} j^2 = \frac{1}{2^{3k} n^3}\left[\frac{2^k n (2^k n + 1)(2(2^k n) + 1)}{6}\right] \\ &= \frac{1}{6}\left(1 + \frac{1}{2^k n}\right)\left(2 + \frac{1}{2^k n}\right) \end{aligned}$$

Thus

$$\int_0^1 x^2 dx = \lim_{k\to\infty}\left[\frac{1}{6}\left(1+\frac{1}{2^k n}\right)\left(2+\frac{1}{2^k n}\right)\right] = \frac{1}{3}.$$

And there you have it, a letter opener.

What now follows is a theorem that is more a collection of properties or formulas that can be invoked once you know a function is integrable or if you have a partition with certain properties.

Theorem 2:

a). For $\epsilon > 0$ and a partition P of an interval $[a, b]$ with the property that

$$U(f, P, \alpha) - L(f, P, \alpha) < \epsilon,$$

then the same property holds for all refinements of P.

b). For a partition $P = \{x_0, x_1, \ldots, x_n\}$ with

$$U(f, P, \alpha) - L(f, P, \alpha) < \epsilon,$$

if for each subinterval $[x_{k-1}, x_k]$ of the partition P there exists points $s_k, t_k \in [x_{k-1}, x_k]$, then

$$\sum_{k=1}^{n} |f(s_k) - f(t_k)|\Delta\alpha_k < \epsilon.$$

c). If $f \in R(\alpha)$ and the hypothesis of part b). holds, then

$$\left|\sum_{k=1}^{n} f(t_k)\Delta\alpha_k - \int_a^b f d\alpha\right| < \epsilon$$

Proof. **a).** For a refinement P^* of P, we have

$$U(f, P^*, \alpha) \le U(f, P, \alpha), \quad \text{and} \quad L(f, P, \alpha) \le L(f, P^*, \alpha).$$

Thus

$$U(f, P^*, \alpha) - L(f, P^*, \alpha) \le U(f, P, \alpha) - L(f, P, \alpha) < \epsilon.$$

b). As $s_k, t_k \in [x_{k-1}, x_k]$, it is clear that $m_k \le f(s_k),\ f(t_k) \le M_k$. Thus

$$\begin{aligned} m_k \le f(s_k) \le M_k \\ -M_k \le -f(t_k) \le -m_k \end{aligned}$$

Thus

$$-(M_k - m_k) \le f(s_k) - f(t_k) \le M_k - m_k,$$

which is equivalent to $|f(s_k) - f(t_k)| \le M_k - m_k$. And so

$$\sum_{k=1}^{n} |f(s_k) - f(t_k)|\Delta\alpha_k \le \sum_{k=1}^{n} (M_k - m_k)\Delta\alpha_k = U(f, P, \alpha) - L(f, P, \alpha) < \epsilon.$$

c). By definition, we have

$$L(f,P,\alpha) \leq \int_a^b f d\alpha \leq U(f,P,\alpha).$$

Similarly, as for $t_k \in [x_{k-1}, x_k]$ we have $m_k \leq f(t_k) \leq M_k$, thus

$$L(f,P,\alpha) \leq \sum_{k=1}^{n} f(t_k)\Delta\alpha_k \leq U(f,P,\alpha).$$

Thus similar to what was done in part b), this is enough to justify

$$\left| \sum_{k=1}^{n} f(t_k)\Delta\alpha_k - \int_a^b f d\alpha \right| \leq U(f,P,\alpha) - L(f,P,\alpha) < \epsilon.$$

□

Finally, enough tool building or toolbox filling depending on how you want to think of it. Now we come to a result where we can definitely put a certain (and probably unsurprising) class of functions inside of the set of integrable functions.

Theorem 3: If f is continuous on the interval $[a,b]$, then $f \in R(\alpha)$ on $[a,b]$.

Proof. The key to this proof is to exploit the fact that continuous functions on compact domains are actually uniformly continuous.

So, to begin, let $\epsilon > 0$ be taken arbitrarily. As $\alpha(a)$, $\alpha(b)$ are both finite, and as f is uniformly continuous on $[a,b]$ for $\frac{\epsilon}{(\alpha(b)-\alpha(a))} > 0$ there exists a $\delta > 0$ such that for all $x, y \in [a,b]$ with $|x-y| < \delta$ implies that

$$|f(x) - f(y)| < \frac{\epsilon}{(\alpha(b)-\alpha(a))}.$$

(**Warning:** Like mentioned in lecture, we are allowed to assume that $\alpha(b) > \alpha(a)$. If $\alpha(a) = \alpha(b)$, then α must be a constant function. But then, $\Delta\alpha = 0$ for any partition P, i.e. in this case the Riemann–Stieltjes integral is very boring as it is always zero)

Now, let P be any partition $P = \{x_0, x_1, \ldots, x_n\}$ with the property that

$$\max_{k\in\{1,\ldots,n\}} \Delta x_k = \max_{k\in\{1,\ldots,n\}} (x_k - x_{k-1}) < \delta.$$

For a given subinterval $[x_{k-1}, x_k]$ of P, as f is continuous on this subinterval (which is a compact set), the extreme value theorem gives us points $s_k, t_k \in [x_{k-1}, x_k]$ such that $f(s_k) = M_k$ and $f(t_k) = m_k$. So, as

$$|s_k - t_k| \leq \max_{k\in\{1,\ldots,n\}} (x_k - x_{k-1}) < \delta$$

we then have that

$$M_k - m_k = f(s_k) - f(t_k) < \frac{\epsilon}{(\alpha(b)-\alpha(a))}$$

And so finally we have

$$\begin{aligned} U(f,P,\alpha) - L(f,P,\alpha) &= \sum_{k=1}^{n} (M_k - m_k)\Delta\alpha_k \leq \frac{\epsilon}{(\alpha(b)-\alpha(a))} \sum_{k=1}^{n} \Delta\alpha_k \\ &= \frac{\epsilon}{(\alpha(b)-\alpha(a))}(\alpha(b)-\alpha(a)) < \epsilon \end{aligned}$$

And thus $f \in R(\alpha)$. □

And so we see that f continuous and α monotonic increasing is enough for f to be Riemann integrable with respect to α. At the beginning of the next class, we will see that switching these properties between the functions f and α gives the same result.

Integration: Day 3

Theorem 1: If f is monotonic on $[a, b]$, and if α is continuous on $[a, b]$, then $f \in R(\alpha)$.

Before we begin the proof, remember that α is still assumed to be monotonic increasing on $[a, b]$.

Proof. As α is continuous on a compact interval $[a, b]$, we have that α is uniformly continuous. Thus for $n \in \mathbb{N}$ and $\epsilon = \frac{\alpha(b)-\alpha(a)}{n}$, there exists a $\delta > 0$ such that for all $x, y \in [a, b]$ with $|x - y| < \delta$, we have that $|\alpha(x) - \alpha(y)| < \frac{\alpha(b)-\alpha(a)}{n}$.

Now, let $P = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a, b]$ such that

$$\max_{k\in\{1,\ldots,n\}} \Delta x_k < \delta$$

Without loss of generality, we will assume that f is monotonic increasing. Then for any given subinterval, $[x_{k-1}, x_k]$ of P we have

$$M_k = f(x_k), \quad m_k = f(x_{k-1})$$

as f takes its largest value at the right most endpoint of the subinterval and similarly, f takes on the smallest value at the left most endpoint of the interval. Because P has the property that the length of any subinterval is less than δ, we have that

$$\begin{aligned} U(f, P, \alpha) - L(f, P, \alpha) &= \sum_{k=1}^{n} (M_k - m_k)\Delta\alpha_k \\ &< \frac{\alpha(b) - \alpha(a)}{n} \sum_{k=1}^{n} [f(x_k) - f(x_{k-1})] \\ &= \frac{[\alpha(b) - \alpha(a)][f(b) - f(a)]}{n} \end{aligned}$$

As n can be taken arbitrarily large, then archimedean property implies that $U(f, P, \alpha) - L(f, P, \alpha) < \epsilon$ for any arbitrary ϵ. Thus, $f \in R(\alpha)$. □

At this point we have proven standard results regarding integrability of functions. In particular, we have seen that continuity and monotonicity of f and α, and vice-versa is enough to guarantee integrability of f with respect to α. But now, we will show that we can lax those criteria substantially.

Theorem 2: Suppose f is bounded on $[a, b]$ and that f has a finite number of discontinuities on the interval $[a, b]$ and that α is continuous at the points of discontinuity of f, then $f \in R(\alpha)$.

The idea behind this proof is as follows: we can not control the difference between the supremum and the infimum on any subinterval of the partition P that contains a discontinuity of f. Thus, we control the size of these subintervals, and 'shrink' them to the point that they 'cancel out' the possibly large $M_k - m_k$ terms that come from f not being continuous. On all other subintervals where f is continuous, this is enough to control the $M_k - m_k$ terms.

Proof. Let $\epsilon > 0$. Let E denote the points at which f is discontinuous in $[a, b]$. And lastly, let $M = \sup_{x\in[a,b]} |f(x)|$, which exists as f is bounded. At each $x_i \in E$, as α is continuous at these points, there exists a $\delta_i > 0$ such that $|x - x_i| < \frac{\delta_i}{2}$ implies that $|\alpha(x) - \alpha(x_i)| < \frac{\epsilon}{|E|}$. Where $|E|$ denotes the cardinality of E, and note that are our assumption is that $|E| < \infty$.

As E is finite, we can define $\delta_1 = \min \frac{\delta_i}{2}$. Also we may assume that none of the intervals $[x_i - \delta_1, x_i + \delta_1]$ do not overlap. If they did, then just shrink δ_1 further until there is no overlap. Hence, now we have a finite number of intervals $[x_i - \delta_1, x_i + \delta_1]$ covering the points $x_i \in E$ in which

$$\alpha(x_i + \delta_1) - \alpha(x_i - \delta_1) < \frac{\epsilon}{|E|}.$$

In particular, this gives that

$$\sum_{j=1}^{|E|} \alpha(x_j + \delta_1) - \alpha(x_j - \delta_1) < \epsilon.$$

Now, let us handle the remainder of the points in the interval $[a, b]$ that is not covered by the intervals $[x_i - \delta_1, x_i + \delta_1]$. Thus, denote

$$K = [a, b]\backslash \left[\bigcup_{j=1}^{|E|} (x_j - \delta_1, x_j + \delta_1) \right].$$

Thus, K is equal to the following

$$K = [a, b] \cap \bigcap_{j=1}^{|E|} [\mathbb{R}\backslash(x_j - \delta_1, x_j + \delta_1)],$$

via the De Morgan laws. Thus K is closed and is a subset of $[a, b]$. Which leads us to the following.

Lemma 1: Closed subsets of compact sets are compact.

Proof. Let $A \subseteq K$ with A closed and K compact. Take $\{U_i\}_{i\in\mathbb{N}}$ to be an arbitrary open cover of A. Then $\{U_i\}_{i\in\mathbb{N}} \cup \mathbb{R}\backslash A$ is an open cover of K. As K is compact, there is a finite subcover $\{V_k\}_{k=1}^n$ from $\{U_i\}_{i\in\mathbb{N}} \cup \mathbb{R}\backslash A$ that covers K.

- If $V_j = \mathbb{R}\backslash A$ for some $1 \leq j \leq n$, then leaving this set out of the finite subcover of K given by $\{V_k\}_{k=1}^n$ gives a finite subcover of the $\{U_i\}$ that covers A.
- If $V_j \neq \mathbb{R}\backslash A$ for all $1 \leq j \leq n$, then $\{V_k\}_{k=1}^n$ is a finite subcover from $\{U_i\}$ which covers K and hence A.

In either case, there is a finite subcover of A. Thus an arbitrary cover of A can be refined into a finite subcover, thus A is compact. □

Back to our proof, on the set K the function f is continuous, and as K is compact we have that f is uniformly continuous on K. Because of this, for our ϵ chosen above, there exists a $\delta_2 > 0$ such that for all $x, y \in K$ with $|x - y| < \delta_2$ implies that $|f(x) - f(y)| < \epsilon$. We now use everything we have constructed to form a very particular partition of $[a, b]$. Define P in the following manner

1. Include the points $x_i - \delta_1$ and $x_i + \delta_1$ in the partition P for each $x_i \in E$.

2. P does not contain any points inside of the intervals $(x_i - \delta_1, x_i + \delta_1)$ for each $x_i \in E$.

3. Include enough points in $P \cap K$ such that the length of the subintervals in K have the property

$$\max_{x_k \in P \cap K} \Delta x_k < \delta_2.$$

Now, because of our construction, there are two types of subintervals coming from the partition P. The first are of the form $[x_i - \delta_1, x_i + \delta_1]$ for each $x_i \in E$. The second are of the form $[y_{k-1}, y_k]$ where $y_{k-1}, y_k \in K$ and $y_k - y_{k-1} < \delta_2$.

For the first type of subintervals $-M \leq m_i, M_i \leq M$, thus $M_i - m_i \leq 2M$. For the second type of subintervals, because of the uniform continuity of f on K, $M_i - m_i < \epsilon$. Thus,

$$\begin{aligned}
U(f, P, \alpha) - L(f, P, \alpha) &= \sum_{y_k \in P} (M_k - m_k)\Delta\alpha_k \\
&= \sum_{y_k \in P \cap K} (M_k - m_k)\Delta\alpha_k + \sum_{j=1}^{|E|} (M_j - m_j)\Delta\alpha_j \\
&< \epsilon \sum_{y_k \in P \cap K} \Delta\alpha_k + 2M \sum_{j=1}^{|E|} [\alpha(x_j + \delta_1) - \alpha(x_j - \delta_1)]. \\
&< \epsilon \sum_{k \in P} \Delta\alpha_k + 2M\epsilon \\
&= \epsilon(\alpha(b) - \alpha(a) + 2M).
\end{aligned}$$

As ϵ was chosen arbitrarily, we can make the expression above as small as we like, thus $f \in R(\alpha)$. □

Now, of course a natural question is if we can do better. We will soon answer this in the affirmative, but we must build a few more tools to make the proof precise.

Note: For what follows, let us work in the case of just Riemann integrability, i.e. $\alpha(x) = x$. We will return to a general Riemann–Stieltjes setting soon.

Defintion 1: A subset A of $\mathbb{R}^n$ has measure 0 if for every $\epsilon > 0$ there is a cover $\{U_1, U_2, U_3, \dots\}$ of A by closed rectangles such that $\sum_{k=1}^{\infty} v(U_k) \leq \epsilon$. In particular, for $n = 1$, the U_k are of the form $U_k = [a_k, b_k]$ and the volume of U_k is $v(U_k) = b_k - a_k$.

Note: You can also use open rectangles (open intervals when $n = 1$) in the definition above and get the same theory in the end.

Example 1:

a). Finite sets have measure 0.

Proof. Let A be a finite subset of $\mathbb{R}$. Thus $|A| = n$ for some $n \in \mathbb{N}$. For each $x_k \in A$ define $U_k = [x_k - \frac{\epsilon}{2n}, x_k + \frac{\epsilon}{2n}]$. Thus the U_k form a cover of A and

$$\sum_{k=1}^{n} v(U_k) = \sum_{k=1}^{n} \frac{\epsilon}{n} = \epsilon.$$

As this can be done for arbitrary ϵ, we have that A is measure 0. □

b). Countable sets and sequences are measure 0.

Proof. Let A be a countable subset of $\mathbb{R}$, then A can be enumerated by the natural numbers, i.e. $\{a_j\}_{j\in\mathbb{N}} = A$. Thus for each $a_j \in A$ define $U_j = [a_j - \frac{\epsilon}{2\cdot 2^j}, a_j + \frac{\epsilon}{2\cdot 2^j}]$. Thus the U_k form a cover of A and

$$\sum_{k=1}^{\infty} v(U_k) = \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = \frac{\epsilon}{2}\sum_{k=0}^{\infty} \frac{1}{2^k} = \frac{\epsilon}{2}\left[\frac{1}{1-\frac{1}{2}}\right] = \epsilon.$$

As this can be done for arbitrary ϵ, we have that A is measure 0. □

c). The rational numbers $\mathbb{Q}$ are measure 0. (Follows from b).)

Integration: Day 4

We continue our foray into the idea of measure 0 today by proving some useful results and examples.

Theorem 1: A countable union of sets of measure 0 is a set of measure 0.

Proof. Let $\{A_k\}_{k\in\mathbb{N}}$ be a countable collection of sets A_k in which each A_k is measure 0. Denote

$$A = \bigcup_{k=1}^{\infty} A_k.$$

Let $\epsilon > 0$. As A_1 is measure 0, there exists an open cover $\{U_{1,1}, U_{1,2}, U_{1,3}, \dots\}$ of A_1 with the property that

$$\sum_{k=1}^{\infty} v(U_{1,k}) < \frac{\epsilon}{2}.$$

Similarly, for each $j \in \mathbb{N}$ there exists a cover $\{U_{j,1}, U_{j,2}, \dots\}$ of A_j such that

$$\sum_{k=1}^{\infty} v(U_{j,k}) < \frac{\epsilon}{2^j}.$$

Now, if we collect all these covers together, i.e. $\{U_{j,k}\}_{j,k\in\mathbb{N}}$, this will be a cover of A, and

$$\sum_{j=1}^{\infty}\sum_{k=1}^{n} v(U_{j,k}) < \sum_{j=1}^{\infty} \frac{\epsilon}{2^j} = \frac{\epsilon}{2}\left[\frac{1}{1-\frac{1}{2}}\right] = \epsilon.$$

Thus A is measure 0. □

Definition 1: A subset A of $\mathbb{R}^n$ has content 0 if for every $\epsilon > 0$ there is a finite cover $\{U_k\}_{k=1}^n$ of A by closed (or open) rectangles such that $\sum_{k=1}^n v(U_k) < \epsilon$.

Note: Clearly a set A being content 0 implies that the set is measure 0. The converse is not in general true, but the next theorem gives us the necessary conditions for content 0 and measure 0 to collapse into the same concept.

Theorem 2: If A is compact and has measure 0, then A has content 0.

Proof. As A has measure 0, there exists a cover $\{U_k\}_{k\in\mathbb{N}}$ of A such that $\sum_{k=1}^{\infty} v(U_k) < \epsilon$ for a given $\epsilon > 0$. As A is compact, there exists a finite subcover of $\{U_k\}_{k\in\mathbb{N}}$, which we will denote by $\{V_k\}_{k=1}^{n}$, and thus

$$\sum_{k=1}^{n} v(V_k) \leq \sum_{k=1}^{\infty} v(U_k) < \epsilon.$$

As ϵ was arbitrary, this implies that A has content 0. □

Okay, at this point, we have the concepts of measure 0 and content 0, both of which exist to give us an idea of sets that have a small 'size' in some regard (In particular, the regard here is measure.) But, the measure of a set should agree with preconceived notions coming from intuition. Like, for example, the measure of an interval $[a, b]$ should be the length of the interval $b-a$. We will not delve into the intricacies of measure theory here (there will be a whole class in your potential future for that), but we would like to check that our ideas of measure 0 and content 0 are not overly powerful. To be more clear, we want to make sure our notion of measure 0 does not include any sets that we inuitively believe have nonzero size, like intervals $[a, b]$.

Theorem 3: If $a < b$, then $[a, b] \subset \mathbb{R}$ does not have content 0. In particular, if $\{U_1, U_2, \ldots, U_n\}$ is a finite cover of $[a, b]$ by closed intervals, then $\sum_{k=1}^{n} v(U_k) \geq b - a$.

Proof. Given the interval $[a, b]$, let $\{U_1, \ldots, U_n\}$ be an open cover of $[a, b]$. We define a partition P of $[a, b]$ in the following manner. Let $a, b \in P$ as always. And let P consist of the endpoints of the intervals U_k, $1 \leq k \leq n$, that lie in the interior of $[a, b]$. Thus $P = \{x_0, x_1, \ldots, x_m\}$, where $x_0 = a$, $x_m = b$, and x_i is an endpoint of some U_k if $0 < i < m$.

Then for P defined in this manner. Any given subinterval $[x_{k-1}, x_k]$ of P lies in at least one U_j. Thus,

$$\sum_{k=1}^{n} v(U_k) \geq \sum_{j=1}^{m} \Delta x_j = b - a.$$

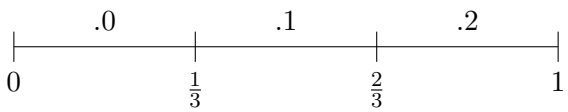
□

At the end of lecture 3, we found that any countable subset of $\mathbb{R}$ had measure 0. Of particular to note is that the converse is NOT a true statement. There exists subsets of $\mathbb{R}$ that are uncountable and measure 0. In particular, we will construct one now.

Example: The Cantor Set

Possibly the easiest way to visualize the construction of the Cantor set is to start with the interval $[0, 1]$.

And break it into thirds as we see above. We now denote these thirds by 0, 1, and 2.



Of interest to note here, is that we are denoting each interval by the common first digit of a numbers ternary expansion. To be more specific, any $x \in [0, \frac{1}{3})$ has 0 as it's first decimal place in it's respective ternary expansion. We now remove the middle third,

And this is the end of the first step. For the next step, we break each of the two remaining intervals into thirds and achieve the common second decimal in the ternary expansions.

And once again remove the middle thirds, to get

Alright, by now I think you get the idea. We repeat this process ad infinitum. At step k we remove the k middle thirds from the prior step, and end up with 2^k intervals. The Cantor set will be what remains after we continue this process a countable number of times. What we can see from the above is that any ternary sequence with a 1 in it will be removed along our process of constructing the Cantor set. Thus, at the end of this process of removing 'middle thirds', each point of the Cantor set will be represented by a countable list made up of 0's and 2's. In particular, if we denote the Cantor set by $\mathfrak{C}$, then we have shown the following

$$\mathfrak{C} = \{0, 2\}^{\mathbb{N}}.$$

Thus, the Cantor set is the set of all $0, 2$ sequences. Because of this, we immediately have that $\mathfrak{C}$ is uncountable.

Note: In a later lecture we will give a more rigorous argument as to why each countable string of 0's and 2's leads to one element of the Cantor set, as long as that element is not a triadic rational.

Let us be a little more formal now in our construction of $\mathfrak{C}$. For the sake of motivation, I will define the following interval $E_\varnothing = (\frac{1}{3}, \frac{2}{3})$, i.e. the first middle third taken out of $[0, 1]$ in the construction process. We now denote

$$E_0 = \left(\frac{1}{9}, \frac{2}{9}\right), \quad E_2 = \left(\frac{7}{9}, \frac{8}{9}\right).$$

The second two middle thirds taken out. In general for $a \in \{0, 2\}^k$, i.e. a is a string of length k, $a = a_1 a_2 \dots a_k$, in which $a_i = 0$ or 2 for $1 \leq i \leq k$, we define

$$E_a = \left(\frac{\sum_{j=1}^k a_j 3^j + 1}{3^{k+1}}, \frac{\sum_{j=1}^k a_j 3^j + 2}{3^{k+1}}\right),$$

i.e. one of the intervals taken out in step $k+1$ of the process. Now, define

$$E = \left[\bigcup_{n=1}^{\infty}\left[\bigcup_{l(a)=n} E_a\right]\right] \cup E_\varnothing$$

where $l(a)$ means the length of the string a. As each E_a and $E_\varnothing$ is open, we have that E is an open set. The Cantor set is $\mathfrak{C} = [0,1]\backslash E$. Thus $\mathfrak{C}$ is a closed set.

Now, take an $\epsilon > 0$, and define the following sets. We define

$$F_\varnothing = \left(\frac{1}{3} + \frac{\epsilon}{4}, \frac{2}{3} - \frac{\epsilon}{4}\right) \subset E_\varnothing.$$

And similarly,

$$F_0 = \left(\frac{1}{9} + \frac{\epsilon}{16}, \frac{2}{9} - \frac{\epsilon}{16}\right) \subset E_0.$$

Continuing on, for a string of length k denoted by a, define

$$F_a = \left(\frac{\sum_{j=1}^{k} a_j 3^j + 1}{3^{k+1}} + \frac{\epsilon}{4^{k+1}}, \frac{\sum_{j=1}^{k} a_j 3^j + 2}{3^{k+1}} - \frac{\epsilon}{4^{k+1}}\right) \subset E_a.$$

And lastly, define

$$F = \left[\bigcup_{n=1}^{\infty}\left[\bigcup_{l(a)=n} F_a\right]\right] \cup F_\varnothing \subset E.$$

Thus $\mathfrak{C} = [0,1]\backslash E \subset [0,1]\backslash F$, i.e. $[0,1]\backslash F$ is a cover of $\mathfrak{C}$ by intervals.

Considering the empty set, $\varnothing$, to be a string of length zero, we have the following. For strings of length k, there are 2^k many sets F_a where $a \in \{0,2\}^k$. And the length of F_a is

$$v(F_a) = \frac{1}{3^{k+1}} - \frac{\epsilon}{2 \cdot 4^k}.$$

Thus the total length of all intervals F_a coming from strings a of length k is

$$2^k v(F_a) = \frac{2^k}{3^{k+1}} - \frac{\epsilon}{2^{k+1}}.$$

Thus, this gives that the length of F is precisely the following

$$\begin{aligned} v(F) &= \sum_{k=1}^{\infty} 2^k v(F_a) = \sum_{k=0}^{\infty} \frac{2^k}{3^{k+1}} - \sum_{k=0}^{\infty} \frac{\epsilon}{2^{k+1}} \\ &= \frac{1}{3}\left[\frac{1}{1-\frac{2}{3}}\right] - \frac{\epsilon}{2}\left[\frac{1}{1-\frac{1}{2}}\right] \\ &= 1 - \epsilon. \end{aligned}$$

And thus the length of $[0,1]\backslash F$ is

$$v([0,1]\backslash F) = v([0,1]) - v(F) = 1 - (1-\epsilon) = \epsilon.$$

Thus the Cantor set can be covered by a collection of intervals of arbitrarily small size. Thus $\mathfrak{C}$ is measure 0.

Integration: Day 5

Definition 1: The oscillation of a bounded function f at $x = a$, denoted by $o(f, a)$ is defined by

$$o(f, a) = \lim_{\delta \to 0}[M(a, f, \delta) - m(a, f, \delta)],$$

where $M(a, f, \delta) = \sup\{f(x) \mid x \in [c, d],\ |x-a| < \delta\}$ and $m(a, f, \delta) = \inf\{f(x) \mid x \in [c, d],\ |x-a| < \delta\}$.

Note: For a bounded function f on an interval $[c, d]$, the above limit always exists. The sets $\{f(x) \mid x \in [c, d],\ |x - a| < \delta\}$ and $\{f(x) \mid x \in [c, d],\ |x - a| < \delta\}$ are clearly bounded above and below as f is, thus $M(a, f, \delta)$ and $m(a, f, \delta)$ exist for any δ. And for $\delta_1 < \delta$, we have that

$$M(a, f, \delta) \geq M(a, f, \delta_1)$$
$$m(a, f, \delta) \leq m(a, f, \delta_1).$$

Thus $M - m$ decreases as δ decreases, and $M - m \geq 0$, thus the monotone convergence theorem implies that the limit in the definition of oscillation always exists.

Theorem 1: A bounded function f on $[c, d]$ is continuous at $x = a \in [c, d]$ if and only if $o(f, a) = 0$.

Proof. $\Rightarrow$ Let f be continuous at $x = a$. Thus for a given $\epsilon > 0$, there is a $\delta > 0$ such that

$$|f(x) - f(a)| < \frac{\epsilon}{2} \quad \text{for} \quad |x - a| < \delta.$$

What this is really saying is that the image of the δ neighborhood about a is contained in the $\frac{\epsilon}{2}$ neighborhood about $f(a)$, or written more compactly,

$$f\left((a - \delta, a + \delta)\right) \subseteq \left(f(a) - \frac{\epsilon}{2}, f(a) + \frac{\epsilon}{2}\right)$$

As $M(a, f, \delta) = \sup\{f(x) \mid |x - a| < \delta\}$, and

$$f(x) \in \left(f(a) - \frac{\epsilon}{2}, f(a) + \frac{\epsilon}{2}\right), \quad \text{for} \quad |x - a| < \delta,$$

we really have that

$$M(a, f, \delta) \in \left[f(a) - \frac{\epsilon}{2}, f(a) + \frac{\epsilon}{2}\right].$$

Thus $|M(a, f, \delta) - f(a)| \leq \frac{\epsilon}{2}$. The same argument shows that $|m(a, f, \delta) - f(a)| \leq \frac{\epsilon}{2}$. Thus we have that

$$\begin{aligned} M(a, f, \delta) - m(a, f, \delta) &= M(a, f, \delta) - f(a) + f(a) - m(a, f, \delta) \\ &= |M(a, f, \delta) - f(a)| + |f(a) - m(a, f, \delta)| \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

As $M(a, f, \delta') - m(a, f, \delta')$ decreases for $\delta' < \delta$ and that this process can be done for any arbitrary value of ϵ, we have that

$$o(f, a) = \lim_{\delta \to 0}[M(a, f, \delta) - m(a, f, \delta)] = 0.$$

$\Leftarrow$ Now, assume that $o(f,a) = 0$. As the oscillation is defined as a limit, we have that for a given $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|M(a,f,\delta') - m(a,f,\delta') - 0| < \epsilon, \quad \text{for} \quad |\delta' - 0| < \delta.$$

And for all $x \in (a-\delta', a+\delta')$, we have

$$m(a,f,\delta') \leq f(x), f(a) \leq M(a,f,\delta')$$

which implies that $|f(x) - f(a)| \leq M(a,f,\delta') - m(a,f,\delta') < \epsilon$. Thus, for $\epsilon > 0$ there exists $\delta' > 0$ such that $|x-a| < \delta'$ implies $|f(x) - f(a)| < \epsilon$. Thus, as ϵ can be taken arbitrarily, we have that f is continuous at $x = a$. □

The following two theorems will be used in our main result classifying exactly what bounded functions are Riemann integrable.

Theorem 2: Let $[a,b] \subseteq \mathbb{R}$ be closed. If $f : [a,b] \to \mathbb{R}$ is any bounded function, and $\epsilon > 0$, then the set $B = \{x \in [a,b] \mid o(f,x) \geq \epsilon\}$ is closed.

Proof. We will prove our result by showing that $\mathbb{R}\backslash B$ is an open set. If $x \in \mathbb{R}\backslash B$, then one of two things is true. Either $x \notin [a,b]$ or $x \in [a,b]$ and $o(f,x) < \epsilon$.

- *Case 1:* If $x \notin [a,b]$, then as $[a,b]$ is a closed set we have that $\mathbb{R}\backslash[a,b]$ is open. Thus, there exists an open interval $U = (c,d)$ containing x with

  $$x \in (c,d) \subseteq \mathbb{R}\backslash[a,b].$$

- *Case 2:* Assume that $x \in [a,b]$ and $o(f,x) < \epsilon$. Call $o(f,x) = \frac{\gamma}{2}$. By the definition of oscillation for $\frac{\gamma}{2}$, there is a $\delta > 0$ such that

  $$|M(x,f,\delta') - m(x,f,\delta') - o(f,x)| < \frac{\gamma}{2} \quad \text{for} \quad \delta' < \delta.$$

  Thus

  $$M(x,f,\delta') - m(x,f,\delta') < o(f,x) + \frac{\gamma}{2} < \gamma < \epsilon.$$

  Let us take any point y in the δ' neighborhood of x, i.e. $y \in (x-\delta', x+\delta')$. Define

  $$\delta'' = \min\{|y - (x-\delta')|, |y - (x+\delta')|\}$$

  We have defined δ'' in such a manner that the δ'' neighborhood of y is contained in the δ' neighborhood of x, i.e.

  $$(y-\delta'', y+\delta'') \subseteq (x-\delta', x+\delta').$$

  This then implies that $M(y,f,\delta'') \leq M(x,f,\delta')$ and $m(y,f,\delta'') \geq M(x,f,\delta')$, thus

  $$M(y,f,\delta'') - m(y,f,\delta'') \leq M(x,f,\delta') - m(x,f,\delta') < \epsilon.$$

  This implies that $o(f,y) < \epsilon$. In particular, we have shown that every point of the δ' neighborhood of x has the property that $o(f,y) < \epsilon$. Thus $(x-\delta', x+\delta') \subseteq \mathbb{R}\backslash B$.

In either case, we have shown for each $x \in \mathbb{R}\backslash B$, the existence of a neighborhood containing x that in entirely contained in $\mathbb{R}\backslash B$. Thus $\mathbb{R}\backslash B$ is an open set. □

Theorem 3: Let A be a closed interval (or a finite collection of closed intervals) and let $f : A \to \mathbb{R}$ be a bounded function such that $o(f, x) < \epsilon$ for a given ϵ and all $x \in A$. Then there exists a partition of P of A with

$$U(f, P) - L(f, P) < \epsilon v(A),$$

where $v(A)$ is the length of A (or the sum of the lengths of the finite intervals making up A.)

Proof. For each $x \in A$, there is an interval $U_x = (x - \delta_x, x + \delta_x)$ containing x with

$$M_{U_x}(f) - m_{U_x}(f) = M(x, f, \delta_x) - m(x, f, \delta_x) < \epsilon.$$

Where $M_{U_x}(f)$ is shorthand for $M(x, f, \delta_x)$. This can be done as $o(f, x) < \epsilon$ for all $x \in A$.

Now the collection of open sets $\{U_x\}_{x\in A}$ is clearly an open cover of A. As A is closed and bounded, and hence compact, there is a finite subcover $\{U_{x_k}\}_{k=1}^n$ that still covers A. Now form a partition of A in the following manner. The partition P will contain the endpoints of A (or the endpoints of all the finite intervals making up A), and P contains the endpoints of the U_{x_k}, $1 \le k \le n$ that are contained in A. Thus each subinterval $[x_{j-1}, x_j]$ of P is contained in at least one U_{x_k} for some k. Thus

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{j=1}^{N} (M_j - m_j)\Delta x_j < \epsilon \sum_{k=j}^{N} \Delta x_j \\ &= \epsilon v(A). \end{aligned}$$

□

Comment: The finite subcover did not obviously come into play in the computation above. Why was it even necessary? Well, we used it in two places. One, to guarantee that $M_k - m_k < \epsilon$ as each subinterval from P was contained in at least one U_{x_j}. Two, if we did not refine from the infinite collection $\{U_x\}_{x\in A}$ to the finite the collection $\{U_{x_k}\}_{k=1}^n$, then we could not form P in the manner made above. (Remember, a partition must be a finite collection of points.)

Okay, back to big picture time. We now have built all the necessary concepts and tools to come to the following result.

Theorem 4: Let $[a, b]$ be a closed interval, and $f : [a, b] \to \mathbb{R}$ a bounded function. Define $B = \{x \in [a, b] \mid o(f, x) > 0\}$, i.e. the collection of the discontinuities of f contained in $[a, b]$. Then $f \in R$ if and only if B is a set of measure 0.

Proof. $\Leftarrow$ Suppose that B has measure 0. Also, define $M = \sup_{x\in[a,b]} |f(x)|$. Let $\epsilon > 0$, and define $B_\epsilon = \{x \in [a, b] \mid o(f, x) \ge \epsilon\}$. Thus $B_\epsilon \subseteq B$, which implies that B_ϵ is also of measure 0. From Theorem 2, we have that B_ϵ is closed. Thus, as B_ϵ is a closed subset of the compact set $[a, b]$, we have that B_ϵ is compact. Thus B_ϵ is content 0 as it is both measure 0 and compact. Thus, there exists a finite collection of open intervals $\{U_i\}_{i=1}^n$ that cover B_ϵ and $\sum_{k=1}^n v(U_k) < \epsilon$.

Now, define $A = [a, b] \setminus \bigcup_{k=1}^n U_k$. As U_k is open, we have that A is closed, and as we are taking a finite number of open intervals from $[a, b]$, we have that A is a finite union of closed intervals. And as $\{U_k\}_{k=1}^n$ cover the set B_ϵ, we have $A \cap B_\epsilon = \varnothing$. Thus for all $x \in A$, we have $o(f, x) < \epsilon$. Thus by Theorem 3, there exists are partition P' of A with

$$U(f, P') - L(F, P') < \epsilon v(A) \le \epsilon(b - a),$$

where the last inequality follows from $A \subseteq [a, b]$.

We now extend P' from a partition of A to a partition of $[a, b]$, which we will denote as P. The partition P will be the partition P' plus a, b, (if they are not already in P') plus the endpoints of the intervals U_k, $1 \leq k \leq n$. To be more formal

$$P = P' \cup \{a, b\} \cup \{\text{endpoints of } U_k, 1 \leq k \leq n\}.$$

Thus P contains two types of subintervals.

1. Subintervals formed by the endpoints of a specific U_k. Call the collection of these subintervals S_1.
2. Subintervals from the partition P' on A. Call the collection of these subintervals S_2.

If we write S to mean an arbitrary subinterval of P, then we have the following.

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{S \in P} (M_S - m_S) v(S) \\ &= \sum_{S \in S_1} (M_S - m_S) v(S) + \sum_{S \in S_2} (M_S - m_S) v(S) \\ &\leq 2M \sum_{k=1}^{n} v(U_k) + [U(f, P') - L(f, P')] \\ &< 2M\epsilon + (b - a)\epsilon. \end{aligned}$$

As ϵ can be taken arbitrarily, we have that $f \in R$.

$\Rightarrow$ Assume that $f \in R$. Keeping with the notation of the first part of this proof, for each $n \in \mathbb{N}$ define

$$B_{\frac{1}{n}} = \{x \in [a, b] \mid o(f, x) \geq \frac{1}{n}\}.$$

Thus, a quick application of the archimedean property gives us that

$$B = \bigcup_{n=1}^{\infty} B_{\frac{1}{n}}.$$

And so, to prove our result, we will show $B_{\frac{1}{n}}$ is content 0 for each $n \in \mathbb{N}$. (Actually measure 0 is enough, but remember, as the $B_{\frac{1}{n}}$ are closed and hence compact, content 0 and measure 0 mean the same thing in this context.)

For a given $\epsilon > 0$, as $f \in R$, there exists a partition $P = \{x_1, x_2, \ldots, x_n\}$ of $[a, b]$ such that

$$U(f, P) - L(f, P) < \frac{\epsilon}{n}.$$

Define the following, take $\mathcal{S} = \{[x_{j-1}, x_j] \subset [a, b] \mid [x_{j-1}, x_j] \cap B_{\frac{1}{n}} \neq \varnothing\}$, i.e. the collection of subintervals from P that have a nontrivial intersection with $B_{\frac{1}{n}}$. Thus, $\mathcal{S}$ is a cover of $B_{\frac{1}{n}}$ by closed intervals. Now for any subinterval U_i in the collection $\mathcal{S}$, we have that $M_{U_i} - m_{U_i} \geq \frac{1}{n}$ as U_i has

nonempty intersection with $B_{\frac{1}{n}}$. Thus

$$\begin{aligned}\frac{1}{n}\sum_{U_i\in\mathcal{S}} v(U_i) &\leq \sum_{U_i\in\mathcal{S}} [M_{U_i} - m_{U_i}]v(U_i)\\ &\leq \sum_{U_i\in P} [M_{U_i} - m_{U_i}]v(U_i)\\ &= U(f,P - L(f,P) < \frac{\epsilon}{n},\end{aligned}$$

where the second inequality follows from summing over all subintervals in the partition instead of just summing over all subintervals in the collection $\mathcal{S}$. Thus

$$\sum_{U_i\in\mathcal{S}} v(U_i) < \epsilon.$$

As the number of subintervals in $\mathcal{S}$ is finite and ϵ can be taken arbitrarily, we have that $B_{\frac{1}{n}}$ is content 0. Thus B is measure 0. □

Integration: Day 6

We begin today by collecting some results and deducing a result about the discontinuities of a monotonic function that will be useful in extending our result from the end of lecture 5 from the Riemann integrability case to the Riemann-Stieltjes case.

Definition 1: Let f be a function defined on (a,b). Consider any point $a \leq x < b$. We write $f(x+) = q$ if $f(t_n) \to q$ as $n \to \infty$ for all sequences $t_n \in (x,b)$ with $t_n \to x$. Similarly, for $a < x \leq b$, we write $f(x-) = q$ if $f(t_n) \to q$ as $n \to \infty$ for all sequences (a,x) with $t_n \to x$.

In particular, this definition is just making formal the notion of limit of f at x from the right and left respectively.

Definition 2: If f is discontinuous at a point a

- if $f(a+)$, $f(a-)$ exist, then f has a discontinuity of the first kind, or simple discontinuity.
- otherwise, f has a discontinuity of the second kind.

Note that there are two ways that simple discontinuities exist at a point a, either $f(a+) \neq f(a-)$ or $f(a+) = f(a-) \neq f(a)$.

Theorem 1: Assume that f is monotonically increasing on (a,b), then $f(x+)$ and $f(x-)$ exist at every point $x \in (a,b)$. More precisely,

$$\sup_{a<t<x} f(t) = f(x-) \leq f(x) \leq f(x+) = \inf_{x<t<b} f(t).$$

And furthermore, if $a < x < y < b$, then $f(x+) \leq f(y-)$.

Note: We are implicitly assuming that f is defined on the entirety of the interval (a,b).

Proof. As f is a monotonically increasing function, the set

$$A = \{f(t) \mid a < t < x\}$$

is bounded above by $f(x)$. Thus $\sup A$ exists, and $\sup A \leq f(x)$. By definition of the supremum, for any $\epsilon > 0$, there must be an element of A that is greater than $\sup A - \epsilon$. As f is monotonically increasing this implies the existence of a $\delta > 0$ such that the following holds,

$$\sup A - \epsilon < f(x - \delta) \leq \sup A.$$

This gives, again by the monotonicity of f, for any t with $x - \delta < t < x$ that

$$\sup A - \epsilon \leq f(x - \delta) \leq f(t) \leq \sup A.$$

And so $|f(t) - \sup A| < \epsilon$ for all $t \in (x - \delta, x)$. As ϵ can be taken arbitrarily, this argument shows that $f(t_n) \to \sup A$ for all sequences $t_n \in (a, x)$ with $t_n \to x$. (As for every ϵ, there exists a δ, which implies the existence of an $N \in \mathbb{N}$, such that for $m > N$, we have have $t_m \in (x - \delta, x)$, and thus $|f(t_m) - \sup A| < \epsilon$. Thus $f(x-)$ exists and in fact we have shown that $f(x-) = \sup A$. Thus

$$\sup_{a<t<x} f(t) = f(x-).$$

The argument for showing why $\inf_{x<t<b} f(t) = f(x+)$ is similar.

Now, take x, y with $a < x < y < b$, now we have the following

$$f(x+) = \inf_{x<t<b} f(t) = \inf_{x<t<y} f(t)$$
$$f(y-) = \sup_{a<t<y} f(t) = \sup_{x<t<y} f(t)$$

as f is monotonically increasing. This clearly implies that $f(x+) \leq f(y-)$. □

Note: A similar theorem holds for monotonically decreasing functions. Just interchange sup and inf above and it is pretty much the same argument.

Corollary : For monotonic functions f on an interval (a, b), we have that f has no discontinuities of the second kind.

This corollary immediately leads us to a very interesting result.

Theorem 2: Let f be monotonic (increasing/decreasing) on (a, b). Then the set of points at which f is discontinuous is at most countable.

Proof. Without loss of generality suppose that f is increasing. Let us call

$$S = \{x \in (a, b) \mid o(f, x) > 0\} = \{x \in (a, b) \mid f \text{ is discontinous at } x\}$$

the set of discontinuities of f on (a, b). As f is defined on the entirety of (a, b), i.e. as $f(x)$ exists for all $x \in (a, b)$, we have that for all $x \in (a, b)$

$$f(x-) \leq f(x) \leq f(x+).$$

Thus, if f is discontinuous at x, then it must be that $f(x-) < f(x+)$. (In other words, it is impossible to have a discontinuity in f at x due to $f(x-) = f(x+) \neq f(x)$.) So, for all $x \in S$, we have $f(x-) < f(x+)$. Now, due to density of the rational numbers, $\mathbb{Q}$, there exists a $r \in \mathbb{Q}$ with $f(x-) < r < f(x+)$. As we associate this rational number r with $x \in S$, we have really created a function $r : S \to \mathbb{Q}$ that associates a rational number to each $x \in S$ with the property that

$$f(x-) < r(x) < f(x+).$$

Now, for $x_1, x_2 \in S$, we can assume without loss of generality that $x_1 < x_2$, but then

$$r(x_1) < f(x_1+) \leq f(x_2-) < r(x_2).$$

Thus unique elements in S map to unique elements of $\mathbb{Q}$, i.e. r is injective. Thus $|S|$ the cardinality of S must be less than or equal to the cardinality of $\mathbb{Q}$, which is countable. □

We end our little tangent here. The primary use of the above result for us is that our weights α in the Riemann–Stieltjes theory can only have a countable number of discontinuities.

For the sake of completeness we present the statement of the following two results that are the analogue of Theorem 5.4 for Riemann–Stieltjes integration. These theorems are presented without proof as the results are beyond the scope of the course.

Theorem 3: Let $f : [a, b] \to \mathbb{R}$ be a bounded function, and let $\alpha : [a, b] \to \mathbb{R}$ be monotonically increasing. Define $E = \{x \in [a, b] \mid o(f, x) > 0\}$ and $F = \{x \in [a, b] \mid o(\alpha, x) > 0\}$ to be the points of discontinuity of f and α respectively in $[a, b]$. We know that F is countable.

- If E is countable and $E \cap F = \varnothing$, then $f \in R(\alpha)$.
- If $E \cap F = S$ and for all $s \in S$ either
  1. $f(s-) = f(s)$ and $\alpha(s+) = \alpha(s)$ or
  2. $f(s+) = f(s)$ and $\alpha(s-) = \alpha(s)$

  then $f \in R(\alpha)$.

Note: Be careful however. Using the notation above, if $E \cap F = \{a\}$, and $\alpha(a-) < \alpha(a) < \alpha(a+)$, then $f \notin R(\alpha)$. In other words, if f and α have even just one point of discontinuity in common, and the left and right limits of α do no equal α's functional value at a, then f is not Riemann integrable with respect to α (this also fails if the left and right limits of f do not agree with f's value at a). This is the exact case of a homework problem you have this week.

OK, at this point, I believe we have sufficiently answered the question of what it means to be integrable. Now, let us move on to some of the important properties of the integral.

Theorem 4:

a). (*Linearity of the integral with respect to f*) If $f_1 \in R(\alpha)$ and $f_2 \in R(\alpha)$, then $f_1 + f_2 \in R(\alpha)$, also $cf_1 \in R(\alpha)$ for all $c \in \mathbb{R}$, and

$$\int_a^b (f_1 + f_2)d\alpha = \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha$$
$$\int_a^b cf_1 d\alpha = c\int_a^b f_1 d\alpha$$

b). (*Monotonicity of the integral*) If $f_1(x) \leq f_2(x)$ on $[a, b]$, then

$$\int_a^b f_1 d\alpha \leq \int_a^b f_2 d\alpha$$

c). (*Additivity on intervals*) If $f \in R(\alpha)$ on $[a,b]$ and if $a < c < b$, then $f \in R(\alpha)$ on $[a,c]$ and on $[c,b]$, and in this case

$$\int_a^b f d\alpha = \int_a^c f d\alpha + \int_c^b f d\alpha.$$

d). (*Boundedness of the integral*) If $f \in R(\alpha)$ on $[a,b]$ and if $|f(x)| \leq M$ on $[a,b]$, then

$$\left| \int_a^b f d\alpha \right| \leq M[\alpha(b) - \alpha(a)].$$

e). (*Linearity of the integral with respect to* α) If $f \in R(\alpha_1)$ and $f \in R(\alpha_2)$, then $f \in R(\alpha_1 + \alpha_2)$ and

$$\int_a^b f d(\alpha_1 + \alpha_2) = \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2.$$

Also, if $f \in R(\alpha)$ and $c \in \mathbb{R}$, then $f \in R(c\alpha)$ and

$$\int_a^b f d(c\alpha) = c \int_a^b f d\alpha.$$

We will present the proof of each part seperately (to avoid to much indentation).

Part a).

Proof. Let us first show the homogeneity of the integral. We will show this in two parts. Let $c \in \mathbb{R}$ with $c > 0$, and take $f \in R(\alpha)$. As $f \in R(\alpha)$, for $\epsilon > 0$ there exists a partition P of $[a,b]$ such that

$$U(f,P,\alpha) - L(f,P,\alpha) < \frac{\epsilon}{c}.$$

Multiplying f by a positive constant c will not effect the value of x at which cf attains a maximum or minimum, it will only scale the maximum and minimum values by c. Thus, $U(cf,P,\alpha) = cU(f,P,\alpha)$, and $L(cf,P,\alpha) = cL(f,P,\alpha)$, which implies that

$$U(cf,P,\alpha) - L(cf,P,\alpha) = c[U(f,P,\alpha) - L(f,P,\alpha)] < c\left[\frac{\epsilon}{c}\right] = \epsilon.$$

Thus $cf \in R(\alpha)$ for $c > 0$.

Now, we will show if $c = -1$ that $-f = cf \in R(\alpha)$. Once again, there is a partition P with the property that

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

For any subinterval $[x_{k-1}, x_k]$ of P, we have the following

$$\begin{aligned} M_k' &= \sup_{x \in [x_{k-1},x_k]} (-f(x)) = -\inf_{x \in [x_{k-1},x_k]} f(x) = -m_k \\ m_k' &= \inf_{x \in [x_{k-1},x_k]} (-f(x)) = -\sup_{x \in [x_{k-1},x_k]} f(x) = -M_k. \end{aligned}$$

Thus $U(-f,P,\alpha) = -L(f,P,\alpha)$ and $L(-f,P,\alpha) = -U(f,P,\alpha)$. Thus,

$$\begin{aligned} U(-f,P,\alpha) - L(-f,P,\alpha) &= -L(f,P,\alpha) - (-U(f,P,\alpha)) \\ &= U(f,P,\alpha) - L(f,P,\alpha) < \epsilon. \end{aligned}$$

Thus $-f \in R(\alpha)$. Between this result and the previous one, we have that $cf \in R(\alpha)$ for all $c \in \mathbb{R}$. And lastly,

$$\begin{aligned}\int_a^b cf d\alpha &= \sup L(cf, P, \alpha) = \sup[cL(f, P, \alpha)] \\ &= c[\sup L(f, P, \alpha)] = c\int_a^b f d\alpha\end{aligned}$$

where the suprema are taken over all partitions of $[a, b]$.

Now, take $f = f_1 + f_2$. For any partition P of the interval $[a, b]$, for a subinterval $[x_{k-1}, x_k]$ of P, we have the following

$$\begin{aligned}m_k(f_1) + m_k(f_2) &\leq m_k(f) \\ M_k(f) &\leq M_k(f_1) + M_k(f_2)\end{aligned}$$

where $m_k(g)$, $M_k(g)$ mean the minimum and maximum of g on the subinterval $[x_{k-1}, x_k]$ respectively. It should be noted that equality only occurs in the inequalities above if f_1 and f_2 achieve their maximum or minimum at the same point in the subinterval. This implies that the following string of inequalities is true.

$$\begin{aligned}L(f_1, P, \alpha) + L(f_2, P, \alpha) &\leq L(f, P, \alpha) \\ &\leq U(f, P, \alpha) \leq U(f_1, P, \alpha) + U(f_2, P, \alpha).\end{aligned}$$

As $f_1, f_2 \in R(\alpha)$, there exists partitions P_1 and P_2 such that

$$\begin{aligned}U(f_1, P_1, \alpha) - L(f_1, P_1, \alpha) &< \frac{\epsilon}{2} \\ U(f_2, P_2, \alpha) - L(f_2, P_2, \alpha) &< \frac{\epsilon}{2}\end{aligned}$$

and these inequalities still hold if we pass to the refinement $P = P_1 \cup P_2$ of the two partitions. This paired with the inequality above gives that

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

Thus $f_1 + f_2 \in R(\alpha)$.

Now to prove the statement about the sum of the integrals. As the integral is defined as the infimum of the upper sums, we have

$$U(f_j, P_j, \alpha) \leq \int_a^b f_j d\alpha + \frac{\epsilon}{2} \quad \text{for } j \in \{1, 2\},$$

and this statement still holds when passing through to the refinement $P = P_1 \cup P_2$. Thus

$$\begin{aligned}\int_a^b f d\alpha &\leq U(f, P, \alpha) \leq U(f_1, P, \alpha) + U(f_2, P, \alpha) \\ &\leq \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha + \epsilon.\end{aligned}$$

As this holds for arbitrary ϵ, we have that

$$\int_a^b f d\alpha \leq \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha.$$

Now replacing all the functions with their negatives $-f, -f_1, -f_2$ gives

$$-\int_a^b f d\alpha \leq -\int_a^b f_1 d\alpha - \int_a^b f_2 d\alpha$$
$$\int_a^b f d\alpha \geq \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha$$

Putting these two inequalities together gives the result.

$$\int_a^b (f_1 + f_2) d\alpha = \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha$$

□

Part b).

Proof. Without loss of generality we only need to show that if $f \geq 0$, then $\int_a^b f d\alpha \geq 0$. As $f \geq 0$, for any partition P of $[a, b]$ and further for any given subinterval $[x_{k-1}, x_k]$ of P, we clearly have $M_k, m_k \geq 0$. Thus $L(f, P, \alpha) \geq 0$. So,

$$\int_a^b f d\alpha = \sup L(f, P, \alpha) \geq L(f, P, \alpha) \geq 0.$$

Now, for the general proof. If $f_1 \leq f_2$, then $f_2 - f_1 \geq 0$, so

$$\int_a^b f_2 d\alpha - \int_a^b f_1 d\alpha = \int_a^b (f_2 - f_1) d\alpha \geq 0$$
$$\int_a^b f_2 d\alpha \geq \int_a^b f_1 d\alpha.$$

□

Part c).

Proof. Take $f \in R(\alpha)$. Thus for $\epsilon > 0$, there exists a partition P such that

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

If $c \in P$, then keep P the same. If not, then refine P by c, i.e. $P' = P \cup \{c\}$. And the same inequality will hold for the refinement

$$U(f, P', \alpha) - L(f, P', \alpha) < \epsilon.$$

The point of this is without loss of generality, we can assume that P is a partition that satisfies the inequality above and that $c \in P$.

Now write P as $P = P_1 \cup P_2$, where $P_1 = P \cap [a, c]$ and $P_2 = P \cap [c, b]$, and thus we see that P_1 is a partition of $[a, c]$ and P_2 is a partition of $[c, b]$. As

$$U(f, P, \alpha) = U(f, P_1, \alpha) + U(f, P_2, \alpha)$$
$$L(f, P, \alpha) = L(f, P_1, \alpha) + L(f, P_2, \alpha)$$

Thus,

$$U(f, P_j, \alpha) - L(f, P_j, \alpha) < \epsilon \text{ for } j \in \{1, 2\}.$$

So, $f \in R(\alpha)$ on $[a, c]$ and $f \in R(\alpha)$ on $[c, b]$.

Now, by the definition of integrability of f, take P to be a partition of $[a, b]$ with the property

$$L(f, P, \alpha) \geq \int_a^b f d\alpha - \frac{\epsilon}{2}, \quad U(f, P, \alpha) \leq \int_a^b f d\alpha + \frac{\epsilon}{2}$$

And take R and S to be partitions of $[a, c]$ and $[c, b]$ given by $R = P \cap [a, c]$ and $S = P \cap [c, b]$. And R and S have the property

$$\begin{aligned} U(f, R, \alpha) &\leq \int_a^c f d\alpha + \frac{\epsilon}{4}, \quad L(f, R, \alpha) \geq \int_a^c f d\alpha - \frac{\epsilon}{4} \\ U(f, S, \alpha) &\leq \int_c^b f d\alpha + \frac{\epsilon}{4}, \quad L(f, S, \alpha) \geq \int_c^b f d\alpha - \frac{\epsilon}{4}. \end{aligned}$$

(This can be done because refining maintains inequalities) And then we have the following

$$\begin{aligned} \int_a^b f d\alpha - \frac{\epsilon}{2} &\leq L(f, P, \alpha) \leq U(f, P, \alpha) \\ &= U(f, R, \alpha) + U(f, S, \alpha) \\ &\leq \int_a^c f d\alpha + \int_c^b f d\alpha + \frac{\epsilon}{2}. \end{aligned}$$

And so

$$\int_a^b f d\alpha - \int_a^c f d\alpha - \int_c^b f d\alpha < \epsilon.$$

Similarly,

$$\begin{aligned} \int_a^c f d\alpha + \int_c^b f d\alpha - \frac{\epsilon}{2} &\leq L(f, R, \alpha) + L(f, S, \alpha) \\ &= L(f, P, \alpha) \leq U(f, P, \alpha) \\ &\leq \int_a^b f d\alpha + \frac{\epsilon}{2}, \end{aligned}$$

which gives us the reverse inequality, hence

$$\left| \int_a^b f d\alpha - \int_a^c f d\alpha - \int_c^b f d\alpha \right| < \epsilon.$$

As ϵ can be taken arbitrarily, we have

$$\int_a^b f d\alpha = \int_a^c f d\alpha + \int_c^b f d\alpha.$$

□

Part d).

Proof. For a bounded set A, we have the result that $|\sup A| \leq \sup |A|$, where $|A| = \{|x| \mid x \in A\}$. And thus

$$\begin{aligned}\left|\int_a^b f d\alpha\right| = |\sup L(f,P,\alpha)| &\leq \sup |L(f,P,\alpha)| \\ &\leq |L(f,P,\alpha)| \leq \sum_{k=1}^{n} |m_k| \Delta\alpha_k.\end{aligned}$$

As $|f(x)| \leq M$ for all $x \in [a,b]$, we have that $|m_k| \leq M$ for all $k \in \{1,2,\ldots,n\}$. Thus,

$$\left|\int_a^b f d\alpha\right| \leq M \sum_{k=1}^{n} \Delta\alpha_k = M[\alpha(b) - \alpha(a)].$$

□

Part e).

Proof. Assume that $f \in R(\alpha_1)$ and $f \in R(\alpha_2)$. For an $\epsilon > 0$ there exists partitions P_1 and P_2 such that

$$\begin{aligned}U(f,P_1,\alpha_1) - L(f,P_1,\alpha_1) &< \frac{\epsilon}{2} \\ U(f,P_2,\alpha_2) - L(f,P_2,\alpha_2) &< \frac{\epsilon}{2}.\end{aligned}$$

And these inequalities still hold for the common refinement $P = P_1 \cup P_2$. Calling $\alpha = \alpha_1 + \alpha_2$, we have

$$U(f,P,\alpha_1) + U(f,P,\alpha_2) = \sum_{k=1}^{n} M_k(\Delta\alpha_{1,k} + \Delta\alpha_{2,k}) = \sum_{k=1}^{n} M_k \Delta\alpha_k = U(f,P,\alpha).$$

and a similar result holds for the lower sums. Thus,

$$U(f,P,\alpha) - L(f,P,\alpha) < \epsilon,$$

so $f \in R(\alpha_1 + \alpha_2)$.

Now for a partition P, using similar type inequalities from part a) and part c), we have

$$\begin{aligned}\int_a^b f d\alpha \leq U(f,P,\alpha) &= U(f,P,\alpha_1) + U(f,P,\alpha_2) \\ &\leq \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2 + \epsilon\end{aligned}$$

Similarly,

$$\begin{aligned}\int_a^b f d\alpha \geq L(f,P,\alpha) &= L(f,P,\alpha_1) + L(f,P,\alpha_2) \\ &\geq \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2 - \epsilon.\end{aligned}$$

Thus

$$\left| \int_a^b f d\alpha - \int_a^b f d\alpha_1 - \int_a^b f d\alpha_2 \right| < \epsilon.$$

As ϵ is arbitrary, we have

$$\int_a^b f d(\alpha_1 + \alpha_2) = \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2.$$

The homogeneity is proven in almost an identical manner as in part a). □

Integration: Day 7

We begin today with a lemma.

Lemma: Suppose $f \in R(\alpha)$ on $[a, b]$ and suppose $m \leq f(x) \leq M$ for $x \in [a, b]$. If ψ is continuous on $[m, M]$, and h is defined as $h(x) = \psi(f(x))$, then $h \in R(\alpha)$ on $[a, b]$.

Proof. Let $\epsilon > 0$. As ψ is continuous on the compact interval $[m, M]$, ψ is actually uniformly continuous. Thus, there is a $\delta > 0$ such that for all $s, t \in [m, M]$ with $|s - t| < \delta$, we have $|\psi(t) - \psi(s)| < \epsilon$. Without loss of generality, we can also assume that $\delta < \epsilon$. (The result we gain from uniform continuity holds with smaller values of δ).

As $f \in R(\alpha)$ there exists a partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ such that

$$U(f, P, \alpha) - L(f, P, \alpha) < \delta^2.$$

For what follows we will use the following notation

$$M_k = \sup_{x \in [x_{k-1}, x_k]} f(x), \quad m_k = \inf_{x \in [x_{k-1}, x_k]} f(x).$$

and

$$M_k' = \sup_{x \in [x_{k-1}, x_k]} h(x), \quad m_k' = \inf_{x \in [x_{k-1}, x_k]} h(x).$$

There are two types of subintervals $[x_{k-1}, x_k]$ within the partition P.

- Subintervals with $M_k - m_k < \delta$, we will call this collection of subintervals S_1.
- Subintervals with $M_k - m_k \geq \delta$, we will call this collection of subintervals S_2.

For k such that $[x_{k-1}, x_k] \in S_1$, we have that $M_k - m_k < \delta$. Thus for all $x, y \in [x_{k-1}, x_k]$, as $m_k \leq f(x), f(y) \leq M_k$, we have that

$$|f(x) - f(y)| < \delta$$

and thus $M_k' - m_k' < \epsilon$ from the uniform continuity of ψ.

For k such that $[x_{k-1}, x_k] \in S_2$, if we denote $K = \sup_{x \in [m,M]} |\psi(x)|$, then we have for any of these subintervals that $M_k' - m_k' \leq K$. Also, we have the following string of inequalities

$$\delta \sum_{k \in S_2} \Delta\alpha_k \leq \sum_{k \in S_2} (M_k - m_k)\Delta\alpha_k < U(f, P, \alpha) - L(f, P, \alpha) < \delta^2.$$

Thus, by dividing δ from both sides, we have that $\sum_{k\in S_2}\Delta\alpha_k < \delta$. And thus, we have that following

$$\begin{aligned} U(h,P,\alpha) - L(h,P,\alpha) &= \sum_{k\in S_1}(M_k' - m_k')\Delta\alpha_k + \sum_{k\in S_2}(M_k' - m_k')\Delta\alpha_k \\ &< \epsilon\sum_{k\in S_1}\Delta\alpha_k + 2K\sum_{k\in S_2}\Delta\alpha_k < \epsilon[\alpha(b)-\alpha(a)] + 2K\delta \\ &< \epsilon[2K + \alpha(b) - \alpha(a)]. \end{aligned}$$

And thus this shows that $h \in R(\alpha)$ on $[a,b]$. □

The use of this lemma is that it increases our library of integrable functions substantially by only making use of functions that we know to be continuous. In fact we will make use of this lemma immediately to prove the following.

Theorem 1: Assume that $f, g \in R(\alpha)$ on $[a,b]$, then

- (*Integrability is closed under multiplication*) The function $fg \in R(\alpha)$.
- (*The triangle inequality for integrals*) The function $|f| \in R(\alpha)$ and

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

Proof. The function $\psi : \mathbb{R} \to [0,\infty)$ given by $\psi(t) = t^2$ is continuous, thus by our lemma if $f \in R(\alpha)$ then $f^2 = \psi \circ f \in R(\alpha)$. We also know that integrability is closed under addition and scalar multiplication. Thus, $f+g \in R(\alpha)$, and similarly for $f-g$. But then

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

which shows that $fg \in R(\alpha)$, and this proves part a).

For part b)., the function $\psi : \mathbb{R} \to [0,\infty)$ given by $\psi(t) = |t|$ is continuous, and thus if $f \in R(\alpha)$, then $|f| = \psi \circ f \in R(\alpha)$. Now, $\int_a^b f d\alpha$ is a number, thus it is either negative or nonnegative. Thus for $c = \pm 1$ we have the following

$$\left|\int_a^b f d\alpha\right| = c\int_a^b f d\alpha.$$

If we now exploit the homogeneity (can bring scalars in and out of an integral on a whim) of the integral, and the fact that as $c = \pm 1$, we have $cf \leq |f|$, then we have

$$\left|\int_a^b f d\alpha\right| = c\int_a^b f d\alpha = \int_a^b cf d\alpha \leq \int_a^b |f| d\alpha,$$

due to the monotonicity of the integral. □

Note: It may seem like I was being a little handwavey above there. Recall to use our lemma, for $m \leq f(x) \leq M$ on $x \in [a,b]$, we must have that $[m,M]$ is contained within the domain of ψ. In the proofs above that was no issue as the domains of the squaring and absolute value functions are all of $\mathbb{R}$. But be careful.

- If $f^3 \in R(\alpha)$, then $f \in R(\alpha)$ as $\psi(t) = t^{1/3}$ is a function $\psi : \mathbb{R} \to \mathbb{R}$. (In fact $f^n \in R(\alpha)$ for n odd implies that $f \in R(\alpha)$.
- However, $f^2 \in R(\alpha)$ does not automatically imply that $f \in R(\alpha)$. We saw a counterexample of this in homework 1, by taking

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \cap [0,1] \\ -1 & \text{if } x \in \mathbb{R} \backslash \mathbb{Q} \cap [0,1] \end{cases}$$

  Then $f^2 \equiv 1$ on $[0,1]$, thus $f^2 \in R(\alpha)$, but $f \notin R(\alpha)$. The general reason for this is that the inverse of squaring $\psi(t) = \sqrt{t}$ exists but as a function $\psi : [0,\infty) \to [0,\infty)$. And we see that the domain of ψ does not contain both the maximal and minimal values of f.

We have covered what it means for a function to be integrable and properties of the integral, but now the next two results we come to in todays lecture (and one on Wednesdays) are more related to computing integrals. To be more specific, we finally come to a point where our generalization to the Stieltjes theory at the beginning really pays off, as we will see how the Stieltjes theory reduces to computing Infinite series or standard Riemann integrals in almost all scenarios.

Definition: Define the following function, often called a step function or the Heavside step function.

$$I(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$$

Theorem 2: (*Integrating with respect to a step function*) For $s \in (a,b)$, and f a bounded function on $[a,b]$ with f continuous at $x = s$. If we define

$$\alpha(x) = I(x - s) = \begin{cases} 0 & x \leq s \\ 1 & x > s \end{cases}$$

Then $f \in R(\alpha)$ and

$$\int_a^b f d\alpha = f(s).$$

Proof. Take the following partition of $[a,b]$ given by $P = \{a, s, x_2, b\}$. Then we have only three subintervals $[a,s]$, $[s,x_2]$, and $[x_2,b]$. It is clear that

$$\begin{aligned} \Delta\alpha_1 &= \alpha(s) - \alpha(a) = 0 \\ \Delta\alpha_2 &= \alpha(x_2) - \alpha(s) = 1 \\ \Delta\alpha_3 &= \alpha(b) - \alpha(x_2) = 0 \end{aligned}$$

Thus $U(f,P,\alpha) - L(f,P,\alpha) = M_2 - m_2$. We can continue this argument as follows. Define the partition $P_3 = \{a, s, x_3, x_2, b\}$. Similar computation shows that the only subinterval where $\Delta\alpha$ does not vanish is the second subinterval. Thus

$$U(f,P_3,\alpha) - L(f,P_3,\alpha) = M_2 - m_2,$$

where M_2 and m_2 are the maximum and minimum of $f(x)$ on the interval $[s,x_3]$. Continuing on we can define for every $n \in \mathbb{N}$ a partition $P_n = \{a, s, x_n, x_{n-1}, \ldots, x_2, b\}$ where

$$U(f,P_n,\alpha) - L(f,P_n,\alpha) = M_2 - m_2,$$

where M_2 and m_2 are the maximum and minimum of $f(x)$ on the interval $[s, x_n]$. As the placement of the new points x_n in each successive partition can be done arbitrarily between x_{n-1} and s, there is no loss in generality in assuming that $x_n \to s$. As f is continuous at s we have that

$$f(s) = \lim_{n\to\infty} M_2([s, x_n]) = \lim_{n\to\infty} m_2([s, x_n]).$$

Thus we see that we can make the difference between the upper and lower sum arbitrarily small for n large enough so $f \in R(\alpha)$, and

$$\int_a^b f d\alpha = \inf U(f, P, \alpha) = \lim_{n\to\infty} M_2([s, x_n]) = f(s).$$

□

The next result is a slight generalization of the one we have just proven.

Theorem 3: Take $\{c_n\}_{n=1}^{\infty}$ to be a sequence of positive terms, i.e. $c_n \geq 0$ for all $n \in \mathbb{N}$. Further assume that $\sum_{n=1}^{\infty} c_n$ converges. If $\{s_n\}_{n=1}^{\infty}$ is a sequence of distinct points, $s_k \in (a, b)$ for all $k \in \mathbb{N}$, and we define

$$\alpha(x) = \sum_{n=1}^{\infty} c_n I(x - s_n)$$

then if f is continuous on $[a, b]$,

$$\int_a^b f d\alpha = \sum_{n=1}^{\infty} c_n f(s_n).$$

Proof. As $\sum_{n=1}^{\infty} c_n$ converges, for $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\sum_{k=N+1}^{\infty} c_k < \epsilon.$$

We break α into two functions α_1 and α_2.

$$\alpha_1(x) = \sum_{k=1}^{N} c_k I(x - s_k), \qquad \alpha_2(x) = \sum_{k=N+1}^{\infty} c_k I(x - s_k).$$

Both functions α_1 and α_2 are monotonically increasing, and thus we can integrate with respect to them.

$$\int_a^b f d\alpha_1 = \int_a^b f d\left(\sum_{k=1}^{N} c_k I(x - s_k)\right)$$

By the linearity of the integral with respect to the weight, we have that

$$\int_a^b f d\alpha_1 = \sum_{k=1}^{N} c_k \int_a^b f d(I(x - s_k)).$$

Invoking theorem 2 gives us that

$$\int_a^b f d\alpha_1 = \sum_{k=1}^{N} c_k f(s_k).$$

As $\alpha_2(a) = 0$, we have that

$$\alpha_2(b) - \alpha_2(a) = \sum_{k=N+1}^{\infty} c_k < \epsilon.$$

Because of this and part iv) of Theorem 6.4 we have

$$\left| \int_a^b f d\alpha_2 \right| < M\epsilon,$$

where $M = \sup_{x\in[a,b]} |f(x)|$. And now

$$\left| \int_a^b f d\alpha - \sum_{k=1}^{N} c_k f(s_k) \right| = \left| \int_a^b f d\alpha - \int_a^b f d\alpha_1 \right| = \left| \int_a^b f d\alpha_2 \right| < M\epsilon.$$

□

From this theorem we see that the theory of Riemann–Stieltjes integration generalizes the theory of infinite series.

If the sequence elements $\{s_n\}_{n=1}^{\infty}$ in the previous theorem are listed in order (i.e. $s_1 < s_2 < \cdots$), then α in the previous theorem can be thought of in the following manner

$$\alpha(x) = \begin{cases} 0 & a \le x \le s_1 \\ c_1 & s_1 < x \le s_2 \\ c_1 + c_2 & s_2 < x \le s_3 \\ \vdots & \vdots \\ \sum_{k=1}^{n} c_k & s_n < x \le s_{n+1} \\ \vdots & \vdots \end{cases}$$

Integration: Day 10

We begin today with an alternate version of the fundamental theorem of calculus.

Theorem 1: (*Fundamental Theorem of Calculus Part II*) If $f \in R$ on $[a, b]$ and if there is a differentiable function F on $[a, b]$ such that $F' = f$, then

$$\int_a^b f dx = F(b) - F(a).$$

Proof. Let $\epsilon > 0$. As $f \in R$, choose a partition $P = \{x_0, x_1, \ldots, x_n\}$ such that

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

On each subinterval $[x_{k-1}, x_k]$ of the partition P, by the mean value theorem, there exists $t_k \in (x_{k-1}, x_k)$ such that

$$\frac{F(x_k) - F(x_{k-1})}{x_k - x_{k-1}} = F'(t_k) = f(t_k).$$

Thus, $F(x_k) - F(x_{k-1}) = f(t_k)\Delta x_k$. And so, from telescoping series

$$\sum_{k=1}^{n} f(t_k)\Delta x_k = F(b) - F(a).$$

From Lecture 2.2 part c). We have that

$$\left|\sum_{k=1}^{n} f(t_k)\Delta x_k - \int_a^b f dx\right| < \epsilon.$$

Putting the previous two statements together with the fact that this can be done for any arbitrary $\epsilon > 0$ gives that

$$\int_a^b f dx = F(b) - F(a).$$

□

A quick immediate result that follows from applying this version of the fundamental theorem of calculus is the integration by parts formula.

Theorem 2: (*Integration by Parts*) Assume that F,G are differentiable, and that $F' = f \in R$ and $G' = g \in R$. Then

$$\int_a^b F(x)g(x)dx = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x)dx.$$

Proof. First define the function $H(x) = F(x)G(x)$. We know that H is differentiable as F and G are. The product rule tells us the explicit form of $H'(x)$,

$$H'(x) = f(x)G(x) + F(x)g(x).$$

As G and F are differentiable on $[a, b]$ and hence continuous, we have that $F, G \in R$ on $[a, b]$. As integrability is closed under sums and products, we have that $H' \in R$. Thus, let us invoke the fundamental theorem of calculus. So,

$$\begin{aligned}
\int_a^b H'(x)dx &= H(b) - H(a) \\
\int_a^b [f(x)G(x) + F(x)g(x)]dx &= F(b)G(b) - F(a)G(a) \\
\int_a^b F(x)g(x)dx &= F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x)dx.
\end{aligned}$$

□

The thing to note about this proof is it shows that integration by parts is precisely nothing more than the product rule 'backwards' in a sense, or an inverse product rule if you prefer to think of differentiation and integration as inverses of each other.

Okay, at this point, the dead horse that is Riemann–Stieltjes integration theory has been sufficiently beaten, and thus it is time, pardon the pun, to move onto greener pastures.

A prototype of what's to come

Def: For each $k \in \mathbb{N}$, let $\mathbb{R}^k$ be the set of all ordered k–tuples

$$\vec{x} = (x_1, x_2, \ldots, x_k).$$

It is well known from your linear algebra class that $\mathbb{R}^k$ is a vector space under vector addition and scalar multiplication,

$$\begin{aligned} \vec{x} + \vec{y} &= (x_1 + y_1, \ldots, x_k + y_k) \\ a\vec{x} &= (ax_1, \ldots, ax_k). \end{aligned}$$

The vector space $\mathbb{R}^k$ is also equipped with a scalar product $\cdot : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ given by

$$\vec{x} \cdot \vec{y} = \sum_{n=1}^{k} x_n y_n,$$

and a norm $\|\cdot\| : \mathbb{R}^k \to [0, \infty)$ given by

$$\|\vec{x}\| = \sum_{n=1}^{k} x_k^2 = \sqrt{\vec{x} \cdot \vec{x}}.$$

It is not difficult to see that the scalar product is symmetric, i.e.

$$\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$$

and bilinear, i.e. linear in each 'slot'

$$\begin{aligned} (a\vec{x} + b\vec{y}) \cdot \vec{z} &= a(\vec{x} \cdot \vec{z}) + b(\vec{y} \cdot \vec{z}) \\ \vec{x} \cdot (a\vec{y} + b\vec{z}) &= a(\vec{x} \cdot \vec{y}) + b(\vec{x} \cdot \vec{z}). \end{aligned}$$

We collect the following results about the norm of vectors in $\mathbb{R}^k$.

Theorem 3: Let $\vec{x}, \vec{y} \in \mathbb{R}^k$ and $\alpha \in \mathbb{R}$, then

a). $\|\vec{x}\| \geq 0$.

b). $\|\vec{x}\| = 0$ if and only if $\vec{x} = \vec{0}$.

c). $\|\alpha\vec{x}\| = |\alpha|\|\vec{x}\|$.

d). $|\vec{x} \cdot \vec{y}| \leq \|\vec{x}\|\|\vec{y}\|$.

e). $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$.

Part d). is often called the Cauchy–Schwarz inequality, and part e). is the standard triangle inequality.

Proof. Parts a), b), and c) are direct consquences of the definition of the scalar product and the fact that squares of real numbers are positive and $\sqrt{\alpha^2} = |\alpha|$. To prove *d*)., fix $\vec{x}, \vec{y} \in \mathbb{R}^k$, and define the function $q : \mathbb{R} \to [0, \infty)$ by

$$q(t) = \|\vec{x} + t\vec{y}\|^2.$$

As $q(t) = (\vec{x} + t\vec{y}) \cdot (\vec{x} + t\vec{y})$, the bilinearity and symmetry of the scalar product gives us that

$$q(t) = t^2\|\vec{y}\|^2 + 2t(\vec{x} \cdot \vec{y}) + \|\vec{x}\|^2.$$

As $q(t) \geq 0$ for all values of t, and as q is a quadratic function, it must be the case that q has at most one real zero. (The graph of $q(t)$ is an upward pointing parabola in which it's vertex may intersect the x–axis.) As such, the discriminant of this quadratic equation is less than or equal to 0. Thus,

$$4(\vec{x} \cdot \vec{y}) - 4\|\vec{x}\|^2\|\vec{y}\|^2 \leq 0,$$

which implies that $|\vec{x} \cdot \vec{y}| \leq \|\vec{x}\|\|\vec{y}\|$.

We now move onto proving e). If we take $t = 1$, then

$$\|\vec{x} + \vec{y}\|^2 = \|\vec{y}\|^2 + 2(\vec{x} \cdot \vec{y}) + \|\vec{x}\|^2.$$

We now immediately make use of the Cauchy–Schwarz inequality we just proved to get

$$\begin{aligned}\|\vec{x} + \vec{y}\|^2 &\leq \|\vec{y}\|^2 + 2\|\vec{x}\|\|\vec{y}\| + \|\vec{x}\|^2 \\ &= (\|\vec{y}\| + \|\vec{x}\|)^2.\end{aligned}$$

Thus, by taking squareroots of both sides, we get

$$\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|.$$

□

Note: The proofs of these two theorems only depended upon properties of a norm as it is related to an inner product that we will define at a later date. No, part of the proofs above relied upon the fact that we were in $\mathbb{R}^k$, and as such, these proofs will be identical when we come back to them at a later date.

It turns out that $\mathbb{R}^k$ is a good prototype for whats to come. In the study of analysis, we often require tools that give us a notion of distance between points in a space, lengths of vectors, angles between vectors, etc in spaces besides Euclidean space. But $\mathbb{R}^k$ is the blueprint and the toy model, the simple version we understand to grasp these concepts while using the intuition gained here to define new tools and topics in more general settings. In particular $\mathbb{R}^k$ is a great example of a metric space, normed space, and inner product space. These topics will all be defined shortly.

Of particular to mention though, is

$$\text{Inner Product spaces } \subset \text{ Normed spaces } \subset \text{ Metric spaces}$$

We will first begin by studying metric spaces and their properties before moving on and adding further structure to study normed spaces and inner product spaces.

Metric Spaces

Def: A set X is a metric space if there exists a function $d : X \times X \to \mathbb{R}$, such that for any two points $p, q \in X$, the function $d(p, q)$ has the following properties

a). $d(p, q) > 0$ if $p \neq q$; $d(p, p) = 0$.

b). $d(p,q) = d(q,p)$.

c). $d(p,q) \leq d(p,r) + d(r,q)$.

The function d is often called the distance or metric on X.

Note: It is common to call the pair (X,d) a metric space instead of calling just the space X metric. Some people prefer this notion as it makes explicit that a metric space is a set X paired with a distance function d.

In the definition above, we see that the conditions on d are sensible for a function that we would consider a distance function. Part a). states that the distance between distinct points should be positive. Part b). says measuring a distance between two points should be independent of which point comes first, i.e. the measurement on a ruler should be independent of if the ruler starts at p and ends at q or vice–versa. And part c)., the triangle inequality, states that notions of distance 'cut out the middle man'. In other words, in a metric space, the shortest distance between two points comes from traveling only between those two points; no adding paths to some third point shortens the distance.

Def: Given a set X, and a function $d : X \times X \to \mathbb{R}$ that satisfies b), c) of the above, and that $d(p,q) \geq 0$, but also has the property that $d(p,q) = 0$ does not necessarily imply that $p = q$, then d is called a pseudometric.

Examples:

a). On the set $\mathbb{R}^n$ define the following

$$d(\vec{x},\vec{y}) = \|\vec{x} - \vec{y}\|,$$

then d is a metric on $\mathbb{R}^n$.

b). Define $C([a,b])$ to be the set of all continuous functions on the set $[a,b]$. Then for $s \in (a,b)$, define

$$d(f,g) = |f(s) - g(s)|,$$

then d is a pseudometric on $C([a,b])$.

c). For $f, g \in C([a,b])$, defining

$$d(f,g) = \sup_{x\in[a,b]} |f(x) - g(x)|$$

is a metric on $C([a,b])$.

d). Take $f, g \in R$ on $[a,b]$ and $\alpha \geq 1$, define

$$d(f,g) = \int_a^b |f(x) - g(x)|^\alpha dx$$

then d is a pseudometric on R. If we defined d on $C([a,b])$, then d would be a metric.

e). The taxi–cab metric on $\mathbb{R}^n$. Define

$$d_1(\vec{x},\vec{y}) = \sum_{k=1}^{n} |x_k - y_k|,$$

then d_1 is a metric on $\mathbb{R}^n$.

f). For $\alpha \geq 1$, define the following on $\mathbb{R}^n$,

$$d_\alpha(\vec{x}, \vec{y}) = \left(\sum_{k=1}^{n} |x_k - y_k|^\alpha \right)^{\frac{1}{\alpha}},$$

then d_α is a metric on $\mathbb{R}^n$. Also, note that $\alpha = 2$ is the euclidean metric we defined in part a).

g). On $\mathbb{R}^n$ define,

$$d_\infty(\vec{x}, \vec{y}) = \max_{1 \leq x \leq n} |x_k - y_k|,$$

then d_∞ is a metric on $\mathbb{R}^n$.

Metric Spaces: Day 2

Today, we begin a journey towards a result of importance. Analysis would mean very little at its core if we could not talk about convergence of objects in a definitive way. Because of this, the notion of completeness of a metric space plays a large role in many questions in analysis. Today we will at least begin the proof of a result that tells us that even if we are currently working in a space that is not complete, then we can move to the completion of our space which, hence it's namesake, is complete.

But first, we must have a formal definition of Cauchy sequences and completeness.

Def: A sequence $\{x_n\}_{n=1}^{\infty}$ in a metric space X is Cauchy if and only if for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that $\forall m, n > N$, we have $d(x_n, x_m) < \epsilon$, where d is the metric of the space X.

Note: At this point you may be thinking a natural question. The definition of Cauchy depends on the metric on X, and as we saw last lecture, a space can have many metrics on it, so do different choices of metric lead to different Cauchy sequences? The answer is of course yes, and we will talk about this in the coming classes.

Def: A metric space (X, d) is complete if and only if every Cauchy sequence in X converges.

A stereotypical staple of proof based math courses is to first learn of an object and then learn about maps that preserves properties when moving between objects of similar types. Think of linear maps from one vector space to another preserving vector space structure, or a homomorphism translating group operations from one group into another.

In a similar manner, as we have just given an absract description of spaces that have notions of distance on them, we can now give a name to the maps that preserve distance.

Def: The metric spaces (X, d) and (Y, σ) are called isometric if and only if there exists an injective map $f : X \to Y$ onto Y such that for all $x, y \in X$,

$$\sigma(f(x), f(y)) = d(x, y).$$

Such maps f are called isometries of X and Y.

We can now state our result.

Theorem 1: Every metric space X with metric d can be isometrically embedded as a dense subset of a complete metric space, called $\widetilde{X}$. This completion is unique up to isometry that fixes X pointwise.

Note: The space $\widetilde{X}$ is often called the metric completion of X.

Proof. Let (X, d) be a metric space. Call

$$\beta = \{(x_n)_{n=1}^{\infty} \mid (x_n)_{n=1}^{\infty} \text{ Cauchy in } X\}$$

the collection of all Cauchy sequences in X. We will use (x_n) as shorthand to mean $(x_n)_{n=1}^{\infty}$. Now, if $(x_n), (y_n) \in \beta$ are two Cauchy sequences in X, then the triangle inequality of the metric d applied twice gives

$$\begin{aligned} d(x_n, y_n) &\leq d(x_n, x_m) + d(x_m, y_n) \\ &\leq d(x_n, x_m) + d(x_m, y_m) + d(y_m, y_n). \end{aligned}$$

Thus

$$d(x_n, y_n) - d(x_m, y_m) \leq d(x_n, x_m) + d(y_n, y_m).$$

Interchanging n with m and vice–versa gives another inequality, which, when put togther with the one above gives

$$|d(x_n, y_n) - d(x_m, y_m)| \leq d(x_n, x_m) + d(y_n, y_m).$$

And as (x_n), (y_n) is Cauchy, there exists $N \in \mathbb{N}$ such that for m, n large enough, i.e. $m, n > N$,

$$|d(x_n, y_n) - d(x_m, y_m)| \leq \epsilon$$

for any arbitrary ϵ. Thus $\{d(x_n, y_n)\}_{n=1}^{\infty}$ is a Cauchy sequence of real numbers. As $\mathbb{R}$ is complete, we have that $\{d(x_n, y_n)\}_{n=1}^{\infty}$ has a definable limit. As (x_n), $(y_n) \in \beta$ were taken arbitrarily, we can define $d^* : \beta \times \beta \to \mathbb{R}$ in the following manner,

$$d^*((x_n), (y_n)) = \lim_{n\to\infty} d(x_n, y_n).$$

Let's take a step back for a moment. What are we doing. Does β extend our space X? Well, yes, to put it bluntly. For any element $x \in X$, if we construct the constant sequence $(x, x, \dots) = \{x\}_{n=1}^{\infty}$, then clearly $(x) \in \beta$ as all constant sequences are Cauchy. Thus every element $x \in X$ has an identification with an element $(x) \in \beta$. So, does β specifically contain X? No. But it contains something that can be identified to X.

Okay, so β extends X in this sense, but is it a metric space? This was the point of creating d^*. The map d^* extended d in some sense. Now, we ask, did d^* inherit any properties from d? (Like being a metric for example.)

Well, for two Cauchy sequences $(x_n), (y_n) \in \beta$,

$$\begin{aligned} d^*((x_n), (y_n)) &= \lim_{n\to\infty} d(x_n, y_n) = \lim_{n\to\infty} d(y_n, x_n) \\ &= d^*((y_n), (x_n)). \end{aligned}$$

Thus d^* is symmetric. And for $(x_n), (y_n)$, and $(z_n) \in \beta$,

$$\begin{aligned} d^*((x_n), (y_n)) &= \lim_{n\to\infty} d(x_n, y_n) \leq \lim_{n\to\infty} [d(x_n, z_n) + d(z_n, y_n)] \\ &= d^*((x_n), (z_n)) + d^*((z_n), (y_n)), \end{aligned}$$

and so d^* also satisfies the triangle inequality. It is easy to check that $d^*((x_n), (y_n)) \geq 0$ for any (x_n) and $(y_n) \in \beta$.

But, $d^*((x_n), (y_n)) = 0$ does not imply that $(x_n) = (y_n)$. Thus d^* is a pseudometric.

Okay, step back number two. Is this a big deal? No. This comes up very often in mathematics, and for good reason. There are times when you have a tool or a map on a space, and you wish this tool or map to have some property, but it's too weak! In our case, $d^*((x_n),(y_n)) = 0$, does not carry enough gumption to force (x_n) to be equal to (y_n). Our space of Cauchy sequences β is too big, or our notion of sequential equality is too fine (i.e. seperates two sequences as different objects very easily, in this case $(x_n) \neq (y_n)$ if just one term differs.) So, what do we do? What we always do, define an equivalence relation. That concept precisely lets us enlargen our concept of equality to as big as we'd like and let's us shrink our space into equivalence classes simultaneously. (Making equality 'grab' more things at a time, and shrinking the space are really one in the same, think about it.)

Thus, define $\sim$ on β as follows: we call $(x_n) \sim (y_n)$ if $\lim_{n\to\infty} d(x_n, y_n) = 0$.

I leave it to you to check that this is in fact an equivalence relation on β. For notation in the remaining proof, $[(x_n)]$ will denote the equivalence class of the sequence (x_n), and $[\beta] = \beta\backslash\sim$ will denote the set of all equivalence classes of elements coming from β.

Now, let (x_n) and $(y_n) \in [(x_n)]$ be two representatives of the equivalence class of (x_n), and let (z_n) be another element of β. Now, once again the triangle inequality of d furnishes the following

$$d(x_n, z_n) \leq d(x_n, y_n) + d(y_n, z_n)$$
$$d(y_n, z_n) \leq d(y_n, x_n) + d(x_n, z_n).$$

Taking limits of both of these expressions as $n \to \infty$ gives that

$$\lim_{n\to\infty} d(x_n, z_n) = \lim_{n\to\infty} d(y_n, z_n).$$

In other words, the distance between (z_n) and any representative of $[(x_n)]$ is the same. This shows that d^* is well–defined when seen as a map $d^* : [\beta] \times [\beta] \to \mathbb{R}$, defined as

$$d^*([(x_n)],[(y_n)]) = d^*((x_n),(y_n)) = \lim_{n\to\infty} d(x_n, y_n).$$

And d^* is a pseudometric on $[\beta]$. And if $d^*([(x_n)],[(y_n)]) = \lim_{n\to\infty} d(x_n, y_n) = 0$, then $[(x_n)] = [(y_n)]$ as $(x_n) \sim (y_n)$. This shows that d^* is a metric on $[\beta]$. Also $[\beta]$ still extends X, i.e. each $x \in X$ is identified as $[(x)]$, the equivalence class of the constant sequence (x).

Thus, we have that $([\beta], d^*)$ is a metric space. We begin the next lecture with

Claim: $([\beta], d^*)$ is a complete metric space.

□

Metric Spaces: Day 3

Let us continue from where we were

Claim: $([\beta], d^*)$ is a complete metric space.

Proof. Let $\{X_k\}_{k=1}^{\infty}$ be a Cauchy sequence in $[\beta]$. For each X_k, let $\{x_k^j\}_{j=1}^{\infty}$ be a representative of X_k, i.e. $(x_k^j) \in X_k$ (as X_k is an equivalence class.)

Let $\epsilon > 0$. As $\{X_k\}_{k=1}^{\infty}$ is Cauchy, there exists $N_1 \in \mathbb{N}$ such that for all $m, n > N_1$,

$$d^*(X_m, X_n) = \lim_{j\to\infty} d(x_m^j, x_n^j) < \frac{\epsilon}{3}.$$

Now, each representative (x_k^j) of X_k is Cauchy, thus there exists $M_k \in \mathbb{N}$ such that for all $m, n > M_k$, we have

$$d(x_k^m, x_k^n) < \frac{1}{k}.$$

For some $j > M_k$ (for example $j = M_k + 1$) pick an x_k^j and define this to be y_k, i.e. $y_k = x_k^{M_k+1}$. And define the constant sequence $\{y_k\}_{n=1}^{\infty} = (y_k, y_k, \dots) = (y_k)$. As this is a constant sequence, it is automatically Cauchy, thus $(y_k) \in \beta$. Define $Y_k \in [\beta]$ to be the element of $[\beta]$ whose representative is (y_k). Because of the definition of y_k, we have

$$d(y_k, x_k^m) < \frac{1}{k} \text{ for } m > M_k.$$

Thus this is really showing that

$$d^*(X_k, Y_k) = \lim_{j\to\infty} d(x_k^j, y_k) < \frac{1}{k}.$$

The point here is this, it is not necessarily true that $X_k = Y_k$, i.e. we do not know that X_k and Y_k are the same equivalence class, but we do know that they are close in the sense of the metric d^*. In particular, if Y_k converged to some element $Z \in [\beta]$, then X_k would also converge to Z due to the triangle inequality,

$$d^*(X_k, Z) \le d^*(X_k, Y_k) + d^*(Y_k, Z).$$

So, we have reduced our proof in some sense. We only need to show that Y_k converges, and this will be our goal for the remainder of the argument.

Let us look at the following sequence, $\{y_n\}_{n=1}^{\infty} = (y_1, y_2, y_3, \dots)$ formed by the diagonal elements from the representatives of Y_k,

$$\begin{aligned} Y_1 &\sim (y_1, y_1, y_1, \dots) \\ Y_2 &\sim (y_2, y_2, y_2, \dots) \\ Y_3 &\sim (y_3, y_3, y_3, \dots) \\ \vdots &\sim \quad \vdots \end{aligned}$$

Let us check that $\{y_n\}_{n=1}^{\infty}$ is actually a Cauchy sequence in X. Let us use the triangle inequality twice

$$\begin{aligned} d(y_m, y_n) &\le d(y_m, x_m^j) + d(x_m^j, y_n) \\ &\le d(y_m, x_m^j) + d(x_m^j, x_n^j) + d(x_n^j, y_n). \end{aligned}$$

By the definition of y_m and y_n, for $j > \max\{M_m, M_n\}$ we have that

$$d(y_m, x_m^j) < \frac{1}{m}, \quad d(y_n, x_n^j) < \frac{1}{n}.$$

As $\{X_k\}_{k=1}^{\infty}$ was Cauchy, for $m, n > N_1$ we have that

$$d^*(X_m, X_n) = \lim_{j\to\infty} d(x_m^j, x_n^j) < \frac{\epsilon}{3}.$$

Thus for j large enough, we can assume that $d(x_m^j, x_n^j) < \frac{\epsilon}{3}$. Thus, for large enough values of j, and $m, n > N_1$ we have

$$d(y_m, y_n) \le \frac{1}{m} + \frac{\epsilon}{3} + \frac{1}{n}.$$

And now, by the archimedean property, there exists some N_2 such that $\frac{1}{N_2} < \frac{\epsilon}{3}$. Thus for $m, n > \max\{N_1, N_2\}$

$$d(y_m, y_n) \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

Thus $\{y_n\}_{n=1}^{\infty}$ is a Cauchy sequence, i.e. $(y_n) \in \beta$. Because of this, we may now define $Y = [(y_n)] \in [\beta]$.

The claim is now that Y_k converges to Y. We see

$$d^*(Y_k, Y) = \lim_{j \to \infty} d(y_k, y_j)$$

As we just saw $\{y_n\}_{n=1}^{\infty} \in \beta$, thus there exists $N \in \mathbb{N}$ such that $m, n > N$ implies that $d(y_m, y_n) < \epsilon$. Thus as $j \to \infty$, for $k > N$ we have $d(y_k, y_j) < \epsilon$. Thus

$$d^*(Y_k, Y) = \lim_{j \to \infty} d(y_k, y_j) < \epsilon, \quad \text{for } \ k > N.$$

Thus for $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for $k > N$, $d^*(Y_k, Y) < \epsilon$. And this is precisely the definition of convergence. Thus $Y_k \to Y$, which shows that X_k converges. Hence $([\beta], d^*)$ is a complete metric space.

In the previous lecture, we saw that X was embedded inside of $[\beta]$, but not exactly. To be more rigorous, there was a 'copy' or identification of X inside of $[\beta]$. In particular for each $x \in X$, we identified x as the equivalence class of the constant sequence $\{x\}_{n=1}^{\infty}$. Written in functional form, define

$$\varphi : X \to [\beta], \quad \varphi(x) = [(x)].$$

In this sense $\varphi(X)$ is the version of X inside of $[\beta]$ that we will work with.

Let us stop for one second, and look at a definition that we will see again in the future.

Def: A point p is a limit point of set E if every neighborhood of p contains points $q \neq p$ with $q \in E$.

An equivalent formulation of this definition is that p is a limit point of a set E if there exists a sequence $\{q_n\}_{n=1}^{\infty}$ of terms $q_n \in E$ for all $n \in \mathbb{N}$ and the sequence q_n converges to p.

Given $Z \in [\beta]$, then Z has a representative sequence $(z_k) \in \beta$. Define $Z_m = [\{z_m\}_{k=1}^{\infty}]$ to be the equivalence class of the constant sequence $(z_m, z_m, z_m, \dots)$. Based off our identification above, $Z_m = [(z_m)] = \varphi(z_m)$. And so, under our identification, $Z_m \in \varphi(X)$, so Z_m is 'effectively' and element of X. And we see

$$d^*(Z, Z_m) = \lim_{j \to \infty} d(z_j, z_m) < \epsilon$$

for m large enough, as $\{z_n\}_{n=1}^{\infty}$, the representative of Z, was a Cauchy sequence to begin with. Thus, Z is the limit of a sequence of terms Z_m that are contained in our identification of X, $\varphi(X)$. Thus, Z is a limit point of $\varphi(X)$. As there was nothing special about Z, we have shown that

$$\varphi(X)' = [\beta],$$

where Y' means the collection of all limit points of a set Y in a metric space. This is precisely what it means to say that $\varphi(X)$ is dense inside of $[\beta]$ with respect to the metric d^*. Thus, the identification of X inside of $[\beta]$ is a dense subset of $[\beta]$.

Lastly, let us see that this metric completion is unique up to isometry that fixes X pointwise.

To do so, let us assume that M is another metric completion of X, i.e. $X \subset M$ and M is complete with respect to the metric d on X and X is dense in M. Thus, for each $m \in M$, there is a sequence $\{x_n\}_{n=1}^{\infty}$ in X such that $x_n \to m$. This lets us define a map from M to $[\beta]$ as

$$f : M \to [\beta] \qquad f(m) = [(x_n)].$$

Let us check that f is well defined. If $\{y_n\}_{n=1}^{\infty}$ is another sequence of elements in X that converges to m, i.e. $y_n \to m$, then

$$\lim_{j\to\infty} d(x_j, y_j) = 0.$$

But then $[(x_n)] = [(y_n)]$, and so each input $m \in M$ has one output, namely the equivalence class $[(x_n)]$ of a sequence (x_n) with $x_n \to m$.

Given $m, n \in M$ with $m \neq n$, and assume that $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ are sequences with $x_n \to m$ and $y_n \to n$. Thus, there exists $N \in \mathbb{N}$ such that

$$d(x_n, x) < \frac{1}{4}d(x, y), \qquad d(y_n, y) < \frac{1}{4}d(x, y)$$

for $n > N$. Thus, by the triangle inequality (used twice)

$$\begin{aligned} d(x, y) &\leq d(x_n, x) + d(x_n, y_n) + d(y_n, y) \\ &< \frac{1}{2}d(x, y) + d(x_n, y_n) \end{aligned}$$

for $n > N$. And thus

$$d^*([(x_n)], [(y_n)]) = \lim_{n\to\infty} d(x_n, y_n) > \frac{1}{2}d(x, y) > 0.$$

Which implies that $f(m) \neq f(n)$. Thus, f is injective.

Now, for $[(x_n)] \in [\beta]$, as $(x_n) \in \beta$, (x_n) is Cauchy in M, thus $x_n \to m$ for some $m \in M$ since M is complete. Then $f(m) = [(x_n)]$, which shows that f is onto.

Lastly, for $x, y \in M$, there exists sequences $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ contained in X with $x_n \to x$ and $y_n \to y$. Then

$$d^*(f(x), f(y)) = d^*([(x_n)], [(y_n)]) = \lim_{n\to\infty} d(x_n, y_n) = d(x, y).$$

Thus f is an isometry. And for $x \in X$, the identification of x in $[\beta]$ is $[(x)]$, and

$$f(x) = [(x)],$$

which states that f fixes X pointwise. And the metric completion $[\beta]$ is unique up to an isometry that fixes X pointwise. □

Metric Spaces: Day 4

The advantage of the language of metric spaces is that it generalizes the structure of the real numbers (with the metric $d(x, y) = |x - y|$) in a manner such that almost all results that hold in $\mathbb{R}$ dependent on the metric will also hold in any metric space (X, d). We will see that this usually reduces to just replacing $|x-y|$ with $d(x, y)$ and the arguments remain similar. Thus, most of today is collecting definitions and results to formalize and make explicit the 'replacement' I mentioned above.

Def: Let X be a metric space, with a metric d.

a). For a point $p \in X$, we define the neighborhood (often shortened to nhood) about p of radius r, for $r \in (0, \infty)$. This is often notated by $N_r(p)$,

$$N_r(p) = \{q \in X \mid d(p, q) < r\}.$$

b). A point p is a limit point of a set E if every neighborhood of p contains points $q \neq p$ with $q \in E$. Put another way, for all $r \in (0, \infty)$,

$$(N_r(p) \backslash \{p\}) \cap E \neq \varnothing.$$

c). For $p \in E$, if p is not a limit point of E, then p is called an isolated point.

d). E is closed if every limit point of E is contained within E.

e). For a set E, a point $p \in E$ is called interior in E if there exists a neighborhood N of p with

$$p \in N \subset E.$$

f). E is open if all points in E are interior points.

g). A set E is bounded if there exists $M \in [0, \infty)$ and a $q \in X$ such that

$$d(p, q) < M, \quad \forall p \in E$$

h). A set E is dense in X if every point of X is a limit point of E, i.e.

$$E' = X.$$

Theorem 1: Every neighborhood is an open set.

Proof. Let's take a neighborhood about a point p, i.e. $N_r(p)$ for some $r > 0$. For $q \in N_r(p)$ we have that $d(q, p) < r$. Let us call $h = r - d(p, q)$, and look at the neighborhood $N_h(q)$ about q. Then for all $x \in N_h(q)$, then

$$d(p, x) \leq d(p, q) + d(q, x) < r - h + h = r.$$

Thus, all elements $x \in N_h(q)$ is contained in $N_r(p)$, and so, q is an interior point of $N_r(p)$. Thus the neighborhood $N_r(p)$ is open. □

Theorem 2: If p is a limit point of E, then every neighborhood of p contains an infinite number of points of E.

Proof. Suppose towards a contradiction that a neighborhood N of p contains only a finite number of points of E, call them $q_1, q_2, \ldots, q_n$. As this is only a finite number of points, we can define

$$r = \min_{1 \leq k \leq n} d(p, q_k).$$

And now, $N_r(p)$, is a neighborhood of p that contains no element of E, thus $(N_r(p) \backslash \{p\}) \cap E = \varnothing$, but this contradicts that p is a limit point of E. □

This theorem has a very direct corollary.

Corollary: A finite set has no limit points.

Def: Let X be a metric space, define the closure of a set E, notated as $\overline{E}$, by

$$\overline{E} = E \cup E'$$

Theorem 3: If X is a metric space and $E \subset X$, then

a). The set $\overline{E}$ is closed.

b). $E = \overline{E}$ if and only if E is closed.

c). $\overline{E} \subset F$ for every closed set F with $E \subset F$.

Proof. **a).** If $p \in X$ and $p \notin \overline{E}$, then $p \notin E$ and $p \notin E'$. As $p \notin E'$ then there exists $N_r(p)$ with $N_r(p) \cap E = \varnothing$. This implies that p is interior in $(\overline{E})^c$, which shows that $(\overline{E})^c$ is an open set. Thus $\overline{E}$ is closed.

b). $\Rightarrow$ If $E = \overline{E}$, then E is closed as $\overline{E}$ is closed.

$\Leftarrow$ By the definition of $\overline{E}$, $E \subseteq \overline{E}$. If E is closed, then $E' \subseteq E$, and thus $\overline{E} \subseteq E$. So $E = \overline{E}$.

c). If F is closed, then $F' \subseteq F$. If $E \subseteq F$, then $E' \subseteq F' \subseteq F$, and thus $\overline{E} \subseteq F$. □

Note: Parts a) and b) of the previous theorem show that taking the closure of a set is an idempotent operation, i.e. taking the closure of a set twice is redundant.

$$\overline{\overline{E}} = \overline{E}.$$

Part c). allows us to give an alternate definition of the closure of a set E as

$$\overline{E} = \bigcap_{\{F \text{ closed},\ E \subset F\}} F.$$

And this shows that the closure of a set E is the smallest closed set containing E.

To change gears slightly, if X is a metric space with the metric d and Y is a subset of X, then Y is also a metric space with a metric $d|_{Y \times Y}$. In other words, Y is a metric space by restricting the metric of X to Y.

But, as the next example will show, openess and closedness are dependent upon the ambient background metric space. In particular, for $Y \subset X$, it is possible for a set to be open in Y and not be open in X.

Ex: Consider the plane $\mathbb{R}^2$ with the Euclidean metric,

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

Let us consider the set $S = \{(x, 0) \mid x \in \mathbb{R}\}$. Effectively, S is a copy of the real line inside of $\mathbb{R}^2$, namely the x–axis. The euclidean metric restricted to S is

$$d((x, 0), (y, 0)) = \sqrt{(x - y)^2 + (0 - 0)^2} = |x - y|.$$

And thus we see that $(S, d|_S)$ is effectively equal to $(\mathbb{R}, |\cdot|)$, i.e. the real line with the usual metric on it. For what follows, $N_r(p)$ means a neighborhood of p of radius r in the normal euclidean metric, and $N_r^S((x,0))$ means a neighborhood of $(x,0)$ of radius r in the restricted metric, $d|_S$.

The set $(0,1)\times\{0\}$ is open in S, as for each $(x,0)\in(0,1)\times\{0\}$ by taking $r=\min\{|x|,|x-1|\}$, we have $N_r^S((x,0))\subset(0,1)\times\{0\}$.

But $(0,1)\times\{0\}$ is not open in $\mathbb{R}^2$. For a point $(x,0)\in(0,1)\times\{0\}$, any neighborhood $N_r((x,0))$ of $(x,0)$ will contain points in $\mathbb{R}^2$ with nonzero y–coordinates (as neighborhoods in $\mathbb{R}^2$ are disks). Thus, no neighborhood of $(x,0)$ is contained in $(0,1)\times\{0\}$, so $(0,1)\times\{0\}$ is not open in $\mathbb{R}^2$.

Note however that every neighborhood $N_r^S((x,0))$ has the form $N_r^S((x,0)) = N_r((x,0))\cap S$. This is no accident, as the next definition and theorem will show.

Def: A set E is open relative to Y if for each $p\in E$, there exists a $r>0$ such that if $d(p,q)<r$ and $q\in Y$, then $q\in E$.

Theorem 4: For $Y\subset X$ with X a metric space. A subset E of Y is open relative to Y if and only if $E = Y\cap G$ where G is open in X.

Proof. $\Rightarrow$ Assume that E is open relative to Y. For each $p\in E$, there exists a radius of a neighborhood $r_p>0$ such that $d(p,q)<r_p$ implies $q\in E$ for all $q\in Y$. Define $V_p = N_{r_p}^X(p)$ to be the neighborhood of p of radius r_p in X. And define,

$$G = \bigcup_{p\in E} V_p.$$

It is clear that G is open in X. By definition, $E\subseteq G\cap Y$. For each $p\in E$, we have $V_p\cap Y\subseteq E$ also by definition. Thus, by taking the union, $G\cap Y\subseteq E$. Thus, $E = G\cap Y$.

$\Leftarrow$ Assume $E = G\cap Y$ for G an open set in X. For each $p\in E$, as G is open, there exists $r>0$ such that $N_r^X(p)\subset G$. But then $N_r^X(p)\cap Y$ is a neighborhood of p relative to Y, hence E is open relative to Y. $\square$

10 Metric Spaces & $\mathbb{R}^n$

Maybe supplement on higher dimensional spaces, metric spaces, topology.

Actually cut out material for relative openness and closedness, make a later section or sections that points out the difference when moving to $\mathbb{R}^n$ or even Banach and Hilbert spaces.

Relative openness and closed ness here first?

Relative openness and closed ness

theorem on structure of relatively open sets

Relative openness and closed ness is not transitive, need a better property

theorem compact has this transitive property that closed and open do not have, closure works but interior does not.

more general things about convex here?

11 Size & Smallness

maybe bring up in size and smallness, with countability, measure, and category. (oxtoby)

Solutions to Exercises

Solutions for section 1.1:

1. Begin with $x - (\sqrt[3]{2} + \sqrt[3]{3}) = 0$. Thus $x - \sqrt[3]{2} = \sqrt[3]{3}$. By cubing both sides we have

$$\begin{aligned}(x - \sqrt[3]{2})^3 &= 3\\ x^3 - 3(\sqrt[3]{2})x^2 + 3(\sqrt[3]{2})^2 x - 2 &= 3\\ -3\sqrt[3]{2}x(x - \sqrt[3]{2}) &= 5 - x^3\end{aligned}$$

At this point we substitute in $\sqrt[3]{3}$ for $x - \sqrt[3]{2}$ from what we had initially, thus

$$\begin{aligned}-3\sqrt[3]{6}x &= 5 - x^3\\ (-3\sqrt[3]{6}x)^3 &= (5 - x^3)^3\\ -162x^3 &= 125 - 75x^3 + 15x^6 - x^9\\ x^9 - 15x^6 - 87x^3 - 125 &= 0\end{aligned}$$

Thus we see that $\sqrt[3]{2} + \sqrt[3]{3}$ is a solution to the polynomial equation $x^9 - 15x^6 - 87x^3 - 125 = 0$, and thus is algebraic.

Solutions for section 1.2:

1. We begin with injectivity. Assume that $f(m, n) = f(p, q)$, thus

$$\begin{aligned}2^{m-1}(2n - 1) &= 2^{p-1}(2q - 1)\\ 2^{m-p}(2n - 1) &= 2q - 1\end{aligned}$$

If $m - p > 0$, then the left hand side is even and the right side is odd and this is impossible. Similarly if $p - m > 0$ then the right hand side is even and the left side is odd and this is impossible. Thus it must be that $m = p$. And this then says that

$$2n - 1 = 2q - 1$$

which shows that $n = q$. Thus $(m, n) = (p, q)$ and so f is injective.

For surjectivity, let x be an odd natural number, then $x = 2q - 1$ for some $q \in \mathbb{N}$ and so

$$f(1, q) = 2^{1-1}(2q - 1) = 2q - 1 = x$$

and so we see that f is onto the odd numbers. If x is an even number then $x = 2^m q$ where q is an odd number, i.e. after you factor out all possible powers of 2 from x you are left with an odd number. As q is of the form $2n - 1$ for some natural number n we have

$$f(m + 1, n) = 2^{m+1-1}(2n - 1) = 2^m(2n - 1) = 2^m q = x$$

and this shows f is onto the even numbers and thus f is surjective.

2. Let $p_1, p_2, \ldots, p_k$ be the first k prime numbers, and define $f : \mathbb{N}^k \to \mathbb{N}$ by

$$f(x_1, x_2, \ldots, x_k) = p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

Then by the unique factorization of whole numbers into primes, if $(x_1, x_2, \ldots, x_k) \neq (y_1, y_2, \ldots, y_k)$ (i.e. $x_j \neq y_j$ for some $1 \leq j \leq k$) then we must have $f(x_1, x_2, \ldots, x_k) \neq f(y_1, y_2, \ldots, y_k)$, and thus f is injective.

3. a). From the fundamental theorem of algebra, a polynomial of degree n has at most n zeroes or solutions to $p(a) = 0$.

   b). A general nth degree polynomial with integer coefficients is given by

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

   with $a_0, a_1, \ldots, a_n$ the $n + 1$ coefficients from $\mathbb{Z}$. From part a). we know there are n roots to this polynomial. Label $\{r_1, r_2, \ldots, r_n\}$ as the n roots of this polynomial $p(x)$. We can now define the surjective function $f : \mathbb{Z}^{n+1} \times \{1, 2, \ldots, n\}$ by

$$f((a_0, a_1, \ldots, a_n), j) = r_j$$

   i.e. f maps $(a_0, a_1, \ldots, a_n)$ to the polynomial $p(x)$, and maps $\{1, 2, \ldots, n\}$ to the roots $\{r_1, r_2, \ldots, r_n\}$ of $p(x) = 0$. It is clear from the definition of f that it is a surjective function. As $\{1, 2, \cdots n\} \subseteq \mathbb{Z}$ and $\mathbb{Z}^m$ is countable for $m \in \mathbb{N}$, this explains why $\mathbb{Z}^{n+1} \times \{1, 2, \ldots, n\}$ is countable.

   c). From part b). for each n we can define $f_n : \mathbb{Z}^{n+1} \times \{1, 2, \ldots, n\} \to$ zeroes of polynomials of degree n with integer coefficients, and we know this map is surjective. Thus the collection of algebraic numbers will be countable if

$$\bigcup_{n=1}^{\infty} (\mathbb{Z}^{n+1} \times \{1, 2, \ldots, n\})$$

   is countable. And this follows from part b). as well since the countable union of countable sets is countable. .

4. If A is an uncountable set and B is a countable subset of A, if it were the case that $A \backslash B$ was also countable then

$$A = B \cup (A \backslash B)$$

shows that A is the union of two countable sets and therefore would be countable, which contradicts that A is uncountable. Thus, when removing a countable set from an uncountable set, one is always left with an uncountable remainder.

Solutions for section 1.3:

1. Using one of the defining properties of being a totally ordered field, we know that $a + c < b + c$ if $a < b$ for any c. Thus starting with $a > 0$, add b to both sides

$$b = 0 + b < a + b$$

and as $b > 0$, the transitivity of $<$ gives the result.

2. a). As $x \neq 0$, it has a multiplicative inverse x^{-1} since our space is a field. Thus

$$y = 1 \cdot y = (x^{-1}x) \cdot y = x^{-1}(xy) = x^{-1}(xz) = (x^{-1}x)z = 1 \cdot z = z$$

b). This follows from what we just proved in part a). with $z = 1$.

c). We have the following

$$y = 1 \cdot y = (x^{-1}x)y = x^{-1}(xy) = x^{-1} \cdot 1 = x^{-1}$$

d). This follows from part c). as $xx^{-1} = 1$, we have that x is the multiplicative inverse of x^{-1} thus $x = (x^{-1})^{-1}$.

Solutions for section 1.4:

1. Using the triangle inequality, we have

$$|x| = |x - y + y| \leq |x - y| + |y|$$

thus we have

$$|x| - |y| \leq |x - y|$$

Switching x and y gives that

$$|y| - |x| \leq |y - x| = |x - y|$$

Thus we have

$$||x| - |y|| = \max(|x| - |y|, |y| - |x|) \leq |x - y|$$

Solutions for section 2.1:

1. Proving the convergence of the following sequences

a). Let $\epsilon > 0$, then by the Archimedean principle, there is a natural number N such that $\frac{1}{2N} < \epsilon$. And then

$$|x_n - 3| = \left|3 + \frac{(-1)^n}{2n} - 3\right| = \left|\frac{(-1)^n}{2n}\right| = \frac{1}{2n} < \frac{1}{2N} < \epsilon$$

for $n > N$. As this can be done for any ϵ we have that $\{x_n\} \to 3$.

b). Let $\epsilon > 0$, by the Archimedean principle there is a $N = \lceil \frac{39}{25\epsilon} - \frac{7}{5} \rceil$. It then follows for $n > N$ that $\frac{39}{5(5n+7)} < \epsilon$, and

$$\left|x_n - \frac{2}{5}\right| = \left|\frac{2n-5}{5n+7} - \frac{2(n+\frac{7}{5})}{5(n+\frac{7}{5})}\right| = \frac{39}{5(5n+7)} < \epsilon$$

thus $\{x_n\} \to \frac{2}{5}$.

c). Let $\epsilon > 0$. By the Archimedean property, there is a natural number N such that $N = \lceil \frac{1}{\epsilon} - 1 \rceil$. For all $n > N$ we have $\frac{1}{n+1} < \epsilon$ and

$$|x_n - 2| = \left| \frac{2n^2 + 3n - 2n^2 - 2n}{n^2 + n} \right| = \left| \frac{n}{n^2 + n} \right| = \frac{1}{n+1} < \epsilon$$

as this can be done for any $\epsilon > 0$ we have that $\{x_n\} \to 2$.

d). Using the formula

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

shows that $x_n = \frac{1}{2} + \frac{1}{2n}$, so by an argument similar to part a). once can see that $\{x_n\} \to \frac{1}{2}$.

2. We prove the first result by induction. For the base case $n = 1$ as $(1+h)^1 = 1 + h$ we see the result holds. We now assume that $(1+h)^n \geq 1 + nh$ for n and we will try to deduce the result for $n + 1$.

$$\begin{aligned}(1+h)^{n+1} &= (1+h)(1+h)^n \geq (1+h)(1+nh) \\ &= 1 + h + nh + nh^2 = 1 + (n+1)h + nh^2 \geq 1 + (n+1)h\end{aligned}$$

As $nh^2 > 0$ the last inequality holds. Thus by the principle of mathematical induction the result holds for all $n \in \mathbb{N}$.

For $0 < r < 1$, we have $1 < \frac{1}{r}$ and thus we can take $h = \frac{1}{r} - 1 > 0$, and

$$r = \frac{1}{\frac{1}{r}} = \frac{1}{1 + \frac{1}{r} - 1} = \frac{1}{1+h}$$

Thus by the inequality we have above,

$$0 < r^n < \frac{1}{(1+h)^n} \leq \frac{1}{1+nh}$$

Taking $N = \lceil \frac{1}{h} \left(\frac{1}{\epsilon} - 1 \right) \rceil$ for $\epsilon > 0$, shows that for all $n > N$

$$0 < r^n < \frac{1}{1+nh} < \epsilon$$

And this shows why $\lim_{n \to \infty} r^n = 0$.

Solutions for section 2.2:

1. The sequence $\{x_n\} = \left\{ \frac{1}{n} \right\}$ and the constant sequence $\{y_n\} = \{0\}$ have the property that $x_n > y_n$ for all $n \in \mathbb{N}$, but $\{x_n\} \to 0$ and $\{y_n\} \to 0$ and $0 \not> 0$.

2. For the squeezing lemma, let $\epsilon > 0$, as $\{a_n\} \to 0$ there exists an N such that

$$a_n = |a_n - 0| < \epsilon$$

for $n > N$. But this implies that $-\epsilon < 0 \leq b_n \leq a_n < \epsilon$ for $n > N$, or equivalently that

$$|b_n - 0| < \epsilon, \text{ for } n > N.$$

And thus $\{b_n\} \to 0$.

For the generalization, we know that $a_n \leq b_n \leq c_n$ for all $n \in \mathbb{N}$. Thus by subtracting, we have that $0 \leq b_n - a_n \leq c_n - a_n$. Using the algebraic limit rules, we have

$$\lim_{n\to\infty}(c_n - a_n) = \lim_{n\to\infty} c_n - \lim_{n\to\infty} a_n = L - L = 0$$

and so by the first version of the squeezing lemma above we have that $\lim_{n\to\infty} b_n - a_n = 0$. This shows that $\{b_n\}$ converges as it is the sum of two convergent sequences, $\{a_n\}$ and $\{b_n - a_n\}$, and by the algebraic limit rules we have

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty}(a_n + (b_n - a_n)) = \lim_{n\to\infty} a_n + \lim_{n\to\infty}(b_n - a_n) = L + 0 = L$$

3. Starting with the following computation

$$\sqrt{n^2+6n} - n \cdot \frac{\sqrt{n^2+6n}+n}{\sqrt{n^2+6n}+n} = \frac{n^2+6n-n^2}{\sqrt{n^2+6n}+n} = \frac{6n}{\sqrt{n^2+6n}+n} \cdot \frac{\frac{1}{n}}{\frac{1}{n}} = \frac{6}{\sqrt{1+\frac{6}{n}}+1}$$

gives us that idea that as $n \to \infty$ that this expression converges to 3. So, let us do some prep work, we have

$$\begin{aligned}
\left|\sqrt{n^2+6n} - n - 3\right| &= \left|\frac{6}{\sqrt{1+\frac{6}{n}}+1} - 3\right| = \left|\frac{6}{\sqrt{1+\frac{6}{n}}+1} - \frac{3(\sqrt{1+\frac{6}{n}}+1)}{\sqrt{1+\frac{6}{n}}+1}\right| \\
&= \left|\frac{3-3\sqrt{1+\frac{6}{n}}}{\sqrt{1+\frac{6}{n}}+1}\right| = \frac{3(\sqrt{1+\frac{6}{n}}-1)}{\sqrt{1+\frac{6}{n}}+1} \cdot \frac{\sqrt{1+\frac{6}{n}}+1}{\sqrt{1+\frac{6}{n}}+1} \\
&= \frac{3(1+\frac{6}{n}-1)}{(\sqrt{1+\frac{6}{n}}+1)^2} = \frac{18}{n(\sqrt{1+\frac{6}{n}}+1)^2} \leq \frac{18}{n}
\end{aligned}$$

as $\sqrt{1+\frac{6}{n}}+1 \geq 1$ for any $n \in \mathbb{N}$. By the archimedean property, for $\epsilon > 0$, there exists N such that $\frac{18}{N} < \epsilon$, and this show that

$$\left|\sqrt{n^2+6n} - n - 3\right| < \epsilon, \text{ for } n > N$$

and so as this can be done for any $\epsilon > 0$ we have that

$$\lim_{n\to\infty} \sqrt{n^2+6n} - n = 3$$

4. a). As $\{x_n\}$ is bounded, there exists a $M > 0$ such that $|x_n| \leq M$ for all $n \in \mathbb{N}$. As $\{y_n\} \to 0$, for $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that for all $n > N$ we have that

$$|y_n| = |y_n - 0| < \frac{\epsilon}{M}$$

But then for $n > N$ we have that

$$|x_n y_n - 0| = |x_n y_n| \leq M|y_n| < M \cdot \frac{\epsilon}{M} = \epsilon$$

As this can be done for any choice of $\epsilon > 0$ we have that $\{x_n y_n\} \to 0$.

b). Take the sequence $\{x_n\} = \{n\}$ and $\{y_n\} = \left\{\frac{1}{n}\right\}$, then we have $\{y_n\} \to 0$, but $\{x_n y_n\} = \{1\}$ is the constant sequence 1.

5. a). Let us first look at the following

$$|y_n - x| = \left|\frac{1}{n}\left(\sum_{k=1}^{n} x_k\right) - x\right| = \left|\frac{1}{n}\left(\sum_{k=1}^{n}(x_k - x)\right)\right| \leq \frac{1}{n}\sum_{k=1}^{n}|x_k - x|$$

using the general triangle inequality. Let $\epsilon > 0$, then as $\{x_n\} \to x$, there exists N such that for $n > N$ we have $|x_n - x| < \frac{\epsilon}{2}$. So, for $n > N$.

$$\begin{aligned}|y_n - x| \leq \frac{1}{n}\sum_{k=1}^{n}|x_k - x| &= \frac{1}{n}\left[\sum_{k=1}^{N}|x_k - x| + \sum_{k=N+1}^{n}|x_k - x|\right] \\ &< \frac{1}{n}\left[\sum_{k=1}^{N}|x_k - x|\right] + \frac{1}{n}(n-N)\frac{\epsilon}{2}\end{aligned}$$

Take $M = \max(|x_1 - x|, |x_2 - x|, \ldots, |x_N - x|)$, then we have

$$|y_n - x| < \frac{MN}{n} + \left(1 - \frac{N}{n}\right)\frac{\epsilon}{2} < \frac{MN}{n} + \frac{\epsilon}{2}$$

as $n > N$. By the Archimedean property, there exists N_1 such that for $n > N_1$ we have $\frac{MN}{n} < \frac{\epsilon}{2}$, thus for $n > \max(N, N_1)$ we have

$$|y_n - x| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

and thus this shows that $\{y_n\} \to x$.

b). The terms of the sequence $\{y_n\}$ are as follows:

$$y_n = \begin{cases} 0, & \text{if } n = 2m \\ -\frac{1}{n}, & \text{if } n = 2m-1 \end{cases}$$

and this shows that $\{y_n\} \to 0$.

6. a). As $\{x_n\} \to L$, for $\epsilon > 0$ there exists an N such that for all $n > N$ we have $|x_n - L| < \epsilon$. By the reverse triangle inequality we have

$$||x_n| - |L|| \leq |x_n - L| < \epsilon$$

for $n > N$ and this can be done for any $\epsilon > 0$ this shows that $\{|x_n|\} \to |L|$.

b). The converse is not true, take $\{x_n\} = \{(-1)^n\}$. Then $\{|x_n|\} = \{1\}$ is the constant sequence of just 1, but $\{x_n\}$ is divergent.

7. a). Assume that $\{x_n\} \to x$, then by the algebraic limit rules we have that

$$\lim_{n\to\infty} x_n^2 = \left(\lim_{n\to\infty} x_n\right)^2 = x^2$$

Assume that $\{x_n^m\} \to x^m$, and then the algebraic limit rules gives that

$$\lim_{n\to\infty} x_n^{m+1} = \left(\lim_{n\to\infty} x_n^m\right)\left(\lim_{n\to\infty} x_n\right) = x^m x = x^{m+1}$$

and thus by the principal of mathematical induction we have that $\{x_n^p\} \to x^p$ for all $p \in \mathbb{N}$.

For a polynomial of degree 1, $p(x) = a_1x + a_0$, then by the algebraic limit rules

$$\lim_{n\to\infty} p(x_n) = \lim_{n\to\infty} (a_1x_n + a_0) = a_1\left(\lim_{n\to\infty} x_n\right) + a_0 = a_1x + a_0 = p(x)$$

Now assume that $p(x)$ is a polynomial of degree m and that $\lim_{n\to\infty} p(x_n) = p(x)$. If $q(x)$ is a polynomial of degree $m+1$, then $q(x) = (b_1x + b_0)p(x)$ where $p(x)$ is a polynomial of degree m, thus by the algebraic limit rules

$$\lim_{n\to\infty} q(x_n) = \lim_{n\to\infty} (b_1x_n+b_0)p(x_n) = \left[\lim_{n\to\infty} b_1x_n + b_0\right]\left[\lim_{n\to\infty} p(x_n)\right] = (b_1x+b_0)p(x) = q(x)$$

and thus by the principal of mathematical induction, we have that the result holds for a polynomial of degree m for any $m \in \mathbb{N}$.

b). As a rational function is of the form $R(x) = \frac{p(x)}{q(x)}$ for polynomials $p(x)$, $q(x)$, by our result above, we have that $\{R(x_n)\} \to R(x)$ for sequences $\{x_n\}$ as long as the value, x, that $\{x_n\}$ converges to is not a zero of the denominator $q(x)$.

8. The first part of this argument with showing how there exists an N such that for all $n > N$ that $b_n \neq 0$ and further that $|b_n| > \frac{|b|}{2}$ is the same as it is in the notes.

Then for $n > N$ we have

$$\begin{aligned}\left|\frac{a_n}{b_n} - \frac{a}{b}\right| = \left|\frac{a_nb - ab_n}{b_nb}\right| &= \frac{|a_nb - ab + ab - ab_n|}{|b_n||b|} \\ &\leq \frac{1}{|b_n||b|}\left[|b||a_n - a| + |a||b_n - b|\right] \\ &< \frac{2}{|b|}|a_n - a| + \frac{2|a|}{|b|^2}|b_n - b|\end{aligned}$$

where the last inequality comes from $|b_n| > \frac{|b|}{2}$ for $n > N$. Let $\epsilon > 0$. As $\{a_n\} \to a$, there is N_1 such that $|a_n - a| < \frac{|b|\epsilon}{4}$ for $n > N_1$. Similarly there is a N_2 such that $|b_n - b| < \frac{|b|^2\epsilon}{4|a|}$ for $n > N_2$. Then for $n > \max(N, N_1, N_2)$ we have

$$\left|\frac{a_n}{b_n} - \frac{a}{b}\right| < \epsilon$$

and as this can be done for any choice of $\epsilon > 0$ we have $\left\{\frac{a_n}{b_n}\right\} \to \frac{a}{b}$.

Solutions for section 2.3:

1. The subsequence $\{u_{3k+1}\} = \{(-1)^{3k+1}\} = \{[(-1)^3]^k \cdot (-1)\} = \{(-1)^{k+1}\}$. Thus we have

$$u_4 = 1, \quad u_7 = -1, \quad u_{10} = 1, \quad u_{13} = -1, \quad u_{16} = 1$$

2. (a) $\{x_{4n}\} = \{\sin(n\pi) + \frac{1}{4n}\} = \{\frac{1}{4n}\}$, and this sequence converges to 0.

   (b) $\{x_{4n+1}\} = \{\sin(n\pi + \frac{\pi}{4}) + \frac{1}{4n+1}\} = \{(-1)^n\frac{\sqrt{2}}{2} + \frac{1}{4n+1}\}$ and this sequence diverges.

(c) $\{x_{4n+2}\} = \{\sin(n\pi + \frac{\pi}{2}) + \frac{1}{4n+2}\} = \{(-1)^n + \frac{1}{4n+2}\}$ and this sequence diverges.

(d) $\{x_{4n+3}\} = \{\sin(n\pi + \frac{3\pi}{4}) + \frac{1}{4n+3}\} = \{(-1)^{n+1}\frac{\sqrt{2}}{2} + \frac{1}{4n+3}\}$ and this sequence diverges.

3. (a) Let $\epsilon > 0$, as $\{u_{2k}\} \to L$, there exists $N_1 \in \mathbb{N}$ such that for all $k > N_1$ we have that

$$|u_{2k} - L| < \epsilon$$

As $\{u_{2k+1}\} \to L$ there exists $N_2 \in \mathbb{N}$ such that for $k > N_2$ we have that

$$|u_{2k+1} - L| < \epsilon$$

Taking $N = \max(N_1, N_2)$, as every natural number n is either even or odd, we have that for $k > N$ with $n = 2k$ or $n = 2k + 1$ that

$$|u_n - L| < \epsilon$$

and thus $\{u_n\} \to L$.

(b) No it does not, the subsequence $\{u_{3k+2}\}$ could converge to some value distinct from L and then $\{u_n\}$ would be divergent overall.

4. The condition required is that $\bigcup_{k=1}^{M} A_k$ equals either $\mathbb{N}$ or equals an infinite tail of the naturals,

$$\{n, n+1, n+2, \dots\}, \quad n \in \mathbb{N}$$

Solutions for section 2.4:

1. (a) For this we will make use of the geometric series formula that

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

It then follows that

$$\sum_{k=q+1}^{p} r^k = r^{q+1} \sum_{k=0}^{p-q-1} r^k = q^{k+1}\left[\frac{1 - r^{p-q}}{1 - r}\right] \leq \frac{r^{q+1}}{1 - r}$$

(b) This follows from, for $p > q$

$$|v_p - v_q| = \left|\sum_{k=q+1}^{p} r^k\right| \leq \frac{|r|^{q+1}}{|1 - r|}$$

As we know that $0 < r < 1$ and $\lim_{n\to\infty} r^n = 0$ in this case, this is enough to show that $\{v_n\}$ is Cauchy.

(c) It follows that for $p > q$

$$u_p - u_q = \sum_{k=q+1}^{p} \frac{1}{k!}$$

And for $k > 2$,

$$k! = k(k-1)(k-2)\cdots 3\cdot 2\cdot 1 > 2\cdot 2\cdots 2\cdot 1 = 2^{k-1}$$
$$\frac{1}{k!} < \frac{1}{2^{k-1}}$$

Thus

$$u_p - u_q = \sum_{k=q+1}^{p} \frac{1}{k!} \leq \sum_{k=q+1}^{p} \frac{1}{2^{k-1}} = 2\sum_{k=q+1}^{p} \frac{1}{2^k} = 2(v_p - v_q)$$

and so

$$|u_p - u_q| \leq 2|v_p - v_q|$$

(d) This follows from part b). If $\{v_n\}$ is Cauchy and we have just shown the bound above, then $\{u_n\}$ is Cauchy.

2. (a) As we know that $a_n \geq 1$ we have that $a_n^2 \geq 1$ and $a_n a_m \geq 1$, thus

$$a_p^3 - a_q^3 = (a_p - a_q)(a_p^2 + a_p a_q + a_q^2) \geq (a_p - a_q)(1+1+1) = 3(a_p - a_q)$$

and the result follows from here.

(b) The bound in part a) shows that if $\{a_n^3$ is Cauchy then $\{a_n\}$ must be Cauchy as well.

3. Let $\epsilon > 0$, as $\{x_n\}$ is Cauchy there is a $N_1 \in \mathbb{N}$ such that for all $p, q > N_1$ we have

$$|x_p - x_q| < \frac{\epsilon}{2}$$

As a subsequence $\{x_{n_k}\} \to L$ we have a $N_2 \in \mathbb{N}$ such that for all $n_k > N_2$ that

$$|x_{n_k} - L| < \frac{\epsilon}{2}$$

Then for $n > \max(N_1, N_2)$ we have that

$$|x_n - L| = |x_n - x_{n_n} + x_{n_n} - L| \leq |x_n - x_{n_n}| + |x_{n_n} - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Thus $\{x_n\} \to L$.

Solutions for section 3.1:

1. We prove the properties one by one,

- Given $\{x_n\} \in C_{\mathbb{Q}}$, as we have that

$$\lim_{n\to\infty} |x_n - x_n| = 0$$

we have that $\{x_n\} \sim \{x_n\}$ and thus $\sim$ is reflexive.

- Given $\{x_n\}, \{y_n\} \in C_{\mathbb{Q}}$ and assume that $\{x_n\} \sim \{y_n\}$, thus we have that $\lim_{n\to\infty} |x_n - y_n| = 0$. As $|x_n - y_n| = |y_n - x_n|$ we have that

$$\lim_{n\to\infty} |y_n - x_n| = 0$$

and thus $\{y_n\} \sim \{x_n\}$ and thus $\sim$ is symmetric.

- Given $\{x_n\}, \{y_n\}, \{z_n\} \in C_{\mathbb{Q}}$ and assume that $\{x_n\} \sim \{y_n\}$ and $\{y_n\} \sim \{z_n\}$. Thus we have that
$$\lim_{n\to\infty} |x_n - y_n| = 0, \qquad \lim_{n\to\infty} |y_n - z_n| = 0.$$
Thus for $\epsilon > 0$, there exists a $N_1 \in \mathbb{N}$ such that for all $n > N_1$, we have that $|x_n - y_n| < \frac{\epsilon}{2}$. Similarly, there exists a $N_2 \in \mathbb{N}$ such that for all $n > N_2$ we have that $|y_n - z_n| < \frac{\epsilon}{2}$. Thus for $N = \max(N_1, N_2)$ we have that for all $n > N$ that
$$|x_n - z_n| = |x_n - y_n + y_n - z_n| \leq |x_n - y_n| + |y_n - z_n| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$
where the triangle inequality was used. As this can be done for any $\epsilon > 0$, we have that $\lim_{n\to\infty} |x_n - z_n| = 0$. Thus $\{x_n\} \sim \{z_n\}$ and thus $\sim$ is transitive.

2. Let $\{x_n\}$ be the original Cauchy sequence, and let $\{x'_n\}$ be the sequence with a finite number of terms changed. Define
$$M = \{k \mid x'_k - x_k \neq 0\}$$
Thus M is precisely the indices of the terms in the sequence $\{x_n\}$ that were replaced by some other value. By assumption, M is a finite set, thus take $N_1 = \max(M)$. As $\{x_n\}$ is Cauchy, we know for $\epsilon > 0$ that there exists a $N_2 \in \mathbb{N}$ such that for all $p, q > N_2$ that $|x_p - x_q| < \epsilon$. Take $N = \max(N_1, N_2)$.

   Then for $p, q > N$, being larger than N_1 means
$$|x'_p - x'_q| = |x_p - x_q|$$
and p, q larger than N_2 means
$$|x'_p - x'_q| = |x_p - x_q| < \epsilon.$$
and as this can be done for any $\epsilon > 0$ we have that $\{x'_n\}$ is Cauchy.

3. Let $\{a_n\} \in \{0,1\}^{\mathbb{N}}$, i.e. $\{a_n\}$ is an element in the space of 0-1 sequences. For $\{q_n\}$ a Cauchy sequence of rational numbers, define a map from the space of 0-1 sequences to a new sequence.
$$f : \{0,1\}^{\mathbb{N}} \to C_{\mathbb{Q}}, \qquad f(\{a_n\}) = \{x_n\} = \left\{q_n + \frac{a_n}{n}\right\}$$
It is easy to see that this is an injective map from 0-1 sequences to a new sequence based at $\{q_n\}$. The fact that this new sequence is also Cauchy will follow as a result of us showing that the sequence $\{x_n\}$ is equivalent to $\{q_n\}$. This follows from
$$|x_n - q_n| = \left|q_n + \frac{a_n}{n} - q_n\right| = \left|\frac{a_n}{n}\right| \leq \frac{1}{n}$$
as $a_n \in \{0,1\}$ for each $n \in \mathbb{N}$. As $\frac{1}{n} \to 0$, we have that $\{x_n\} \sim \{q_n\}$. As the map from 0-1 sequences to the $\{x_n\}$ is injective, there are at least as many x_n as 0-1 sequences, and as $\{0,1\}^{\mathbb{N}}$ is uncountable the claim follows.

Solutions for section 3.5:

1. We will give a proof for the case $x > 1$, the case $0 < x < 1$ is very similar.

   As $x > 1$ we have that $1^2 = 1 < x$ and $x < x^2$, so let us initialize two sequences, $y_1 = 1$ and $z_1 = x$. Our goal will be to create two sequences $\{y_n\}$ and $\{z_n\}$ in which $\{y_n\}$ is a sequence of lower approximations and $\{z_n\}$ is a sequence of upper approximations.

   Thus define $m_2 = \frac{y_1+z_1}{2}$. There are three possibilities by the trichotomy law: if $m_2^2 = x$ then we are done, if $m_2^2 < x$ then define $y_2 = m_2$ and $z_2 = z_1$, if $m_2^2 > x$, then define $y_2 = y_1$ and $z_2 = m_2$.

   Continue on in this fashion and for each $n \in \mathbb{N}$ define $m_n = \frac{y_{n-1}+z_{n-1}}{2}$. There are three possibilities by the trichotomy law: if $m_n^2 = x$ then we are done, if $m_n^2 < x$ then define $y_n = m_n$ and $z_n = z_{n-1}$, if $m_n^2 > x$, then define $y_n = y_{n-1}$ and $z_n = m_n$.

   Either this process terminates at a finite step and we have found the square root of x is equal to m_n for some n, or we have two sequences $\{y_n\}$ and $\{z_n\}$.

   Generally, we have

   $$z_n - y_n = \begin{cases} z_{n-1} - m_{n-1} \\ m_{n-1} - y_{n-1} \end{cases} = \frac{z_{n-1} - y_{n-1}}{2}$$

   and from this we see that

   $$z_n - y_n = \frac{x-1}{2^{n-1}}$$

   which shows that $\{z_n - y_n\}$ converges to 0.

   The argument for why $\{y_n\}$ is a Cauchy sequence is effectively identical to the proof of the LUB property in this section. Thus $\{y_n\}$ and $\{z_n\}$ are Cauchy.

   Thus by the completeness of the reals, we have the existence of $y, z \in \mathbb{R}$ such that $\{y_n\} \to y$ and $\{z_n\} \to z$. And as we know $\{z_n - y_n\} \to 0$ we have that $z = y$.

   Each term in $\{y_n\}$ was chosen with the property that $y_n^2 < x$, and similarly each term of $\{z_n\}$ was chosen with the property that $z_n^2 > x$. Thus by the algebraic limit rules and how we know limits interact with orderings, we have that $y^2 \leq x$ and $z^2 \geq x$. As $z = y$ we have $y^2 = x$ and thus y is the square root of x.

2. The proof is identical is the power of 2 is replaced by a power n, thus this process can be used to generate nth roots of positive numbers.

Appendix: More on Cardinality

Cardinal Equivalences & Comparisions

The notion of cardinality equivalence forms an equivalence relation (equivalence concept) on sets

Theorem 103. *Let X be a set. For A and B subsets of X, define a relation $\approx$ by saying $A \approx B$ if and only if $|A| = |B|$. Then $\approx$ is an equivalence relation on $P(X)$.*

Proof. We show that $\approx$ satisfies the required properties of an equivalence relation.

- Let A be a subset of X, i.e. $A \in P(X)$. Then the identity map on A, $\mathrm{Id}_A : A \to A$ given by $\mathrm{Id}_A(a) = a$ for all $a \in A$ is clearly a bijection from A to itself. Thus, $|A| = |A|$. As $A \approx A$, and A is an arbitrary element of $P(X)$, we have that $\approx$ is reflexive.
- Let A and B be subsets of X and assume that $A \approx B$. Thus, $|A| = |B|$ and there exists a bijection $f : A \to B$. As f is a bijective function, we have that the inverse of f exists. And $f^{-1} : B \to A$ is a bijection. Thus, $|B| = |A|$, so $B \approx A$. So, $\approx$ is a symmetric relation.
- Let A, B, and C be subsets of X and assume that $A \approx B$ and $B \approx C$. Thus as $|A| = |B|$ and $|B| = |C|$, there exists bijections $f : A \to B$ and $g : B \to C$. We have that $g \circ f : A \to C$ and by a result in a previous lecture, $g \circ f$ is a bijection as f and g are. Thus, $|A| = |C|$, i.e. $A \approx C$. Therefore $\approx$ is transitive.

Thus our result is shown. □

This result may seem obvious and unecessary, but it is just the opposite. When the sets A and B are finite, the statement $|A| = |B|$ is nothing more than $m = m$ for some $m \in \mathbb{N}$, which is obvious. But, momentarily, we will begin talking about infinite sets in detail, and it is important that we can say the cardinality of two infinite sets are equal with definitive meaning even when the cardinalities of these sets are not equal to any natural number. [73]

With that being said, cardinality equality is not the only relation we can put on $P(X)$ coming from the concept of cardinality.

Definition 61. *Let X be a set. For two subsets A and B of X we define a relation $\preceq$ on $P(X)$ by $A \preceq B$ if and only if $|A| \leq |B|$ if and only if there exists an injective function $f : A \to B$.* [74].

Theorem 104. *Let X be a set. The relation $\preceq$ is reflexive and transitive on $P(X)$.*

Proof. To prove the result, simply check the two properties.

- Take $A \in P(X)$. As mentioned before, the identity map on A, $\mathrm{Id}_A : A \to A$, given by $\mathrm{Id}_A(a) = a$ for all $a \in A$ is a bijection from A to itself. As every bijection is also an injection, we have that there exists an injective map from A to A. Thus, $A \preceq A$ and $\preceq$ is reflexive.

[73] In our theorem, we restricted ourselves to subsets of $P(X)$ for some set X, so $\approx$ would be an equivalence relation. This is because the collection of all sets is too large to be a set and is often called a *class*. Because of this, in general, $\approx$, is called an *equivalence concept* and not an equivalence relation. Of interest to note, the collection of all sets of a fixed cardinality is also too large to be a set.

[74] Once again, this 'relation' does exist on the class of all sets, but it is not formally proper to call it a relation, hence the momentary restriction to $P(X)$

- Now, let A, B, and C be subsets of X, and assume that $A \preceq B$ and $B \preceq C$. Thus, there exists an injective map $f : A \to B$ and there exists and injective map $g : B \to C$. In a prior lecture we proved that if f and g are both injective, then $g \circ f$ is injective. Thus, there exists an injective map $g \circ f : A \to C$. Thus, $A \preceq C$ and $\preceq$ is transitive.

□

When working with abstract sets and functions between them it is sometimes easier to show a function has one property over another. In particular, given two sets A and B it may be easier to show there is a surjection $g : B \to A$ than an injection $f : A \to B$ and vice–versa. Because of this, as our definition above relies on the existence of an injective map, it would be nice to have a similar condition in the definition involving surjectivity to facilitate comparing A and B under $\preceq$. This leads to the following theorem.

Theorem 105. *Let X be a set. For A and B two subsets of X, $A \preceq B$ if and only if there exists a surjective map $g : B \to A$.*

Proof. $\Rightarrow$ Assume that $A \preceq B$. In other words, assume that there exists an injective map $f : A \to B$. As f is injective, it is bijective when the codomain is restricted to $\text{Ran}(f)$. Thus, $f : A \to \text{Ran}(f)$ has an inverse $f^{-1} : \text{Ran}(f) \to A$. Taking α to be an element of A, we define $g : B \to A$ by

$$g(x) = \begin{cases} f^{-1}(x) & \text{if } x \in f(A) = \text{Ran}(f) \\ \alpha & \text{if } x \in B \backslash f(A) \end{cases}$$

As $f^{-1} : \text{Ran}(f) \to A$ is a bijection, we have that g is surjective. Thus, there exists a surjective map $g : B \to A$.

$\Leftarrow$ Assume there exists a surjective map $g : B \to A$. (*Warning:* This proof will require the Axiom of Choice) As g is surjective, we have that $g^{-1}(\{a\}) \neq \varnothing$ for every $a \in A$. Thus, by the axiom of choice, we may assume that for each $a \in A$ one unique element $b_a \in B$ was chosen from each $g^{-1}(\{a\})$. Now, define $f : A \to B$ by $f(a) = b_a$. For $a, c \in A$, if we assume that $f(a) = f(c)$, then $b_a = b_c$. Thus, it must be that $a = c$, as only one unique b_a was chosen from each $g^{-1}(\{a\})$, otherwise $g(b_a) = a$ and $g(b_a) = c$ so $a \neq c$ would imply g is not a function. Thus, $a = c$. Therefore f is injective. Thus, there exists an injective map $f : A \to B$. Thus, $A \preceq B$. □

Remark 9. *We have the following:*

- *Because of the theorem above, to show $A \preceq B$ we must show there exists an injective map $f : A \to B$ or there exists a surjective map $g : B \to A$.*
- *Also, note that the definition of $A \preceq B$ and the theorem above only require the existence of certain functions. For $A \preceq B$ it is <u>not</u> required that <u>every</u> map from A to B is injective or that <u>every</u> map from B to A is surjective.*

Now, at this point you may be wondering for a set X if $\preceq$ is a partial order on $P(X)$. [75]. We have already shown that $\preceq$ is reflexive and transitive, but we have not shown anti–symmetry. The relation $\preceq$ is anti–symmetric in a particular way, [76] but this is not an easy result to show. This result is given a name.

[75] Once again, to avoid classes

[76] i.e. $A \preceq B$ and $B \preceq A$ will not imply $A = B$

Theorem 106. *(*Cantor–Schröder–Bernstein Theorem*) Let X be a set. For A and B subsets of X, if $A \preceq B$ and $B \preceq A$, then $A \approx B$.*

The proof of this result will be shown at a later part of the appendix.

Definition 62. *For a set X, and for subsets A and B of X we define $A \prec B$ (equivalently $|A| < |B|$) if $A \preceq B$ and $A \not\approx B$, (equivalently, $|A| \leq |B|$ and $|A| \neq |B|$). In terms of our language involving functions, $A \prec B$ if there exists a function $f : A \to B$ that is injective and there exists no function $g : B \to A$ that is surjective.*

We have successfully created (or recreated) a notion of counting elements in sets. It reduces to standard counting when sets are finite, but the language we have created using functions has the distinct advantage that we can now compare infinite sets with relative ease.

Infinite Sets

Definition 63. *A set A is called* **finite** *if $A = \varnothing$ or A can be put into bijection with a set of the form $\{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$.*

Now we can finally define an infinite set. (albeit in maybe a slightly disappointing way.)

Definition 64. *A set is* **infinite** *if it is not finite.*

At this point, let us stop and finally give a formal proof of a result we have assumed before. [77]

Theorem 107. *The set $\mathbb{N}$ is an infinite set.*

Proof. By way of contradiction, assume that $\mathbb{N}$ is a finite set. Thus, assume that there exists an $m \in \mathbb{N}$ and a bijection $f : \mathbb{N} \to \{1, 2, \ldots, m\}$. As f is onto, we have that $f^{-1}(\{k\}) \neq \varnothing$ for all $k \in \{1, 2, \ldots, m\}$. In a prior lecture, we saw that f being injective implies that $f^{-1}(\{k\})$ contains at most one element for each $k \in \{1, 2, \ldots, m\}$. Thus, as f is bijective, we have that $f^{-1}(\{k\})$ contains exactly one element for each $k \in \{1, 2, \ldots, m\}$. Because of this, let n_k be the element of $\mathbb{N}$ in $f^{-1}(\{k\})$ for each $k \in \{1, 2, \ldots, m\}$. Thus, $f^{-1}(\{k\}) = \{n_k\}$ for each $k \in \{1, 2, \ldots, m\}$.

Define $n = n_1 + n_2 + \cdots + n_m = \sum_{k=1}^{m} n_k$. As n is a sum of numbers in the naturals, it is clear that $n \in \mathbb{N}$. It is also clear that $n > n_k$ for all $k \in \{1, 2, \ldots, m\}$. As $f : \mathbb{N} \to \{1, 2, \ldots, m\}$ is surjective, we have that $f(n) \in \{1, 2, \ldots, m\}$. Therefore, for some $l \in \{1, 2, \ldots, m\}$ we have that $n \in f^{-1}(\{l\}) = \{n_l\}$. Thus, $f(n) = f(n_l) = l$, and $n \neq n_l$. This contradicts the injectivity of f. Thus we have reached a contradiction, and therefore $\mathbb{N}$ can not be put into bijection with any set of the form $\{1, 2, \ldots, m\}$ for $m \in \mathbb{N}$. □

The definition of an infinite set is a little restrictive, or may not be simple to apply in a given situation. The following result gives us a collection of results or properties that is equivalent to saying a set is infinite.

[77] If we had constructed the naturals formally this result would not be required as the construction of $\mathbb{N}$ uses the axiom of infinity and shows the naturals as the intersection of all *inductive* sets, but we did not do this.

Theorem 108. *Let A be a set, the following are equivalent.*

a). A is infinite.

b). A contains a sequence of distinct terms.

c). there is a bijection from A to a proper subset of itself.

Proof. We prove this result in a cycle. We will show $a)$. implies $b)$., $b)$. implies $c)$., and $c)$. implies $a)$.. This is enough to assert the equivalence of all statements.

So, $a) \implies b)$. Assume that A is an infinite set. By definition, as A is not finite, $A \neq \varnothing$. Thus, as A is not empty we may pick an element, a_1, from A. If $A\backslash\{a_1\} = \varnothing$, then there exists a bijection $f : \{1\} \to A$, given by $f(1) = a_1$. But this would imply that A is finite, thus we must have $A\backslash\{a_1\} \neq \varnothing$. Therefore there exists $a_2 \in A\backslash\{a_1\}$. Clearly, $a_1 \neq a_2$. Thus, we see that subtracting the element a_1 from A leaves us with a nonempty set that we can pick an element a_2 from. Really we have just proven the base case of an induction process.

Let us make clear what our statement $P(n)$ is. The statement $P(n)$ will be

$$P(n): \ A\backslash\{a_1, \ldots, a_n\} \neq \varnothing.$$

It is implicit in the statement of $P(n)$ that $a_1, \ldots, a_n$ are all distinct as repitition of elements is not allowed in sets. Thus, we showed that $P(1)$ is true. Now, via strong induction, we assume that $P(k)$ is true for $1 \leq k \leq n$, and each $a_k \in A$ was chosen such that $a_k \in A\backslash\{a_1, \ldots, a_{k-1}\}$. (Strong induction is not really necessary here, but I am using it as it makes it clear how each a_k is chosen in step k.)

Thus, by our induction hypothesis $A\backslash\{a_1, \ldots, a_n\} \neq \varnothing$. Thus, we may choose an element in $A\backslash\{a_1, \ldots, a_n\}$ and label it a_{n+1}. If $A\backslash\{a_1, \ldots, a_{n+1}\} = \varnothing$, then there exists a bijection $f : \{1, \ldots, n+1\} \to A$, but this contradicts A being infinite. Thus, $A\backslash\{a_1, \ldots, a_{n+1}\} \neq \varnothing$. Thus, $P(n+1)$ is true. So, by induction, we have that $P(m)$ is true for all $m \in \mathbb{N}$. Then by construction, $\{a_k\}_{k=1}^{\infty}$, is a sequence of distinct terms in A. [78]

Now, to prove $b) \implies c)$. Assume that A has a sequence of distinct terms, $\{a_k\}_{k=1}^{\infty}$ contained within itself. Call $B = A\backslash\{a_k \mid k \in \mathbb{N}\}$. We will define a bijection in the following manner. Take every element from the sequence $\{a_k\}_{k=1}^{\infty}$ and shift forward by one element. In other words, map a_1 to a_2, a_2 to a_3, and so on. Define $f : A \to A\backslash\{a_1\}$ by

$$f(x) = \begin{cases} a_{k+1} & \text{if } x = a_k \\ x & \text{if } x \in B \end{cases}$$

Note that f is just the identity map on B, and f maps $\{a_k\}_{k=1}^{\infty}$ one–to–one and onto $\{a_k\}_{k=2}^{\infty}$. Thus, f is a bijection from A to $A\backslash\{a_1\}$, which is a proper subset of A.

Lastly, to prove $c) \implies a)$. Let C be a proper subset of A, i.e. $C \subset A$, and assume that there is a bijection $g : C \to A$. By way of contradiction, also assume that A is finite. Thus, there is a bijection $f : A \to \{1, 2, \ldots, m\}$ for some $m \in \mathbb{N}$. We require a lemma

Lemma 109. *If $f : A \to B$ is a bijection, and $C \subset A$ is a proper subset of A, then $f(C)$ is a proper subset of B.*

[78] The Axiom of Choice was haunting this entire argument

Proof. By way of contradiction, assume that $f(C) = B$. As C is a proper subset of A, $A \backslash C \neq \varnothing$. Thus, there exists $a \in A \backslash C$. As $f : A \to B$, let $b \in B$ be such that $b = f(a)$. By assumption, as $f(C) = B$, there exists $c \in C$ such that $f(c) = b = f(a)$, and clearly, $a \neq c$. Thus, f is not injective. Hence, we have reached a contradiction. □

Hence, by the lemma, $f : A \to \{1, 2, \ldots, m\}$ maps C to a proper subset of $\{1, 2, \ldots, m\}$. In other words, $f(C) \subset \{1, 2, \ldots, m\}$, $|f(C)| < m$, and the map $f|_C : C \to f(C)$ is also a bijection. (Notation: The symbol $f|_C$ means restricting the function $f : A \to B$ to act only on C, i.e. $f|_C$ is a function with domain C.)

Thus, $f|_C : C \to f(C)$ is a bijection from C to $f(C) \subset \{1, 2, \ldots, m\}$. And $f \circ g : C \to \{1, 2, \ldots, m\}$ is a bijection from C to $\{1, 2, \ldots, m\}$. Thus, by the transitivity of $\approx$ we have $|f(C)| = |\{1, 2, \ldots, m\}|$, but

$$m > |f(C)| = |\{1, 2, \ldots, m\}| = m,$$

which is a contradiction. [79] □

Corollary 110. *We obtain the following from this result*

- *A set is infinite if and only if it has an infinite subset.*
- *If D is an infinite set, then $|\mathbb{N}| \leq |D|$, or equivalently $\mathbb{N} \preceq D$.*

The second statement is saying under the partial order $\preceq$, $\mathbb{N}$ is least. In other words, $\mathbb{N}$ can be thought of as the smallest infinite set.

Countable Sets

Definition 65. *Let A be a set. The set A is said to be* **countably infinite** *if A and $\mathbb{N}$ have equivalent cardinality, $A \approx \mathbb{N}$. A set A is called* **countable** *if it is finite or countably infinite.*

Example 46.

Let $\mathbb{E} = \{2, 4, 6, 8, \ldots\}$ denote the even numbers as a subset of $\mathbb{N}$. Defining the map $f : \mathbb{N} \to \mathbb{E}$ by $f(n) = 2n$, we see the following. Given $m \in \mathbb{E}$, m is of the form $m = 2q$ for some $q \in \mathbb{N}$. Thus,

$$f(q) = 2q = m,$$

and thus f is surjective. For $m, n \in \mathbb{N}$, assuming $f(m) = f(n)$ leads to,

$$\begin{aligned} f(m) &= f(n) \\ 2m &= 2n \\ m &= n. \end{aligned}$$

Thus f is injective. Now that we have shown that $f : \mathbb{N} \to \mathbb{E}$ is bijective, we have $|\mathbb{N}| = |\mathbb{E}|$, and thus $\mathbb{E}$ is countable.

In particular, we have the more general result...

Theorem 111. *Every subset of a countable set is countable.*

[79] This argument in a roundabout way, is the pidgeonhole principle.

Proof. Let A be a countable set. Let E be a subset of A. If E is finite, then the result is clear. Thus, assume that E is infinite. The goal is to show that E is countably infinite. Let $f : A \to \mathbb{N}$ be a bijection. We look at $f(E)$ as a subset of $\mathbb{N}$. First of all, note that the restriction of f^{-1} to $f(E)$ is a bijection from $f(E)$ to E, written in function notation, $f^{-1}|_{f(E)} : f(E) \to E$.

Now, $f(E) \neq \varnothing$, thus by the well ordering principle $f(E)$ has a least element. We will call this least element $e_1 \in f(E)$. Now, $f(E)\backslash\{e_1\}$ is a nonempty subset of $\mathbb{N}$, and thus once again, by the well ordering principle $f(E)\backslash\{e_1\}$ has a least element. Call this e_2. We can continue on in this fashion. Define e_n to be the least element of $f(E)\backslash\{e_1, e_2, \ldots, e_{n-1}\}$. This process continues indefinitely, as if $f(E)\backslash\{e_1, \ldots, e_m\} = \varnothing$ for some $m \in \mathbb{N}$, then $f(E)$ is a finite set, which implies E is finite contradictory to our previous assumption that E is infinite. Thus, by construction, $\{e_k\}_{k=1}^{\infty}$ is a sequence of distinct elements of $f(E)$. What about $f(E)\backslash\{e_k\}_{k=1}^{\infty}$? If $f(E)\backslash\{e_k\}_{k=1}^{\infty} \neq \varnothing$, then there is some term $f \in f(E)$ and $f \neq e_k$ for any $k \in \mathbb{N}$.

Define the following subset of $\{e_k\}_{k=1}^{\infty}$,

$$B = \{e_k \mid f < e_k\}.$$

Then B is a nonempty subset of $f(E)$ and hence $\mathbb{N}$. Thus B has a least element. In other words, there is some least e_k above f. Say e_m is the smallest of the $\{e_k\}_{k=1}^{\infty}$ above f. Then, as $f \in f(E)$, we have that f is the least element of $f(E)\backslash\{e_1, \ldots, e_{m-1}\}$ and not e_m, and this contradicts the definition of e_m. Thus, it must be that $f(E)\backslash\{e_k\}_{k=1}^{\infty} = \varnothing$.

Now, define $g : \mathbb{N} \to f(E)$ by $g(n) = e_n$. As $f(E)\backslash\{e_k\}_{k=1}^{\infty} = \varnothing$ and $\{e_k\}_{k=1}^{\infty}$ is a sequence of distinct terms, we have that g is a bijection. Finally we have that $f^{-1}|_{f(E)} \circ g : \mathbb{N} \to E$ is a bijection. Thus, $|\mathbb{N}| = |E|$, and so E is countably infinite. □

There are many things we can do with countable sets as we will soon see, but first we will prove a very useful lemma.

Lemma 112. *Given a set X. For $\{A_k\}_{k\in\mathbb{N}}$ an arbitrary collection of subsets of X, we can define the following sets.*

$$B_1 = A_1, \quad B_2 = A_2\backslash A_1, \quad B_3 = A_3\backslash(A_1 \cup A_2)$$

in general define

$$B_n = A_n\backslash\left(\bigcup_{k=1}^{n-1} A_k\right).$$

Then $B_i \cap B_j = \varnothing$ if $i \neq j$ and

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n.$$

Proof. Given B_i and B_j with $i \neq j$. Either $i > j$ or $j > i$. Without loss of generality assume that $i > j$. Then $B_j = A_j\backslash\left(\bigcup_{k=1}^{j-1} A_k\right) \subseteq \bigcup_{k=1}^{i-1} A_k$. Then, by the definition of B_i, as B_i has no element in common with $\bigcup_{k=1}^{i-1} A_k$ we have that B_i has no element in common with B_j, thus $B_i \cap B_j = \varnothing$.

By construction, $B_j \subseteq A_j$ for every $j \in \mathbb{N}$, thus it is clear that

$$\bigcup_{k=1}^{\infty} B_k \subseteq \bigcup_{k=1}^{\infty} A_k.$$

Now, take $x \in \bigcup_{k=1}^{\infty} A_k$. Then $x \in A_j$ for some $j \in \mathbb{N}$. One of two things holds,

i). If $x \in A_j$ and $x \notin \bigcup_{k=1}^{j-1} A_k$, then $x \in B_j$.

ii). If $x \in A_j$ and $x \in \bigcup_{k=1}^{j-1} A_k$, then

$$D = \{k \mid x \in A_k\}$$

Then by our assumption, D is a nonempty subset of $\mathbb{N}$, and as such has a least element. Also, by our assumptions above the least element of D is at most $j-1$. Thus, call l the least element of D. Thus $x \notin A_k$ for $k < l$ and $x \in A_l$. Then $x \in B_l$.

In either case, $x \in B_k$ for some $k \in \mathbb{N}$. Thus, $x \in \bigcup_{k=1}^{\infty} B_k$. Thus,

$$\bigcup_{k=1}^{\infty} A_k \subseteq \bigcup_{k=1}^{\infty} B_k.$$

□

The point of the lemma above is that in the following proofs we lose no generality in assuming that the collections of sets that we are working with are mutually disjoint, i.e. there is no overlap between sets in the collection.

Theorem 113. *The union of two countable sets is countable.*

Proof. Let A_1 and A_2 be two countable sets. We prove the argument in three cases. By the lemma above, we lose no generality in assuming that $A_1 \cap A_2 = \varnothing$.

Case 1. Assume that A_1 and A_2 are finite sets. Then there exists $m, n \in \mathbb{N}$ such that $f : \{1, 2 \ldots, m\} \to A_1$ is a bijection and $g : \{1, 2, \ldots, n\} \to A_2$ is a bijection. We know define the following function, $h : \{1, 2, \ldots, m+n\} \to A_1 \cup A_2$ by

$$h(x) = \begin{cases} f(x), & \text{if } x \in \{1, 2, \ldots, m\} \\ g(x-m), & \text{if } x \in \{m+1, m+2, \ldots m+n\} \end{cases}$$

As there is no overlap, i.e. $A_1 \cap A_2 = \varnothing$, we have that h is a bijection as f and g are. Thus $A_1 \cup A_2$ is finite, hence countable.

Case 2. Assume that one of A_1 and A_2 is infinite. Without loss of generality, let A_1 be finite and A_2 be countably infinite. Thus there exists an $m \in \mathbb{N}$ such that $f : \{1, 2, \ldots, m\} \to A_1$ is a bijection. As A_2 is countably infinite, there exists $g : \mathbb{N} \to A_2$ that is a bijection. Define the function $h : \mathbb{N} \to A_1 \cup A_2$ by

$$h(x) = \begin{cases} f(x) & \text{if } x \in \{1, 2, \ldots, m\} \\ g(x-m) & \text{if } x \in \{m+1, m+2, \ldots\} \end{cases}$$

Thus h maps the first m numbers from $\mathbb{N}$ to A_1 and maps $\{m+1, m+2, \ldots\}$ to A_2. What we are exploiting here is that removing any finite number of elements of $\mathbb{N}$ still leaves us with a countably infinite set. To be more rigorous, subtracting a finite number of elements of $\mathbb{N}$ gives a subset of $\mathbb{N}$. By the theorem above, this subset is countable. If it was finite, then $\mathbb{N}$ could be written as the union of two finite sets which is finite. As this is a contradiction, it must be that a subset of $\mathbb{N}$ obtained by subtracting a finite number of elements from $\mathbb{N}$ is countably infinite. Once again, h is a bijection as f and g are. Thus, $A_1 \cup A_2$ is countably infinite, hence countable.

Case 3. Assume that both A_1 and A_2 are countably infinite. We begin with a lemma.

Lemma 114. *Let $\mathbb{O} = \{1, 3, 5, 7, \dots\}$ denote the odd numbers in $\mathbb{N}$. Then $\mathbb{O}$ is countable.*

Proof. Define the map $f : \mathbb{N} \to \mathbb{O}$ by $f(n) = 2n - 1$. Clearly, f maps into $\mathbb{O}$. Given any $p \in \mathbb{O}$, as p is odd, it is of the form $p = 2m - 1$ for some $m \in \mathbb{N}$. Then

$$p = 2m - 1 = f(m).$$

Thus f is surjective. Now, take $m, n \in \mathbb{N}$ and assume that $f(m) = f(n)$, then

$$\begin{aligned} f(m) &= f(n) \\ 2m - 1 &= 2n - 1 \\ 2m &= 2n \\ m &= n. \end{aligned}$$

Thus f is injective. As f is a bijection, we have that $\mathbb{O}$ is countably infinite, hence countable. □

As A_1 and A_2 are countably infinite, there are bijections $f_1 : \mathbb{N} \to A_1$ and $f_2 : \mathbb{N} \to A_2$. From an example above and the lemma directly above, we have that the even numbers $\mathbb{E}$ and the odd numbers $\mathbb{O}$ are countable. Thus there exists bijections $g_1 : \mathbb{E} \to \mathbb{N}$ and $g_2 : \mathbb{O} \to \mathbb{N}$. Because of this $f_1 \circ g_1 : \mathbb{E} \to A_1$ is a bijection from the even numbers to A_1, and $f_2 \circ g_2 : \mathbb{O} \to A_2$ is a bijection from the odd numbers $\mathbb{O}$ to A_2. Now, we define a function $h : \mathbb{N} \to A_1 \cup A_2$ by

$$h(x) = \begin{cases} (f_1 \circ g_1)(x) & \text{if } x \in \mathbb{E} \\ (f_2 \circ g_2)(x) & \text{if } x \in \mathbb{O} \end{cases}$$

As $A_1 \cap A_2 = \varnothing$ and as $f_1 \circ g_1$ and $f_2 \circ g_2$ are bijections, we have that h is a bijection. In words, h maps the even numbers to A_1 and the odd numbers to A_2. As h is a bijection from $\mathbb{N}$ to $A_1 \cup A_2$, $A_1 \cup A_2$ is countably infinite, and therefore countable. □

This result can be used immediately to derive a slightly stronger result. In particular, the previous theorem proves the nontrivial part of the induction process in the next theorem.

Theorem 115. *For any $m \in \mathbb{N}$, if $\{A_k\}_{k=1}^m$ is a collection of countable sets, i.e. each A_k is countable for $1 \leq k \leq m$, then the union*

$$\bigcup_{k=1}^{m} A_k$$

is countable.

Proof. We will prove this result by induction. Let $P(n)$ be the statement that

$$P(n) : \bigcup_{k=1}^{n} A_k \text{ is countable.}$$

Let us first check the base case. By assumption, every member of $\{A_k\}_{k=1}^m$ is countable, thus A_1 is countable. Hence, $P(1)$ is true.

Now, assume that $P(n)$ is true. Thus, assume that $\bigcup_{k=1}^n A_k$ is countable, for a collection of n countable sets. Now, let $\{A_k\}_{k=1}^{n+1}$ be a collection of $n + 1$ countable sets. As,

$$\bigcup_{k=1}^{n+1} A_k = \left[\bigcup_{k=1}^{n} A_k\right] \cup A_{n+1}.$$

By the induction hypothesis $\bigcup_{k=1}^{n} A_k$ is a countable set. By the above, $\bigcup_{k=1}^{n+1} A_k$ is written as the union of $\bigcup_{k=1}^{n} A_k$ and A_{n+1}, which are two countable sets. Thus, by the previous theorem, $\bigcup_{k=1}^{n+1} A_k$ is countable as it is the union of two countable sets. Thus, $P(n+1)$ is true. Thus, by the principle of mathematical induction, we have that $P(m)$ is true for all $m \in \mathbb{N}$. □

At this point, you may have noticed that many of these proofs seem overly technical or overly abstract, but the importance of these early results is that they give a library or a toolbox or results we can use to give much shorter proofs in the future. In particular, because of the result above, we can prove that a set is countable if we can write it as a finite union of countable sets instead of explicitly finding a bijection between the set in question and $\mathbb{N}$.

Example 47. *The integers $\mathbb{Z}$ are countable.*

Using the theorem we have just proved if we let $\mathbb{N}^-$ denote the negative natural numbers, i.e. $\mathbb{N}^- = \{-1, -2, -3, \dots\}$ then clearly, $\mathbb{N}^-$ is countable. As

$$\mathbb{Z} = \mathbb{N}^- \cup \{0\} \cup \mathbb{N}.$$

Thus $\mathbb{Z}$ is the union of the three countable sets $\mathbb{N}$, $\mathbb{N}^-$, and $\{0\}$. Thus, by the prior theorem $\mathbb{Z}$ is countable.

The last few theorems have really been building to this result. We now extend the prior theorem to the countably infinite case.

Theorem 116. *Let $\{A_k\}_{k=1}^{\infty}$ be a countably infinite collection of countable sets A_k. Then*

$$\bigcup_{k=1}^{\infty} A_k \text{ is countable.}$$

Proof. Let us begin by defining the following set.

$$J = \{k \mid A_k \text{ is countably infinite}\}.$$

In other words, J is the set of the indices k such that A_k is not finite and therefore countably infinite. We prove the main result by cases.

Case 1: The set J is finite.

As J is finite, $\mathbb{N}\backslash J$ as a not finite subset of $\mathbb{N}$ is countably infinite. (This follows from a prior theorem) And for every $k \in \mathbb{N}\backslash J$, A_k is finite. We will show that

$$\bigcup_{k \in \mathbb{N}\backslash J} A_k \text{ is countable.}$$

As $\mathbb{N}\backslash J$ is countable, there exists a bijection $g : \mathbb{N}\backslash J \to \mathbb{N}$. In other words, we can label subscripts of k in $\mathbb{N}\backslash J$ as $k_1, k_2, k_3, \dots$ and so on. From our definition of bijective, a set being countably infinite is analogous to saying you can label all terms of a countably infinite set with natural numbers. Thus, assume all elements of $\mathbb{N}\backslash J$ have been labeled in this way, i.e. assume

$$\{A_k\}_{k \in \mathbb{N}\backslash J} = \{A_{k_n}\}_{n=1}^{\infty}.$$

In this case, A_{k_n} is finite for all $n \in \mathbb{N}$. Thus for every $n \in \mathbb{N}$ there exists some $m_n \in \mathbb{N}$ and a bijection $f_n : \{1, 2, \dots, m_n\} \to A_{k_n}$. Using these maps we will construct a bijection from $\mathbb{N}$ to

$\bigcup_{k\in\mathbb{N}\setminus J} A_k$ as follows, map the first m_1 natural numbers to A_{k_1} by f_1, the next m_2 natural numbers to A_{k_2} by f_2, and so on. The map h would look as follows.

$$h(x) = \begin{cases} f_1(x) & \text{if } x \in \{1, 2, \ldots, m_1\} \\ f_2(x - m_1) & \text{if } x \in \{m_1 + 1, \ldots, m_1 + m_2\} \\ \vdots & \vdots \\ f_n\left(x - \sum_{k=1}^{n-1} m_k\right) & \text{if } x \in \{\left(\sum_{k=1}^{n-1} m_k\right) + 1, \ldots, \sum_{k=1}^{n} m_k\} \\ \vdots & \vdots \end{cases}$$

As each f_n is a bijection from $\{1, 2, \ldots, m_n\}$ to A_{k_n}, we have that h is a bijection, and thus

$$\bigcup_{k\in\mathbb{N}\setminus J} A_k = \bigcup_{n=1}^{\infty} A_{k_n} \text{ is countable}$$

In this case, as J is finite. We have that

$$\bigcup_{n=1}^{\infty} A_n = \left[\bigcup_{n\in J} A_n\right] \cup \left(\bigcup_{n\in\mathbb{N}\setminus J} A_n\right)$$

is a finite union of countable sets, as J is a finite set and $\bigcup_{k\in\mathbb{N}\setminus J} A_k$ is countable.

Case 2: The set J is countably infinite.

The set $\mathbb{N}\setminus J$ is the following set,

$$\mathbb{N}\setminus J = \{k \mid A_k \text{ is finite}\}.$$

If $\mathbb{N}\setminus J$ is finite, then $\bigcup_{k\in\mathbb{N}\setminus J} A_k$ is a finite union of finite sets, hence countable. If $\mathbb{N}\setminus J$ is infinite, then $\bigcup_{k\in\mathbb{N}\setminus J} A_k$ is a countable union of finite sets. In this case, the same method in case 1 can be used to create a bijection from $\mathbb{N}$ to $\bigcup_{k\in\mathbb{N}\setminus J} A_k$. Thus, $\bigcup_{k\in\mathbb{N}\setminus J} A_k$ is countably infinite, hence countable. In either situation $\bigcup_{k\in\mathbb{N}\setminus J} A_k$ is countable. As in a prior theorem we showed that the union of two countable sets is countable, we only must now show that $\bigcup_{k\in J} A_k$ is countable.

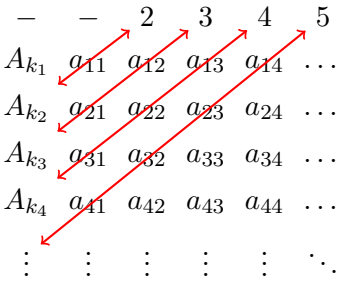
From a prior theorem, as $J \subseteq \mathbb{N}$ and J is infinite, we have that J is countable. Hence, there is a bijection $f : \mathbb{N} \to J$. In other words, we can label the indices in J by the natural numbers, i.e.

$$\{A_k\}_{k\in J} = \{A_{k_n}\}_{n=1}^{\infty}.$$

Now, as each A_{k_n} is countably infinite, there exists bijections $g_n : \mathbb{N} \to A_{k_n}$ for every $n \in \mathbb{N}$. In other words, for each index k_n we can label the elements of A_{k_n} by the natural numbers. We do so as follows,

$$\begin{array}{rl} A_{k_1}: & a_{11}a_{12}a_{13}\ldots \\ A_{k_2}: & a_{21}a_{22}a_{23}\ldots \\ A_{k_3}: & a_{31}a_{32}a_{33}\ldots \\ \vdots & \vdots \ddots \ddots \end{array}$$

To construct a bijection between $\mathbb{N}$ and this array of elements, we will count along diagonals.



Let's note some patterns we see here.

- The only element in diagonal 1 is a_{11} and $1+1=2$. The only elements in diagonal 2 are a_{12} and a_{21} and $1+2=2+1=3$. In general, the terms in diagonal m are the terms a_{ij} with $i+j=m+1$.
- Diagonal 1 contains one element, diagonal 2 contains two elements, and so on. In general diagonal m contains m elements.
- Before reaching diagonal 2 we have counted 1 element. Before reaching diagonal 3 we have counted 3 elements. Before reaching diagonal 4 we have counted 6 elements. In general, before reaching diagonal m we have counted

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

  elements.

Thus, we construct a map from $\mathbb{N}$ to $\bigcup_{n=1}^{\infty}A_{k_n}$ in the following manner. We count by diagonal first and then column second. Thus 1 maps to a_{11}, and then 2 maps to a_{21} and 3 to a_{12}, then 4 to a_{31}, 5 to a_{22}, 6 to a_{13}, and so on and so on. For a general $n\in\mathbb{N}$, let S be the following set,

$$S=\left\{p\in\mathbb{N}\mid n\leq\sum_{l=1}^{p}l\right\},$$

then $S\neq\varnothing$ as n is a finite number for any fixed $n\in\mathbb{N}$. Hence, by the well-ordering principle, S has a least element, call it m. And m has the property that

$$\sum_{l=1}^{m-1}l<n\leq\sum_{l=1}^{m}l.$$

(Note: as convention $\sum_{l=1}^{0}l=0$ in this proof.) The number m will be the diagonal that n is mapped to. As $\sum_{l=1}^{m-1}l$ many naturals were used to map to the previous $m-1$ diagonals, it is sensible that n will map to diagonal m. If n is exactly one larger than $\sum_{l=1}^{m-1}l$, then n will map to the element in column 1 of diagonal m. If n is two larger than $\sum_{l=1}^{m-1}l$, then n will map to the element in column 2 of diagonal m. Thus, the column entry that n maps to is $n-\sum_{l=1}^{m-1}l$. If we let a_{ij} denote the element that n maps to, recall that $i+j=m+1$ if a_{ij} is in diagonal m. Thus, we have that

$$i+\left(n-\sum_{l=1}^{m-1}l\right)=m+1$$

Thus n maps to row $i = m + 1 - n + \sum_{l=1}^{m-1} l$. Using that $\sum_{l=1}^{m-1} l = \frac{(m-1)m}{2}$, we can write the map $h : \mathbb{N} \to \bigcup_{n=1}^{\infty} A_{k_n}$, as

$$h(n) = a_{m+1-n+\frac{(m-1)m}{2}, n-\frac{(m-1)m}{2}}.$$

Below we see how each natural number maps to elements in the collection $\bigcup_{n=1}^{\infty} A_{k_n}$.

$$\begin{array}{cccccccc}
 & _ & _ & _ & _ & _ & _ & \\
A_{k_1} & \overset{1}{a_{11}} & \overset{3}{a_{12}} & \overset{6}{a_{13}} & \overset{10}{a_{14}} & \ldots \\
A_{k_2} & \overset{2}{a_{21}} & \overset{5}{a_{22}} & \overset{9}{a_{23}} & a_{24} & \ldots \\
A_{k_3} & \overset{4}{a_{31}} & \overset{8}{a_{32}} & a_{33} & a_{34} & \ldots \\
A_{k_4} & \overset{7}{a_{41}} & a_{42} & a_{43} & a_{44} & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

By construction of the map h, it is clear that h is surjective. Now, take $p, q \in \mathbb{N}$, and assume that $h(p) = h(q)$. As $h(p) = h(q)$, we have that p and q map to the same diagonal, call it m. Then

$$a_{m+1-p+\frac{(m-1)m}{2}, p-\frac{(m-1)m}{2}} = a_{m+1-q+\frac{(m-1)m}{2}, q-\frac{(m-1)m}{2}}.$$

From this we have two equations.

$$m + 1 - p + \frac{(m-1)m}{2} = m + 1 - q + \frac{(m-1)m}{2}, \qquad p - \frac{(m-1)m}{2} = q - \frac{(m-1)m}{2}$$

and both imply that $p = q$. Thus, h is injective. Thus, $h : \mathbb{N} \to \bigcup_{n=1}^{\infty} A_{k_n}$ is a bijection, thus $\bigcup_{k \in J} A_k$ is countable. So,

$$\bigcup_{n=1}^{\infty} A_n = \left[\bigcup_{k \in J} A_k\right] \cup \left(\bigcup_{k \in \mathbb{N} \setminus J} A_k\right).$$

Thus, $\bigcup_{n=1}^{\infty} A_n$, is the countable union of two countable sets and is therefore countable. □

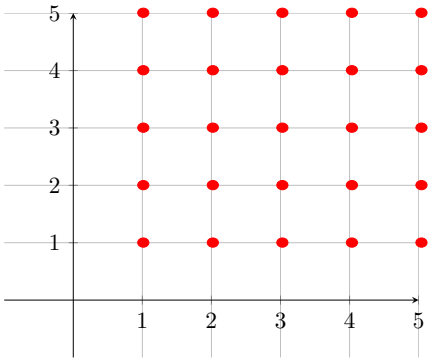
Let us immediately apply this result to prove another theorem.

Theorem 117. *The set $\mathbb{N} \times \mathbb{N}$ is countable.*

Proof. Remember that the cartesian product of $\mathbb{N}$ with itself is,

$$\mathbb{N} \times \mathbb{N} = \{(m, n) \mid m, n \in \mathbb{N}\},$$

the collection of ordered pairs (m, n) with $m, n \in \mathbb{N}$. For the sake of intuition, we can graph $\mathbb{N} \times \mathbb{N}$ as



Define the following sets, $A_k = \{k\} \times \mathbb{N}$ for each $k \in \mathbb{N}$. We have that A_k is countable for each $k \in \mathbb{N}$, as $f_k : \mathbb{N} \to A_k$ given by $f_k(n) = (k, n)$ is a bijection. Thus, A_k is countable for each $k \in \mathbb{N}$. As,

$$\mathbb{N} \times \mathbb{N} = \bigcup_{k=1}^{\infty} A_k,$$

we have that $\mathbb{N} \times \mathbb{N}$ is a countable union of countable sets and is therefore countable. □

We can finally give an argument for why $\mathbb{Q}$ is countable.

Example 48. *The rational numbers, $\mathbb{Q}$, are countable.*

Let $\mathbb{Q}^+$ denote the positive rational numbers and $\mathbb{Q}^-$ denote the negative rational numbers. Thus,

$$\mathbb{Q}^+ = \{\frac{a}{b} \mid a, b \in \mathbb{N},\ b \neq 0,\ \gcd(a, b) = 1\}.$$

We first show that $\mathbb{Q}^+$ is countable. Define $f : \mathbb{Q}^+ \to \mathbb{N} \times \mathbb{N}$ by $f(\frac{a}{b}) = (a, b)$. For $\frac{a}{b}, \frac{p}{q} \in \mathbb{Q}^+$, if we assume $f(\frac{a}{b}) = f(\frac{p}{q})$, then

$$f(\frac{a}{b}) = f(\frac{p}{q})$$
$$(a, b) = (p, q)$$

which implies that $a = p$ and $b = q$, thus $\frac{a}{b} = \frac{p}{q}$. Thus f is injective. This implies that $\mathbb{Q}^+ \preceq \mathbb{N} \times \mathbb{N} \approx \mathbb{N}$.

Now, define the map $g : \mathbb{N} \to \mathbb{Q}^+$ by $g(n) = n = \frac{n}{1}$. Clearly, g is injective, thus $\mathbb{N} \preceq \mathbb{Q}^+$. Now, the CSB theorem implies that $\mathbb{Q}^+ \approx \mathbb{N}$. Thus $\mathbb{Q}^+$ is countable.

Lastly, define the map $f : \mathbb{Q}^+ \to \mathbb{Q}^-$ by $f(x) = -x$. Clearly, f is a bijection. Thus, $\mathbb{Q}^- \approx \mathbb{Q}^+$, and so $\mathbb{Q}^-$ is countable. Therefore, as

$$\mathbb{Q} = \mathbb{Q}^- \cup \{0\} \cup \mathbb{Q}^+,$$

it is clear that $\mathbb{Q}$ is a finite union of countable sets and is therefore countable.

Theorem 118. *For any $m \in \mathbb{N}$, $\mathbb{N}^m$ is countable.*

Proof. Recall that

$$\mathbb{N}^m = \underbrace{\mathbb{N} \times \mathbb{N} \times \cdots \times \mathbb{N}}_{m \text{ times}}.$$

We will prove this result by induction. Let $P(n)$ be the statement

$$P(n) :\ \mathbb{N}^n \text{ is countable}$$

The base case $n = 1$, $P(1)$ is just the statement $\mathbb{N}$ is countable, which is true by definition. Thus, the base case holds. Now, assume that $\mathbb{N}^n$ is countable, and let us deduce that $\mathbb{N}^{n+1}$ is countable.

As $\mathbb{N}^n$ is countable, there exists a bijection $f : \mathbb{N}^n \to \mathbb{N}$. Using this, we define a map $g : \mathbb{N}^{n+1} \to \mathbb{N} \times \mathbb{N}$ as follows,

$$g((m_1, m_2, \ldots, m_{n+1})) = (f((m_1, m_2, \ldots, m_n)), m_{n+1})$$

We show that g is a bijection. For any $(p, q) \in \mathbb{N}\times\mathbb{N}$, as f is a bijection, there exists $(m_1, m_2, \ldots, m_n) \in \mathbb{N}^n$ such that $f((m_1, \ldots, m_n)) = p$. Thus,

$$g((m_1, \ldots, m_n, q)) = (p, q).$$

So, g is surjective. Now for $(p_1, \ldots, p_{n+1}), (q_1, \ldots, q_{n+1}) \in \mathbb{N}^{n+1}$ with $g((p_1, \ldots, p_{n+1})) = g((q_1, \ldots, q_{n+1}))$, we have

$$(f((p_1, \ldots, p_n), p_{n+1})) = (f((q_1, \ldots, q_n)), q_{n+1})$$

Thus, $f((p_1, \ldots, p_n)) = f((q_1, \ldots, q_n))$ and $p_{n+1} = q_{n+1}$. The injectivity of f gives that $p_k = q_k$ for $1 \leq k \leq n$. Thus,

$$(p_1, \ldots, p_{n+1}) = (q_1, \ldots, q_{n+1}).$$

and so g is injective. Thus g is bijective, so $\mathbb{N}^{n+1} \approx \mathbb{N} \times \mathbb{N}$. From the prior theorem, $\mathbb{N} \times \mathbb{N} \approx \mathbb{N}$, hence $\mathbb{N}^{n+1} \approx \mathbb{N}$. Thus $\mathbb{N}^{n+1}$ is countable. Therefore, by the principle of mathematical induction, we have that $\mathbb{N}^m$ is countable for every $m \in \mathbb{N}$. □

Warning! Note that the above theorem does **not** imply $\mathbb{N}^{\mathbb{N}}$ is countable. The theorem above explicitly states that $\mathbb{N}^m$ is countable for any finite value of m. We will soon see that even a two element set 'raised' to the $\mathbb{N}$ (in the sense brought up at the end of lecture 6) is not countable.

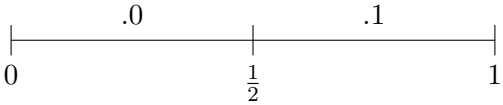
$|\{0, 1\}^{\mathbb{N}}| = |\mathbb{R}|$

Theorem 119. $\{0, 1\}^{\mathbb{N}} \approx \mathbb{R}$.

Proof. Every natural number, n, has a unique binary expansion, i.e. there is a finite sequence of zeros and ones that represent n. To be more explicit, for each $n \in \mathbb{N}$, there exists an $m \in \mathbb{N}$ and a sequence $a_m a_{m-1} \cdots a_1 a_0$ with $a_k \in \{0, 1\}$, such that

$$n = \sum_{k=0}^{m} a_k 2^k.$$

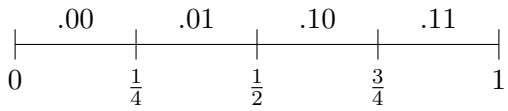
We extend this notion of a binary expansion to $\mathbb{R}$. Let us first explain how any number in the interval $[0, 1]$ can be expanded in negative powers of 2. For $x \in [0, 1]$, we cut the interval $[0, 1]$ into halves and ask which half x falls into. If $x \in [0, \frac{1}{2}]$ we define $a_{-1} = 0$, and if $x \in [\frac{1}{2}, 1]$ we define $a_{-1} = 1$. This is shown in the following picture.



We now continue this process to define a_{-2}.

- If $a_{-1} = 0$, then we break the interval $[0, \frac{1}{2}]$ into halves. If $x \in [0, \frac{1}{4}]$ we define $a_{-2} = 0$, and if $x \in [\frac{1}{4}, \frac{1}{2}]$ we define $a_{-2} = 1$.
- If $a_{-1} = 1$, then we break the interval $[\frac{1}{2}, 1]$ into halves. If $x \in [\frac{1}{2}, \frac{3}{4}]$ we define $a_{-2} = 0$, and if $x \in [\frac{3}{4}, 1]$ we define $a_{-2} = 1$.

This shown in the following picture.



In general, to define the next term in the sequence, you take the interval that x lies in at the current stage and break it into halves. If x lies in the left half, then you define the next term to be 0. If x lies in the right half, then you define the next term to be 1. Continuing this process indefinitely gives a constructive manner to build a sequence of zeros and ones that represents a nested family of intervals that decrease in length down to zero. Thus, this process gives a way to build a 0–1 sequence for every real $x \in [0,1]$. To be more explicit, for any $x \in [0,1]$ there exists a sequence $\{a_{-k}\}_{k=1}^{\infty}$ with $a_{-k} \in \{0,1\}$ and

$$x = \sum_{k=1}^{\infty} \frac{a_{-k}}{2^k}.$$

This is called the binary expansion for a number in the interval $[0,1]$.

As we mentioned before every $n \in \mathbb{N}$ has a finite binary expansion $n = a_m a_{m-1} \cdots a_1 a_0$. Given any $x \in \mathbb{R}^+ = (0,\infty)$, we have that $x = \lfloor x \rfloor + y$, where $\lfloor x \rfloor$ is the flooring function or least integer function, and $y \in [0,1]$. As $\lfloor x \rfloor \in \mathbb{N}$, putting the binary expansion of an abitrary natural number together with the binary expansion of an arbitrary number in $[0,1]$, we have that every $x \in \mathbb{R}^+$ has a binary expansion.

$$x = a_m a_{m-1} \cdots a_1 a_0 . a_{-1} a_{-2} \cdots .$$

To be more explicit, for every $x \in \mathbb{R}^+$, there exists an $m \in \mathbb{N}$ and a sequence $\{a_k\}$. $-\infty < k \leq m$ with $a_k \in \{0,1\}$, such that

$$x = \sum_{k=-\infty}^{m} a_k 2^k.$$

At this point, I have given an argument for the existence of a binary expansion for an element $x \in \mathbb{R}^+$, but I have made no mention of uniqueness. To save time we will not go through this next part in a completely rigorous fashion, but in general the binary expansion of a number $x \in \mathbb{R}^+$ is not unique if x has a particular form. The argument above fails to give a unique binary expansion for any number x that is a dyadic rational number. The dyadic rationals (here called $\mathbb{D}$) are

$$\mathbb{D} = \{\frac{p}{2^q} \mid p, q \in \mathbb{Z},\ \gcd(p, 2^q) = 1\}.$$

In particular, the dyadic rationals are the members of the rational numbers whose denominator is a power of 2, for example $\frac{1}{2}$, $\frac{3}{8}$, $\frac{127}{256}$, etc. The reason that dyadic rationals do not have a unique binary expansion is that they lie 'on the cut' in our process of creating a binary expansion.

To be more clear, for an example look at $\frac{1}{2} \in [0,1]$. To find the binary expansion of $\frac{1}{2}$, we cut $[0,1]$ into the halves $[0,\frac{1}{2}]$ and $[\frac{1}{2},1]$ and ask which half is $\frac{1}{2}$ in? Well, both. In other words, we are left with two options.

- We say $\frac{1}{2}$ is in the left half $[0,\frac{1}{2}]$ and then choose the right half for all future cuts, i.e. the binary expansion of $\frac{1}{2}$ is

$$\frac{1}{2} = .01111111111\ldots$$

- We say $\frac{1}{2}$ is in the right half $[\frac{1}{2}, 1]$ and then choose the left half for all future cuts, i.e. the binary expansion of $\frac{1}{2}$ is
$$\frac{1}{2} = .1 = .1000000000\ldots$$

Well, both of these expansions are valid, and this shouldn't be surprising to anyone that knows geometric series

$$\begin{aligned} .01111111\ldots &= \sum_{k=2}^{\infty} \frac{1}{2^k} = \frac{1}{4}\sum_{k=0}^{\infty} \frac{1}{2^k} \\ &= \frac{1}{4}\left[\frac{1}{1-\frac{1}{2}}\right] = \frac{1}{4}\cdot 2 \\ &= \frac{1}{2} = (1)\cdot\frac{1}{2} + \sum_{k=2}^{\infty}(0)\frac{1}{2^k} = .10000000\ldots \end{aligned}$$

This is the case for every dyadic rational. For every dyadic rational, at some point in the binary sequence expansion, the tail of the sequence will look like $0111111\ldots$ or $1000000\ldots$. So, we make a choice! For every dyadic rational, we choose as it's binary sequence expansion the expansion with the tail of the form $1000000\ldots$.

So, back to the big picture here. Every $x \in \mathbb{R}^+$ that is not a dyadic rational has a unique binary expansion. And for every $x \in \mathbb{D} \cap \mathbb{R}^+$ we choose its binary expansion to be of the form above. Thus, we have really constructed an injective function $f : \mathbb{R}^+ \to \{0,1\}^{\mathbb{N}}$. As the reader can check, the map $g : \mathbb{R} \to \mathbb{R}^+$ given by $g(x) = e^x$ is a bijection. Thus, $f \circ g : \mathbb{R} \to \{0,1\}^{\mathbb{N}}$ is an injection from $\mathbb{R}$ to $\{0,1\}^{\mathbb{N}}$. Thus, $\mathbb{R} \preceq \{0.1\}^{\mathbb{N}}$.

Making use of the standard decimal expansion (base 10), we construct a map from $\{0,1\}^{\mathbb{N}}$ to $\mathbb{R}$. In particular, define $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ by

$$f(a_1a_2a_3\ldots) = \sum_{k=1}^{\infty} \frac{a_k}{10^k}.$$

Let $a = a_1a_2a_3\ldots$ and $b = b_1b_2b_3\ldots$ be two elements of $\{0,1\}^{\mathbb{N}}$. Assume that $a \neq b$, thus

$$\{j \in \mathbb{N} \mid a_j \neq b_j\}$$

is a nonempty subset of the naturals, and hence by the well ordering principle, has a least element, call it p. Thus p is the first term in the sequences for which a and b differ, i.e. $a_k = b_k$ for $1 \leq k \leq p-1$. Now, without loss of generality, assume that $a_p = 1$ and $b_p = 0$. We want to show that f is injective, thus we aim to show that $f(a) \neq f(b)$.

Well, $f(a)$ has a term of $\frac{1}{10^p}$ and $f(b)$ does not, so intuition makes it seem that $f(a) \geq f(b)$. The question is, in the worst possible scenario, could $f(b)$ 'make up' this missing $\frac{1}{10^p}$ term with the remaining terms in the sequence of b. Well, let me be clearer about worst possible case scenario. Worst possible case scenario would be if $f(a)$ stopped growing after the pth term, i.e. $a_k = 0$ for $k > p$, and $f(b)$ grew as much as possible, i.e. $b_k = 1$ for $k > p$. In this case, the amount $f(b)$ gains from $b_k = 1$ for $k > p$ is

$$\sum_{k=p+1} \frac{1}{10^k} = \frac{1}{10^{p+1}}\sum_{k=0}^{\infty} \frac{1}{10^k} = \frac{1}{10^{p+1}}\left[\frac{1}{1-\frac{1}{10}}\right] = \frac{1}{10^{p+1}}\cdot\frac{10}{9} = \frac{1}{9\cdot 10^p} < \frac{1}{10^p}$$

So $f(b)$ can not 'catch up' to $f(a)$ because of the $\frac{1}{10^p}$ gap, even in the worst possible case scenario. Thus, $f(a) \neq f(b)$. So, f is injective. Thus, $\{0,1\}^{\mathbb{N}} \preceq \mathbb{R}$.

Thus, the Cantor-Schroder-Bernstein theorem immediately implies that $\{0,1\}^{\mathbb{N}} \approx \mathbb{R}$. □

Corollary 120. *$\mathbb{R}$ is uncountable.*

More on uncountable sets

Example 49. *Now that we have shown that $\mathbb{R}$ is uncountable, we can build up some more examples of uncountable sets very quickly.*

a). The interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ is uncountable. The function $\arctan : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ is a bijection, thus $(-\frac{\pi}{2}, \frac{\pi}{2}) \approx \mathbb{R}$.

b). The interval $(0,1)$ is uncountable. The function $g : (0,1) \to (-\frac{\pi}{2}, \frac{\pi}{2})$ given by $g(x) = \pi x - \frac{\pi}{2}$ is a bijection as the reader can verify. Thus $(0,1) \approx (-\frac{\pi}{2}, \frac{\pi}{2})$.

c). For $a, b \in \mathbb{R}$ with $a < b$, the interval (a,b) is uncountable. The function $h : (0,1) \to (a,b)$ given by $h(x) = (b-a)x + a$ is a bijection as the reader can check. Thus $(0,1) \approx (a,b)$.

Example 50. *For $a, b \in \mathbb{R}$ with $a < b$, the interval $[a,b]$ is uncountable.*

It is clear that $(a,b) \subseteq [a,b] \subseteq \mathbb{R}$, thus

$$(a,b) \preceq [a,b] \preceq \mathbb{R} \approx (a,b)$$

Thus, the CSB theorem immediately gives that $[a,b] \approx (a,b) \approx \mathbb{R}$. Thus, $[a,b]$ is uncountable. Later, I will give an explicit bijection between $\mathbb{R}$ and $[a,b]$.

Theorem 121. *The power set of the naturals is uncountable, i.e. $P(\mathbb{N}) \approx \mathbb{R}$.*

Proof. To prove this result, we will actually show $P(\mathbb{N}) \approx \{0,1\}^{\mathbb{N}}$. Let us define a map from $P(\mathbb{N})$ to $\{0,1\}^{\mathbb{N}}$. For an element $A \in P(\mathbb{N})$, A contains some of the elements of $\mathbb{N}$. Let us define $f : P(\mathbb{N}) \to \{0,1\}^{\mathbb{N}}$ in the following manner.

$$f(A) = a_1 a_2 a_3 \ldots,$$

where the a_k are defined by

$$a_k = \begin{cases} 1 & \text{if } k \in A \\ 0 & \text{if } k \notin A \end{cases}$$

Thus for some examples,

$$\begin{aligned} f(\{2\}) &= 01000000\ldots \\ f(\mathbb{E}) &= 010101010101\ldots \\ f(\mathbb{O}) &= 1010101010\ldots \\ f(\{2,3,5,7\}) &= 01101010000000\ldots \end{aligned}$$

We show that f is a bijection.

Given $b \in \{0,1\}^{\mathbb{N}}$, i.e. $b = b_1 b_2 b_3 \ldots$, define the following set

$$B = \{j \in \mathbb{N} \mid b_j = 1\}.$$

then it is clear that $B \in P(\mathbb{N})$ and $f(B) = b$. Thus f is surjective.

Now, take $A, B \in P(\mathbb{N})$ and assume that $f(A) = f(B)$. Denote

$$f(A) = a_1 a_2 a_3 \ldots$$
$$f(B) = b_1 b_2 b_3 \ldots$$

then $a_k = b_k$ for all $k \in \mathbb{N}$. Thus $k \in A$ iff $k \in B$ for all $k \in \mathbb{N}$, but this is equivalent to saying $A = B$. Thus f is injective. □

We lastly present a proof of the explicit bijection between an interval of the form (a, b) and an interval of the form $[c, d]$.

Lemma 122. *For any two real numbers, $a, b \in \mathbb{R}$ with $a < b$, $|(a, b)| = |(0, 1)|$.*

Proof. The line containing the points $(a, 0)$ and $(b, 1)$ has the equation

$$y = \frac{1}{b-a}x - \frac{a}{b-a}.$$

As this equation is linear, it is easy to see that it is a bijection from the interval (a, b) to the interval $(0, 1)$. □

Replacing the symbols a and b in the previous proof with any two other symbols (for example C and D) also shows that $|[C, D]| = |[0, 1]|$. Thus, it is sufficient for us to prove that $|(0, 1)| = |[0, 1]|$.

Define the map $F : (0, 1) \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ by

$$F(x) = \pi\left(x - \frac{1}{2}\right).$$

It is easy to see that F is a bijection.

Define $G : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$ by $G(x) = \tan x$. Via calculus we have the following,

$$\lim_{x \to -\frac{\pi}{2}^+} \tan x = -\infty \qquad \lim_{x \to \frac{\pi}{2}^-} \tan x = \infty.$$

As G is continuous on the domain $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, by the Intermediate Value Theorem, we have that G is onto. Now, take x and y in the domain of G with $G(x) = G(y)$. So,

$$\begin{aligned} G(x) &= G(y) \\ \tan x &= \tan y \\ \tfrac{\sin x}{\cos x} &= \tfrac{\sin y}{\cos y} \\ \sin x \cos y &= \sin y \cos x \\ \sin x \cos y - \cos x \sin y &= 0 \\ \sin(x - y) &= 0. \end{aligned}$$

Now, as $-\dfrac{\pi}{2} < x, y < \dfrac{\pi}{2}$, we have that $-\pi < x - y < \pi$. Since $\sin(x - y) = 0$, we have that $x - y = k\pi$ for some $k \in \mathbb{Z}$. As $-\pi < x - y < \pi$, it must be that $x - y = 0$. So, $x = y$, and G is 1–1. Thus G is a bijection.

Now take $f = G \circ F$, so $f : (0, 1) \to \mathbb{R}$. As f is a composition of bijections, f is itself a bijection.

Now define, $g : \mathbb{R} \to [0, 1]$ in the following manner.

$$g(x) = \begin{cases} f^{-1}(x) & \text{if } x \in \mathbb{R}\backslash[\{0\} \cup \mathbb{N}] \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x = 1 \\ f^{-1}(x - 2) & \text{if } x \in [2, \infty) \cap \mathbb{N} \end{cases}$$

As the image of $f^{-1}(x - 2)$ on $[2, \infty) \cap \mathbb{N}$ will return the same values of $f^{-1}(x)$ on $\mathbb{N} \cup \{0\}$, we see from f being a bijection that g will map $\mathbb{R}\backslash[\{0, 1\}]$ to $(0, 1)$ in a 1–1 and onto fashion. Thus, it is clear that g is a bijection from $\mathbb{R}$ to $[0, 1]$.

Then taking $h = g \circ f$, we have that h is a composition of bijections, and thus is a bijection with $h : (0, 1) \to [0, 1]$. Thus $|(0, 1)| = |[0, 1]|$.

Proof of the Cantor–Schröder–Bernstein Theorem

Here we provide two proofs of the Cantor–Schröder–Bernstein Theorem.

Theorem 123. *Let A, B be sets. If $f : A \to B$ and $g : B \to A$ are 1–1 functions, then there exists a bijection $h : A \to B$.*

I will be considering 0 as a natural number throughout this proof.

Proof. Begin by defining the following sets. Let $A_0 = A$, and $B_0 = B$. Now define the following, for all $n \in \mathbb{N}$.

$$\begin{aligned} A_{2n} &= g(A_{2n-1}), & A_{2n+1} &= f(A_{2n}) \\ B_{2n} &= f(B_{2n-1}), & B_{2n+1} &= g(B_{2n}). \end{aligned}$$

So, just to get our definitions straight,

$$A_{2n} = g(A_{2n-1}) = \{g(x) \mid x \in A_{2n-1}\} \subseteq A.$$

There are some things we can notice here. First of all, the function f will always map subsets of A to subsets of B, and similarly, g will always map subsets of B to subsets of A. It is not true in general that a function will map a proper subset to a a proper subset, for example,

$$[0, \infty) \subset \mathbb{R} \text{ but } f([0, \infty)) = f(\mathbb{R}),$$

if $f(x) = x^2$. However, if the function is 1–1, then it will map proper subsets to proper subsets. As $B_1 = g(B)$, and g is not onto (as the proof would be trivial in this case), we have that $A_0 \supseteq B_1$. If we apply f to both sides, we obtain that $f(A_0) \supseteq f(B_1)$, or in other words, $A_1 \supseteq B_2$. And applying g to both sides of $A_1 \supseteq B_2$ gives that $A_2 \supseteq B_3$. Continuing in this fashion, we see that for all $n \in \mathbb{N}$, (I'm including 0 in the naturals)

$$A_n \supseteq B_{n+1} \;\to\; A_{n+1} \supseteq B_{n+2}.$$

Similarly, as $A_1 = f(A)$, and f is not onto, we have that $B_0 \supseteq A_1$. Taking g of both sides gives that $B_1 \supseteq A_2$, and taking f of both sides of this gives $B_2 \supseteq A_3$. So, continuing in this fashion, we see that for all $n \in \mathbb{N}$,

$$B_n \supseteq A_{n+1} \ \rightarrow \ B_{n+1} \supseteq A_{n+2}.$$

So, really what we have found in a pseudo–inductive way, is that $A_n \supseteq B_{n+1}$ and $B_n \supseteq A_{n+1}$ for all $n \in \mathbb{N}$. So, using this, we get the following two inclusions.

$$A_0 \supseteq B_1 \supseteq A_2 \supseteq B_3 \cdots$$

and

$$B_0 \supseteq A_1 \supseteq B_2 \supseteq A_3 \cdots$$

Lemma 124. *The following is true*

$$\bigcap_{n=1}^{\infty} A_{2n} = \bigcap_{n=0}^{\infty} B_{2n+1}, \quad \bigcap_{n=1}^{\infty} B_{2n} = \bigcap_{n=0}^{\infty} A_{2n+1}.$$

Proof. This is a standard set inclusion argument. Take $x \in \bigcap_{n=1}^{\infty} A_{2n}$. Then $x \in A_{2n}$ for all $n \in \mathbb{N}$ with $n \geq 1$. But, then by the inclusions above, we have that

$$x \in A_{2n} \subseteq B_{2n-1} = B_{2(n-1)+1}.$$

So $x \in B_{2n+1}$ for all $n \in \mathbb{N}$. Thus $x \in \bigcap_{n=0}^{\infty} B_{2n+1}$. Thus, we have

$$\bigcap_{n=1}^{\infty} A_{2n} \subseteq \bigcap_{n=0}^{\infty} B_{2n+1}$$

The other inclusion is shown in a similar way. The second set equality, of $\bigcap_{n=1}^{\infty} B_{2n} = \bigcap_{n=0}^{\infty} A_{2n+1}$ is proved in the exact same way. □

Due to the lemma, we will assign the following names to these objects. From now on we will notate the sets in the following ways,

$$A_\infty = \bigcap_{n=1}^{\infty} A_{2n}, \quad B_\infty = \bigcap_{n=1}^{\infty} B_{2n}.$$

Now, given an element $x \in A$, if $x \notin A_0 \backslash B_1$, then clearly x must be in B_1 as $A_0 = A$. If it so happens that $x \notin A_0 \backslash B_1$ and $x \notin B_1 \backslash A_2$, then x must be in A_2. Now, if x is not contained in any of the sets in the collection

$$\{A_0 \backslash B_1, \ B_1 \backslash A_2, \ A_2 \backslash B_3\}$$

then x is contained in $B_3 = B_1 \cap B_3$. (The fact that $B_3 = B_1 \cap B_3$ is due to how the sets are nested. Now, if x is not contained in any of the sets in the collection

$$\{A_0 \backslash B_1, \ B_1 \backslash A_2, \ \ldots, A_{2k} \backslash B_{2k+1}\}$$

i.e. if

$$x \in \left(\left(\bigcup_{n=0}^{k} A_{2n} \backslash B_{2n+1} \right) \cup \left(\bigcup_{n=0}^{k-1} B_{2n+1} \backslash A_{2n+2} \right) \right)^c$$

then x is contained in $B_{2k+1} = \bigcap_{n=0}^{k} B_{2n+1}$. Similarly, if x is not contained in any of the sets in the collection

$$\{A_0 \backslash B_1,\ B_1 \backslash A_2,\ \ldots, B_{2k+1} \backslash A_{2k+2}\}$$

i.e. if

$$x \in \left(\left(\bigcup_{n=0}^{k} A_{2n} \backslash B_{2n+1} \right) \cup \left(\bigcup_{n=0}^{k} B_{2n+1} \backslash A_{2n+2} \right) \right)^c$$

then x is contained in $A_{2k+2} = \bigcap_{n=1}^{k+1} A_{2n}$. So, putting all of this together, we have that if x is not contained in any of the sets in the collection

$$\{A_0 \backslash B_1,\ B_1 \backslash A_2,\ \ldots\}$$

then

$$x \in \bigcap_{n=0}^{\infty} B_{2n+1} = \bigcap_{n=1}^{\infty} A_{2n} = A_\infty.$$

So, all of this work was to show that

$$\{A_\infty, A_0 \backslash B_1,\ B_1 \backslash A_2, A_2 \backslash B_3, \ldots\}$$

partitions the set A. Similarly, one can see that

$$\{B_\infty, B_0 \backslash A_1, A_1 \backslash B_2, B_2 \backslash A_3, \ldots\}$$

partitions the set B.

Lemma 125. *For a function $F : X \to Y$, and any subset $A \subseteq X$, we have that*

$$F(X) \backslash F(A) \subseteq F(X \backslash A)$$

with set equality if F is 1–1.

Proof. Take $y \in F(X) \backslash F(A)$. As $y \in F(X)$, there exists some $z \in X$ such that $y = F(z)$. We must have that $z \notin A$, or $y \in F(A)$ by definition, thus $z \in X \backslash A$, and so $y \in F(X \backslash A)$. So, $F(X) \backslash F(A) \subseteq F(X \backslash A)$.

Now, assume that F is 1–1, and take $y \in F(X \backslash A)$. Then $y = F(d)$ for some $d \notin A$. If it were the case that $y \in F(A)$, then $y = F(c)$ for some $c \in A$, and as F is 1–1,

$$F(d) = y = F(c) \to c = d,$$

which is a clear contradiction as $c \in A$ and $d \notin A$. So $y \notin F(A)$, thus $F(X \backslash A) \subseteq F(X) \backslash F(A)$. □

By the lemma above we obtain the following.

$$f(A_{2n} \backslash B_{2n+1}) = f(A_{2n}) \backslash f(B_{2n+1}) = A_{2n+1} \backslash B_{2n+2}.$$

As g is 1–1, we have that g^{-1} exists and is defined from $B_1 \to B$, and hence defined on any subset of B_1. As g^{-1} is 1–1, we have that, also by the lemma,

$$g^{-1}(B_{2n+1} \backslash A_{2(n+1)}) = g^{-1}(B_{2n+1}) \backslash g^{-1}(A_{2(n+1)}) = B_{2n} \backslash A_{2n+1},$$

where $g^{-1}(A_{2(n+1)}) = g^{-1}(g(A_{2(n+1)-1})) = A_{2(n+1)-1} = A_{2n+1}$.

Thus f maps $A_{2n}\backslash B_{2n+1}$ 1–1 and onto $A_{2n+1}\backslash B_{2n+2}$ for all $n \in \mathbb{N}$, and g^{-1} maps $B_{2n+1}\backslash A_{2(n+1)}$ 1–1 and onto $B_{2n}\backslash A_{2n+1}$ for all $n \in \mathbb{N}$.

Lastly, given $x \in A_\infty$, we have that $x \in A_{2n}$ for all $n \in \mathbb{N}$ with $n \geq 1$. Thus $f(x) \in f(A_{2n}) = A_{2n+1}$ for all $n \in \mathbb{N}$ with $n \geq 1$. So

$$f(x) \in \bigcap_{n=1}^{\infty} A_{2n+1} = \bigcap_{n=0}^{\infty} A_{2n+1} = B_\infty.$$

(The first set equality is true as A_1 contains all other A_i with odd index.) So $f(A_\infty) \subseteq B_\infty$. Now, if $y \in B_\infty$, then $y \in B_{2n}$ for all $n \in \mathbb{N}$ with $n \geq 1$. But, by our first lemma, B_∞ also equals $\bigcap_{n=0}^{\infty} A_{2n+1}$, thus $y \in A_{2n+1}$ for all $n \in \mathbb{N}$. Each A_{2n+1} is defined by $A_{2n+1} = f(A_{2n})$. Thus, for every $n \in \mathbb{N}$, there is an $x_{2n} \in A_{2n}$ with $f(x_{2n}) = y$. As f is a 1–1 function, we have that

$$x_0 = x_2 = x_4 = \cdots.$$

So, call $x = x_0 = x_2 = x_4 = \cdots$. We have that $x \in A_{2n}$ for all $n \in \mathbb{N}$. Thus,

$$x \in \bigcap_{n=0}^{\infty} A_{2n} = \bigcap_{n=1}^{\infty} A_{2n} = A_\infty.$$

(Once again, the first set equality comes from the nesting of all A_{2n} inside of A_0.) Thus, as $y = f(x)$, we have that $y \in f(A_\infty)$, and therefore $B_\infty \subseteq f(A_\infty)$ as y was taken arbitrarily. So, $f(A_\infty) = B_\infty$. Or, in other words, f maps A_∞ 1–1 and onto B_∞.

So, finally, define the function $h : A \to B$, by

$$h(x) = \begin{cases} f(x) & \text{if } x \in A_{2n}\backslash B_{2n+1},\ n \in \mathbb{N} \\ g^{-1}(x) & \text{if } x \in B_{2n+1}\backslash A_{2(n+1)},\ n \in \mathbb{N} \\ f(x) & \text{if } x \in A_\infty \end{cases}$$

By the work above, and the fact that $\{A_\infty, A_0\backslash B_1,\ B_1\backslash A_2, A_2\backslash B_3, \dots\}$ partitions A, and $\{B_\infty, B_0\backslash A_1, A_1\backslash B_2, B_2\backslash A$ partitions B, we have that h is a bijection from A to B.

□

We now present a second proof of the CSB theorem from Stephen Willard's book titled *General Topology* [W]

We will first proof the following as a Lemma.

Lemma 126. *Let A, B be sets. Suppose that with each subset C of A there is associated a subset C' of A in such a way that $C \subseteq D$ implies that $C' \subseteq D'$. Then $E = E'$ for some $E \subseteq A$.*

Proof. We start by defining a particular collection of subsets of A. Define

$$\mathcal{C} = \{C \in P(A) \mid C \subseteq C'\}.$$

Now, take E to be the following.

$$E = \bigcup_{A \in \mathcal{C}} A$$

So, E is just the union of all sets in the collection $\mathcal{C}$. As E is a union of subsets of A, it is clear that E is itself a subset of A. Now, given an arbitrary $x \in E$, we have that $x \in C$ for some $C \in P(A)$ with $C \subseteq C'$ by the definition of E. So $x \in C'$. By our assumption, we have that

$$C \subseteq E \quad \Longrightarrow \quad C' \subseteq E'.$$

So $x \in C' \subseteq E'$. Thus $E \subseteq E'$. Now, via our assumption again, we have that

$$E \subseteq E' \quad \Longrightarrow \quad E' \subseteq (E')'.$$

Thus $E' \in \{C \in P(A) \mid C \subseteq C'\} = \mathcal{C}$. So,

$$E' \subseteq \bigcup_{A \in \mathcal{C}} A = E.$$

So $E' \subseteq E$, which gives that $E = E'$. □

Now, we prove the theorem. Assume that $f : A \to B$ and $g : B \to A$ are 1–1 functions. We will show there is a bijection $h : A \to B$.

Proof. We begin with a definition. For a subset C of A, define $C' = A\backslash(g(B\backslash f(C)))$. As I have mentioned earlier, a function will map subsets to subsets, and a 1–1 function will map proper subsets to proper subsets. So, given two subsets C, D of A with the property that $C \subseteq D$, we have the following.

$$\begin{array}{rll} C & \subseteq D & \\ f(C) & \subseteq f(D) & \text{As } f \text{ is a function} \\ B\backslash f(D) & \subseteq B\backslash f(C) & \text{Definition of set compliment} \\ g(B\backslash f(D)) & \subseteq g(B\backslash f(C)) & \text{As } g \text{ is a function} \\ A\backslash(g(B\backslash f(C))) & \subseteq A\backslash(g(B\backslash f(D))) & \text{Definition of set compliment} \\ C' & \subseteq D' & \end{array}$$

So, with our definition of a prime of a set, we see that $C \subseteq D$ implies that $C' \subseteq D'$. Then our lemma applies. Thus there exists a subset E of A with the property that $E = E'$.

We now define our function $h : A \to B$ in the following manner.

$$h(x) = \begin{cases} f(x) & \text{if } x \in E \\ g^{-1}(x) & \text{if } x \in A\backslash E \end{cases}$$

We have a few questions we must address. Is h well–defined, is it 1–1, is it onto? Let us first turn to the well-definedness. Clearly, as f was defined on the entirety of A, it will be defined on E, a subset of A. The question is, is g^{-1} defined on $A\backslash E$. Well, g^{-1} is defined from the image of g in A, $g(B)$ to B. Look at the following.

$$\begin{array}{rl} B\backslash f(E) & \subseteq B \\ g(B\backslash f(E)) & \subseteq g(B) \\ A\backslash g(B) & \subseteq A\backslash(g(B\backslash f(E))) = E' = E \\ A\backslash E & \subseteq g(B). \end{array}$$

We see that $A\backslash E$ is contained in $g(B)$, thus g^{-1} is defined on $A\backslash E$. Also, g^{-1} is 1–1 where it is defined, thus g^{-1} is 1–1 on $A\backslash E$. Also, f is 1–1 on E. Thus $h : A \to B$ is 1–1. We only must show that h is onto. As $E = E' = A\backslash(g(B\backslash f(E)))$, we have

$$A\backslash E = g(B\backslash E).$$

Thus $g^{-1}(A\backslash E) = B\backslash f(E)$. Thus, as f maps E to $f(E)$, and g^{-1} maps $A\backslash E$ to $B\backslash f(E)$, we have that $h(A) = B$. Thus h is onto, and the proof is complete. □

The Continuum Hypothesis

We begin with an important result.

Theorem 127. *For any set A, $A \prec P(A)$.*

Proof. To begin, define a map $f : A \to P(A)$ by $f(a) = \{a\}$. For $a, b \in A$, assume that $f(a) = f(b)$. Then

$$\{a\} = f(a) = f(b) = \{b\}.$$

By the definition of set equality, as two sets are equal precisely if they contain the same elements, it must be that $a = b$. Thus f is injective. As there exists an injective map $f : A \to P(A)$, we have that $A \preceq P(A)$.

Now, to show $A \not\approx P(A)$. To do this, one must show that no map $g : A \to P(A)$ is surjective. Thus, let g be a function, $g : A \to P(A)$, and assume that g is surjective. We will deduce a contradiction. For the function g, define the following set,

$$J = \{x \in A \mid x \notin g(x)\}.$$

Thus J is the subset of A containing elements x that are not contained in their image under g, i.e. $x \notin g(x)$. As g is surjective, there exists $b \in A$ such that $g(b) = J$. We now ask a simple question: Is $b \in J$ or is $b \notin J$?

- If $b \in J$, then $b \notin g(b) = J$. This is clearly a contradiction.
- If $b \notin J$, then $b \in g(b) = J$. This is also a clear contradiction.

Thus b is not in J or its complement J^c, thus $b \notin J \cup J^c = A$, and this is a contradiction. So, no map $g : A \to P(A)$ is surjective, thus $A \not\approx P(A)$. □

In the case that A is a finite set the result above is clear as we have proven that $|P(A)| = 2^{|A|}$ in this case. The interesting part of this result comes from the fact that this result is applicable to any set A, even if A is not finite. In particular, by applying the power set iteratively to the set $\mathbb{N}$, we have

$$\mathbb{N} \prec P(\mathbb{N}) \prec P(P(\mathbb{N})) \prec P(P(P(\mathbb{N}))) \prec \cdots.$$

In this way, higher and higher notions of infinite with respect to cardinality can be constructed.

Let us look at the first comparison more deeply. We have $\mathbb{N} \prec P(\mathbb{N})$, and we have shown earlier that $P(\mathbb{N}) \approx \mathbb{R}$. So, a natural question to ask is, "Is there a set with cardinality strictly in between $\mathbb{N}$ and $\mathbb{R}$?" This is a question that Cantor considered in his studies of cardinality. He asserted that

the answer to this question is negative, and his assertion is called the Continuum Hypothesis. Let us state this explicitly.

The Continuum Hypothesis: There is no set S with $\mathbb{N} \prec S \prec \mathbb{R}$, or equivalently, there is no set S with

$$|\mathbb{N}| < |S| < |\mathbb{R}|.$$

Cantor conjectured that the continuum hypothesis is true. Now, of course if it was true, then it would not be a hypothesis. The reason that it is called a hypothesis is as follows. In 1939, Kurt Gödel proved that on the basis of our axioms (the Zermelo–Fraenkel axioms) the continuum hypothesis could not be disproved. And in 1963, Paul Cohen showed that the continuum hypothesis could not be proved from the ZF axioms either. Thus, the continuum hypothesis is a question whose answer is outside of the realm of what is provable under the Zermelo–Fraenkel axioms.

Appendix: Propositions about Fields & Totally Ordered Fields

Similar to section 1.3, much of this material comes from [R].

Proposition 128. *The axioms for addition imply the following:*

a) If $x + y = x + z$ then $y = z$. This is often called the cancellation law.

b) If $x + y = x$ then $y = 0$, i.e. the additive identity is unique.

c). If $x + y = 0$ then $y = -x$, i.e. additive inverses are unique.

d) $-(-x) = x$.

Proof. For part a) we have

$$y = 0 + y = (-x + x) + y = -x + (x + y) = -x + (x + z) = (-x + x) + z = 0 + z = z$$

For part b), if we let $z = 0$ in part a) then we see that $x + y = x$ implies that $y = 0$. For part c), taking $z = -x$ in part a) implies that $y = -x$. Lastly to prove d), using part c) by replacing x with $-x$ and $y = x$ implies that x is the additive inverse of $-x$, i.e. $x = -(-x)$. □

Proposition 129. *The axioms for multiplication imply the following:*

a) If $x \neq 0$ and $xy = xz$, then $y = z$. This is often called the cancellation law.

b) If $x \neq 0$ and $xy = x$, then $y = 1$ i.e. the multiplication identity is unique.

c). If $x \neq 0$ and $xy = 1$, then $y = \frac{1}{x}$ i.e. multiplicative inverses are unique.

d) If $x \neq 0$, then $x = \frac{1}{\frac{1}{x}}$.

Proof. Left as an exercise □

Proposition 130. *The field axioms imply the following for $x, y, z \in X$.*

a) $0 \cdot x = 0$.

b) If $x \neq 0$ and $y \neq 0$ then $xy \neq 0$.

c) $(-x)y = -(xy) = x(-y)$

d) $(-x)(-y) = xy$

Proof. For the proof of a), we have

$$0 \cdot x = (0 + 0) \cdot x = 0 \cdot x + 0 \cdot x$$

and thus from our previous proposition it must be $0 \cdot x = 0$. For part b). if xy did equal zero, then we would have

$$1 = \frac{1}{y} \cdot \frac{1}{x} \cdot x \cdot y = \frac{1}{y} \cdot \frac{1}{x} \cdot 0 = 0$$

and this is a contradiction, so it must be that $xy \neq 0$. For part c), using what we saw in part a) we have

$$(-x)y + xy = (-x + x)y = 0y = 0$$

and by a previous proposition this implies that $(-x)y = -(xy)$. By a similar argument (letting x be y and vice-versa) shows why $x(-y) = -xy$. For part d) using part c) immediately gives

$$(-x)(-y) = -[x(-y)]$$

and using part c) again gives

$$(-x)(-y) = -[x(-y)] = -[-(xy)]$$

and by a previous proposition we have $-(-xy) = xy$ and this proves part d). □

Proposition 131. *Let $a, b, c, d \in X$ with X a totally ordered field.*

i. If $a \neq 0$, then $a^2 > 0$. In particular, $1 > 0$.

ii. If $a > 0$, then $a^{-1} > 0$.

iii. If $a > b$ and $c > 0$, then $ac > bc$. If $a > b$ and $c < 0$, then $ac < bc$.

iv. If $0 < a < b$, then $0 < b^{-1} < a^{-1}$.

v. If $a < b$ and $c < d$, then $a + c < b + d$.

vi. If $0 < a < b$ and $0 < c < d$, then $a \cdot c < b \cdot d$.

vii. If $a > 0$ and $b > 0$ then $a + b > 0$.

Proof. For part i. if $a > 0$, then $a \cdot a > 0$ by X being a totally ordered field. If $a < 0$, then $-a > 0$ and so $a^2 = (-a)(-a) > 0$ by part d) of the previous propostion and X being a totally ordered field. As $1^2 = 1$, we have $1 > 0$.

For part ii) we have $(a^{-1})^2 > 0$ from part i) and the properties of X being a totally ordered field means that as $a > 0$ and $(a^{-1})^2 > 0$ then $a^{-1} = a(a^{-1})^2 > 0$.

For part iii) as $a - b > 0$ and $c > 0$ we have $(a - b)c > 0$ and the distributive law then gives $ac > bc$. The proof of the second statement is very similar.

For part vi) starting with $0 < a < b$, multiplying across by $\frac{1}{b}$ gives $0 < \frac{a}{b} < 1$ and then multiplying across by $\frac{1}{a}$ gives the result. Note that this is using part i. and part iii.

For part v) X being a totally ordered field gives $a + c < b + c$ from $a < b$ and $b + c < b + d$ from $c < d$. The transitivity of the order gives that $a + c < b + d$.

For part vi) from part iii) we have $ac < bc$ as $a < b$ and $bc < bd$ as $c < d$, and by the transitivity of the order we have $ac < bd$.

Part vii) is left as an exercise. □

Appendix: Induction

Introduction & Examples

Sometimes, we would like to infer that a statement is true for all possible natural numbers. Like, for example, showing that $2n$ is even for all $n \in \mathbb{N}$. Well, this example is trivial as being even is precisely having a factor of 2. But, at other times, there are statements that we would like to prove true for all possible values of $n \in \mathbb{N}$ in which it is not obvious as to how to prove them directly. In these cases, we will fall back on a very important tool called induction.

Consider the statement $P(n)$, that is dependent on $n \in \mathbb{N}$, given by

$$P(n) : \sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Now, there is a direct proof of this, but this will not be given at this time as we wish to show the power of induction.

So, what we need is a new idea that can aggregate all these infinite statements and verify their truth all at once. If the statements in question are recursive or iterative, then we can think of the infinite collection of statements as an infinite sequence of dominos in a row. To clarify this concept a little, in regards to the example above the relation between $P(n)$ and $P(n+1)$ is given by the fact that

$$\sum_{k=1}^{n+1} k = \left(\sum_{k=1}^{n} k\right) + (n+1).$$

In other words, from statement $P(n)$ we can reach statement $P(n+1)$ by just adding the next term, $n+1$. Thus, it is believable that the truth of statement $P(n+1)$ is dependent on the truth of statement $P(n)$.

This is what induction will do formally. Given the truth of the first statement, $P(1)$, i.e. making sure the first domino falls, and then given that the truth of $P(n)$ implies the truth of $P(n+1)$, i.e. the nth domino falling causes the $(n+1)$th to fall, it sufficient to imply that all the dominos fall.

Remark 10. *Induction gives us a way to bypass the notion of infinity. There is no way in reality that we could wait for all the dominos to fall or verify that they have all fallen. But Induction gives us sufficient conditions to verify the truth of an infinite collection of statements regardless, i.e. induction let's us assume this domino process already occurred a priori and deduce the result. This assuming an infinite process has already taken place happens often in mathematics, in fact many of you have done this already. (Think limits in a calculus class)*

Mathematical induction can only be applied when very specific conditions are met. Be careful, because general inductive reasoning leads to contradictions quickly. Consider the phrase, "A dog has 4 legs and a dog is an animal, thus all animals have 4 legs."

What follows is the formal definition of the principle of mathematical induction as well as two other important statements.

The Principle of Mathematical Induction (PMI) – For a collection of statements $P(n)$ dependent upon $n \in \mathbb{N}$, if $P(1)$ is true and $P(n)$ being true implies that $P(n+1)$ is true, then $P(m)$ is true for all $m \in \mathbb{N}$. Or equivalently written as statements,

$$(P(1) \wedge (P(n) \implies P(n+1))) \implies (P(m)\ \forall m \in \mathbb{N}).$$

When doing a proof by induction, you

- Check the base case: Show or check that $P(1)$ is true.
- Assume the induction hypothesis, $P(n)$ is true, and show that it implies that $P(n+1)$ is true.

Once you have shown both of these things, you invoke the principal of mathematical induction (PMI), and simply state that the result is shown.

What follows are two more important concepts.

Well Ordering Principle for $\mathbb{N}$ (WOP) - Every nonempty subset of the natural numbers, $S \subseteq \mathbb{N}$, $S \neq \varnothing$, has a least element.

The Principle of Strong Induction (PSI) - Let $P(n)$ be a sequence of statements dependent upon natural numbers n. If $P(1)$ is true and $P(k)$ true for all $1 \leq k \leq n$ implies $P(n+1)$ is true, then $P(m)$ is true for all $m \in \mathbb{N}$. Written in terms of statements

$$[P(1) \wedge ((P(k), \forall\ 1 \leq k \leq n) \implies P(n+1))] \implies (P(m)\ \forall m \in \mathbb{N}).$$

Strong Induction is sometimes also called *complete induction*.

As it will turn out, all three of these concepts are logicially equivalent. As such, strong induction as a name is a bit of misnomer as it is not actually stronger than the principle of mathematical induction, however it is better suited to handle particular problems. The proof of the equivalency of these will be left to the following section.

Example 51. *We are asked to verify*

$$P(n) : \sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

for all $n \in \mathbb{N}$. So, we check the base case, by plugging in $n = 1$ to both sides,

$$P(1) : 1 = \sum_{k=1}^{1} k = \frac{1(1+1)}{2} = 1,$$

and we see that $P(1)$ is true. Now, assume the induction hypothesis, i.e. assume that $P(n)$ is true. Thus we know that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$. We wish to show that

$$P(n+1) : \sum_{k=1}^{n+1} k = \frac{(n+1)(n+2)}{2},$$

is true. So, we begin with the left–hand side (LHS) of statement $P(n+1)$ and deduce the right–hand side (RHS).

$$\begin{aligned}
\sum_{k=1}^{n+1} k &= \left(\sum_{k=1}^{n} k\right) + (n+1) && \textit{Definition of sum} \\
&= \frac{n(n+1)}{2} + (n+1) && \textit{Induction Hypothesis} \\
&= \frac{n(n+1) + 2(n+1)}{2} && \textit{Algebra} \\
&= \frac{(n+1)(n+2)}{2} && \textit{RHS.}
\end{aligned}$$

Thus $P(n+1)$ is true. So, by PMI we have that $P(m)$ is true for all $m \in \mathbb{N}$.

Remark 11. *In the proof above, do not start with statement $P(n+1)$ and subtract $(n+1)$ from both sides and deduce that $P(n)$ is true. This is bad form, as in doing this you are assuming what you a trying to prove, namely assuming $P(n+1)$ is true in the first place. Never assume what you are trying to prove.*

What follows is a few more examples involving induction processes.

Theorem 132. *The size of the power set for a set with n elements is 2^n.*

Proof. Before we begin the proof, let me remark that the actual set we take the power set of does not matter. The theorem we are proving only asks about the size of a set, but states nothing about the character of the elements in that set. So, in that vain, as we only need a set with n objects, I will prove the result for

$$P(\{1, 2, \ldots, n\}),$$

the power set on the set of the first n natural numbers. But we could just as easily prove the result for the power set on the first n letters from some alphabet, the power set of the first n elements in the periodic table, etc. However, natural numbers do have the distinct advantage that they are not as easily exhausted as finite lists such as the alphabet or the periodic table though. We will come back to this concept later.

Not surprisingly, we will prove the result by induction. The statements $P(n)$ will be

$$P(n): \quad |P(\{1, 2, \ldots, n\})| = 2^n.$$

So, we first check the base case, $n = 1$. This follows simply from

$$P(\{1\}) = \{\varnothing, \{1\}\}.$$

and thus $|P(\{1\})| = 2^1$, and so $P(1)$ is true.

Next, assume the induction hypothesis, $|P(\{1, 2, \ldots, n\})| = 2^n$. Now, we look at

$$P(\{1, 2, \ldots, n, n+1\}).$$

If we take any element $A \in P(\{1, 2, \ldots, n, n+1\})$, then one of two things happens. Either $n+1 \in A$ or $n+1 \notin A$.

- If $n+1 \notin A$, then $A \in P(\{1, 2, \ldots, n\})$.
- If $n+1 \in A$, then $A \backslash \{n+1\} \in P(\{1, 2, \ldots, n\})$.

As $P(\{1, 2, \ldots, n\}) \subseteq P(\{1, 2, \ldots, n, n+1\})$, the above shows that $P(\{1, 2, \ldots, n+1\})$ can be written as

$$P(\{1, 2, \ldots, n+1\}) = \{A \mid A \in P(\{1, 2, \ldots, n\})\} \cup \{A \cup \{n+1\} \mid A \in P(\{1, 2, \ldots, n\})\}.$$

and more importantly there is no overlap between these two sets,

$$\{A \mid A \in P(\{1, 2, \ldots, n\})\} \cap \{A \cup \{n+1\} \mid A \in P(\{1, 2, \ldots, n\})\} = \varnothing$$

This last statement can be seen by the following, if

$$B \in \{A \mid A \in P(\{1, 2, \ldots, n\})\} \cap \{A \cup \{n+1\} \mid A \in P(\{1, 2, \ldots, n\})\}$$

then $B \in P(\{1, 2, \ldots, n\})$ and $n + 1 \in B$, but this is nonsense, so clearly the intersection is empty.

As $P(\{1, 2, \ldots, n + 1\})$ is formed by $\{A \mid A \in P(\{1, 2, \ldots, n\})\}$ and $\{A \cup \{n + 1\} \mid A \in P(\{1, 2, \ldots, n\})\}$ we can count the number of elements in $P(\{1, 2, \ldots, n + 1\})$ by counting the number of elements in these two sets. Now, this is important, because the intersection of these two sets is empty, we will not 'double–count' any element. So

$$|P(\{1, 2, \ldots, n + 1\})| = |\{A \mid A \in P(\{1, 2, \ldots, n\})\}| + |\{A \cup \{n + 1\} \mid A \in P(\{1, 2, \ldots, n\})\}|.$$

Now, by the induction hypothesis,

$$|\{A \mid A \in P(\{1, 2, \ldots, n\}\}| = |P(\{1, 2, \ldots, n\})| = 2^n$$
$$|\{A \cup \{n + 1\} \mid A \in P(\{1, 2, \ldots, n\})\}| = |P(\{1, 2, \ldots, n\})| = 2^n$$

Thus,

$$|P(\{1, 2, \ldots, n + 1\})| = 2^n + 2^n = 2(2^n) = 2^{n+1}.$$

Thus, $P(n+1)$ is true. So, now by the principle of mathematical induction, the result holds for all $m \in \mathbb{N}$. □

Example 52. *The sum of the first n odd numbers equals n^2.*

Let us first write this as a collection of statements over the naturals.

$$P(n) : \sum_{k=1}^{n} (2k - 1) = n^2$$

We will perform a proof by induction.

Proof. We first check the base case

$$1 = \sum_{k=1}^{1} (2k - 1) = 1^2 = 1$$

and we see that it is true.

We now assume that $P(n)$ is a true statement, i.e.

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

and deduce $P(n + 1)$. To do this we will start with the left hand side (LHS) of the statement $P(n + 1)$, and get to the right hand side (RHS).

$$\begin{aligned} \sum_{k=1}^{n+1} (2k - 1) &= \sum_{k=1}^{n} (2k - 1) + [2(n + 1) - 1] \\ &= n^2 + 2n + 1 = (n + 1)^2 \end{aligned}$$

And so we see that $P(n)$ being true implies that $P(n + 1)$ is true.

And so, now that we have shown this, by the principle of mathematical induction we have that $P(m)$ is true for all $m \in \mathbb{N}$. □

Example 53. *Let us see why* $5|(n^5 - n)$ *for all* $n \in \mathbb{N}$.

We see that we have a collection of statements

$$P(n) : 5|(n^5 - n)$$

Proof. We proceed by induction. We first check the base case $n = 0$.

$$5|(0^5 - 0)$$

and we see this is true as any nonzero number divides 0.

Let us now see how $P(n)$ implies $P(n+1)$ via a direct proof. Let us assume $P(n)$ is true, i.e. $5|(n^5 - n)$ thus there is a $p \in \mathbb{Z}$ such that $n^5 - n = 5p$, and see if we can deduce $P(n+1)$, thus let us look at

$$\begin{aligned}(n+1)^5 - (n+1) &= n^5 + 5n^4 + 10n^3 + 10n^2 + 5n + 1 - n - 1 \\ &= n^5 - n + 5n^4 + 10n^3 + 10n^2 + 5n \\ &= 5p + 5n^4 + 10n^3 + 10n^2 + 5n \\ &= 5(p + n^4 + 2n^3 + 2n^2 + n)\end{aligned}$$

As $p + n^4 + 2n^3 + 2n^2 + n \in \mathbb{Z}$ we have $5|[(n+1)^5 - (n+1)]$, thus $P(n+1)$ is true.

Thus by the principle of mathematical induction, we have that $P(m)$ is true for all $m \in \mathbb{N}$. □

Example 54. *Show that* $4|5^n - 1$ *for all* $n \in \mathbb{N}$.

To maybe give a hint of the equivalency between the well ordering principle and the principle of mathematical induction, we will prove this by the well-ordering principle.

Proof. Let us define the folllowing collection of statements as

$$P(n) : 4|5^n - 1$$

and define the set $S = \{n \mid P(n) \text{ is true}\}$. If $S = \mathbb{N}$ then there is nothing to prove, our claim is true. If $S \neq \mathbb{N}$, then $\mathbb{N}\backslash S \neq \varnothing$, and thus by the well ordering principle, the set $\mathbb{N}\backslash S$ has a least element, call it l.

For $n = 1$, we have $4|(5 - 1)$, so $1 \in S$ and thus $l \neq 1$ as $l \in \mathbb{N}\backslash S$. Now since, $l > 1$ we have that $l - 1 \in \mathbb{N}$. By the definition of l being the least element of $\mathbb{N}\backslash S$, it must be that $l - 1 \in S$. So $P(l-1)$ is true, thus $4|5^{l-1} - 1$. And so there is a $p \in \mathbb{Z}$ such that $5^{l-1} - 1 = 4p$. But this then implies

$$\begin{aligned}5^l - 1 &= 5^l - 5 + 5 - 1 \\ &= 5(5^{l-1} - 1) + 4 \\ &= 5(4p) + 4 \\ &= 4(5p + 1)\end{aligned}$$

So this shows that $4|(5^l - 1)$, so $l \in S$. And so we have a clear contradiction as

$$l \in S \cap (\mathbb{N}\backslash S) = \varnothing$$

So it must be the case that $\mathbb{N}\backslash S = \varnothing$, thus $S = \mathbb{N}$ and we have proven our result. □

Theorem 133. (The Binomial Theorem) *For real numbers x and y with n a natural number, we have*

$$(x+y)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^{n-k} y^k.$$

where $\left(\begin{array}{c} n \\ k \end{array} \right)$ *is the binomial coefficient, often referred too as 'n' choose 'k', written*

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!(n-k)!}.$$

Proof. We aim to prove the result with induction, thus we will first show the base case, $n = 1$.

$$x + y = (x+y)^1 = \sum_{k=0}^{1} \left(\begin{array}{c} 1 \\ k \end{array} \right) x^{1-k} y^k = \left(\begin{array}{c} 1 \\ 0 \end{array} \right) x + \left(\begin{array}{c} 1 \\ 1 \end{array} \right) y = x + y$$

as n choose 0 and n choose 1 are identical (both are 1. Also keep in mind that $0! = 1$ is commonplace. The rationale behind this involves the Bohr-Mollerup theorem if you want to set 0! on some solid foundation.)

We now proceed to the general proof. We will assume as our inductive hypothesis that

$$(x+y)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^{n-k} y^k$$

and try to deduce the result for $n+1$. We will begin with the left hand side

$$\begin{aligned}
(x+y)^{n+1} = (x+y)(x+y)^n &= (x+y) \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^{n-k} y^k \\
&= \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^{n+1-k} y^k + \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^{n-k} y^{k+1}. \\
&\quad \text{by the distributive property} \\
&= \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^{n+1-k} y^k + \sum_{m=1}^{n+1} \left(\begin{array}{c} n \\ m-1 \end{array} \right) x^{n+1-m} y^m. \\
&\quad \text{by re-indexing the second sum with } m = k+1.
\end{aligned}$$

At this point we will do something a little strange. We are really about to exploit the outer edges of Pascal's triangle, or put more bluntly, the fact that n choose 0 is equivalent to m choose 0 for any n and m. (Similar for n choose n and m choose m.) We will peel off the first term from the first sum and the last from the second to get

$$\left(\begin{array}{c} n \\ 0 \end{array} \right) x^{n+1} + \sum_{k=1}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^{n+1-k} y^k + \sum_{m=1}^{n} \left(\begin{array}{c} n \\ m-1 \end{array} \right) x^{n+1-m} y^m + \left(\begin{array}{c} n \\ n \end{array} \right) y^{n+1}.$$

Now, the point of this was that the sums in the middle now have indices that end and begin in the same place (both start at 1 and go to n.) Also, k and m are just counters, there is nothing special

about them, there is no reason we can not just use one of the letters to do the counting in both sums. When we do this we get

$$(x+y)^{n+1} = x^{n+1} + \sum_{k=1}^{n} \left[\begin{pmatrix} n \\ k \end{pmatrix} + \begin{pmatrix} n \\ k-1 \end{pmatrix} \right] x^{n+1-k} y^k + y^{n+1}.$$

We now come to a lemma. A lemma that says nothing more than the property of Pascal's triangle that is taught in grade school. That to find the terms in the next row of the triangle we only must sum terms from the prior row, i.e.

Lemma 134.

$$\begin{pmatrix} n+1 \\ k \end{pmatrix} = \begin{pmatrix} n \\ k \end{pmatrix} + \begin{pmatrix} n \\ k-1 \end{pmatrix}.$$

Proof. We will prove this lemma quickly from the definition. Start with the right hand side

$$\begin{aligned} \begin{pmatrix} n \\ k \end{pmatrix} + \begin{pmatrix} n \\ k-1 \end{pmatrix} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\ &= \frac{n!(n-k+1)+n!k}{k!(n+1-k)!} = \frac{n!(n-k+1+k)}{k!(n+1-k)!} \\ &= \frac{(n+1)!}{k!(n+1-k)!} = \begin{pmatrix} n+1 \\ k \end{pmatrix} \end{aligned}$$

□

Back to our proof. The coefficient in our sum can be combined now by the lemma.

$$(x+y)^{n+1} = x^{n+1} + \sum_{k=1}^{n} \begin{pmatrix} n+1 \\ k \end{pmatrix} x^{n+1-k} y^k + y^{n+1}.$$

And like we mentioned, exploiting Pascal's triangle

$$(x+y)^{n+1} = \begin{pmatrix} n+1 \\ 0 \end{pmatrix} x^{n+1} + \sum_{k=1}^{n} \begin{pmatrix} n+1 \\ k \end{pmatrix} x^{n+1-k} y^k + \begin{pmatrix} n+1 \\ n+1 \end{pmatrix} y^{n+1}.$$

The term on the left and far right are precisely the '0'th and '$n+1$'th term of the sum, so fold them back in to get

$$(x+y)^{n+1} = \sum_{k=0}^{n+1} \begin{pmatrix} n+1 \\ k \end{pmatrix} x^{n+1-k} y^k.$$

And this is right hand side of our claim, thus the proof is complete. □

Before the next set of examples, let us see a quick definition. If we look at the following quadratic equation,

$$x^2 - x - 1 = 0.$$

one can see the solutions of which are

$$\alpha = \frac{1+\sqrt{5}}{2}, \qquad \beta = \frac{1-\sqrt{5}}{2}$$

This will come up in the next example. The number α is what most people fondly call the 'golden ratio.'

Definition 66. *The Fibonacci sequence is a sequence that is recursively defined as follows. The first two terms of the sequence are $F_1 = 1$ and $F_2 = 1$, and the general nth term is defined by*

$$F_n = F_{n-1} + F_{n-2}.$$

So, the Fibonacci sequence is a perfect example of a case in which the nth step is dependent on more than just the immediately prior step, it is dependent on the prior two steps. As such, for many inductive proofs involving the Fibonacci sequence it is much easier to use strong induction.

Example 55. *The nth term in the Fibonacci sequence is given by*

$$F_n = \frac{\alpha^n - \beta^n}{\sqrt{5}}$$

Proof. Because of the recursive nature of the Fibonacci sequence mentioned above, this collection of statements will be easier to prove by strong induction.

The first term in the Fibonacci sequence that is defined by the recursive process is $F_3 = F_2 + F_1$. Thus, when we check our base case, we will check that both F_1 and F_2 obey the formula in the claim.

$$F_1 = \frac{\alpha - \beta}{\sqrt{5}} = \frac{1}{\sqrt{5}}\left[\frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2}\right] = 1.$$

As α and β are solutions to $x^2 - x - 1 = 0$, we have that $\alpha^2 = \alpha + 1$, and $\beta^2 = \beta + 1$. Thus,

$$F_2 = \frac{\alpha^2 - \beta^2}{\sqrt{5}} = \frac{\alpha + 1 - (\beta + 1)}{\sqrt{5}} = F_1 = 1.$$

Thus we see that the base case $n = 1$ holds as well as case $n = 2$ to get us to our first recursive step.

Now we assume that $F_k = \frac{\alpha^k - \beta^k}{\sqrt{5}}$ for all $1 \leq k \leq n$ and show that the claim holds for $n + 1$.

$$\begin{aligned}
F_{n+1} &= F_n + F_{n-1} \\
F_{n+1} &= \frac{\alpha^n - \beta^n}{\sqrt{5}} + \frac{\alpha^{n-1} - \beta^{n-1}}{\sqrt{5}} && \text{Induction Hypothesis} \\
F_{n+1} &= \frac{\alpha^{n-1}(\alpha + 1) - \beta^{n-1}(\beta + 1)}{\sqrt{5}} \\
F_{n+1} &= \frac{\alpha^{n-1}(\alpha^2) - \beta^{n-1}(\beta^2)}{\sqrt{5}} && \text{As } \alpha^2 = \alpha + 1, \text{ same for } \beta \\
F_{n+1} &= \frac{\alpha^{n+1} - \beta^{n+1}}{\sqrt{5}}
\end{aligned}$$

Thus, by strong induction, the result holds for all $m \in \mathbb{N}$. □

Example 56. *Show that for the Fibonacci sequence we have*

$$F_{n+1}^2 - F_{n+1}F_n - F_n^2 = (-1)^n$$

for all natural numbers n.

Proof. We will prove this result by induction (not strong induction). So, let us check the base case

$$\begin{aligned} F_2^2 - F_2F_1 - F_1^2 &= (-1)^1 \\ 1^2 - 1(1) - 1^2 &= -1 \\ -1 &= -1 \end{aligned}$$

We will now, as our induction step, assume that

$$F_{n+1}^2 - F_{n+1}F_n - F_n^2 = (-1)^n$$

and deduce the following

$$F_{n+2}^2 - F_{n+2}F_{n+1} - F_{n+1}^2 = (-1)^{n+1}$$

So we start with the left hand side

$$\begin{aligned} F_{n+2}^2 - F_{n+2}F_{n+1} - F_{n+1}^2 &= F_{n+2}^2 - 2F_{n+2}F_{n+1} + F_{n+1}^2 + F_{n+2}F_{n+1} - 2F_{n+1}^2 \\ &= (F_{n+2} - F_{n+1})^2 + F_{n+1}(F_{n+2} - 2F_{n+1}) \\ &= F_n^2 + F_{n+1}(F_n + F_{n+1} - 2F_{n+1}) \\ &= F_n^2 + F_{n+1}F_n - F_{n+1}^2 \\ &= -(F_{n+1}^2 - F_{n+1}F_n - F_n^2) \\ &= -(-1)^n = (-1)^{n+1} \end{aligned}$$

thus we see that the inductive step holds. Thus, by the principle of mathematical induction we have that the claim holds for all $n \in \mathbb{N}$. □

Example 57. *Let us see why* $12|n^4 - n^2$ *for all* $n \in \mathbb{N}$.

Proof. We will give a proof by induction, so we first check the base case when $n = 1$. As $1^4 - 1^2 = 0$ and any integer divides 0 we have that $12|1^4 - 1^2$ hence the result holds.

Now let us assume the result holds for n, i.e. $12|n^4 - n^2$, so there exists a $p \in \mathbb{N}$ such that $n^4 - n^2 = 12p$ and let us try and deduce the result for $n + 1$. In other words, let us show that $12|(n + 1)^4 - (n + 1)^2$. So let us look at the expression

$$\begin{aligned} (n+1)^4 - (n+1)^2 &= n^4 + 4n^3 + 6n^2 + 4n + 1 - (n^2 + 2n + 1) \\ &= n^4 - n^2 + 4n^3 + 6n^2 + 2n \\ &= 12p + 2n(2n^2 + 3n + 1) \\ &= 12p + 2n(n + 1)(2n + 1) \end{aligned}$$

Now the result will be complete if we can show why $n(n+1)(2n+1)$ is divisible by 6. This can be done in many ways, but I'll give you two as example

- The expression $n(n + 1)(2n + 1)$ is definitely even as $n(n + 1)$ will always be even regardless of n being even or odd, so we only need to see why a factor of 3 is present. And this comes down to three cases.
  - If n has a remainder of 0 when divided by 3, then n contains a factor of 3.

  - If n has a remainder of 1 when divided by 3, then $2n+1$ contains a factor of 3. In this case $n = 3m+1$ and

$$2n+1 = 2(3m+1)+1 = 6m+2+1 = 6m+3 = 3(2m+1)$$

  - If n has a remainder of 2 when divided by 3, then $n+1$ has a factor of 3.

  Thus in any case $n(n+1)(2n+1)$ has a factor of 2 and 3, thus is divisible by 6.

- This result can also follow from

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

  and this immediately tells you that $n(n+1)(2n+1)$ is divisible by 6 as the left hand side is a sum of integers.

Making use of the second bullet point we have

$$(n+1)^4 - (n+1)^2 = 12p + 2\left(6\sum_{k=1}^{n} k^2\right) = 12\left(p + \sum_{k=1}^{n} k^2\right)$$

and this shows that $12|(n+1)^4 - (n+1)^2$. Thus the inductive step is proven and the result now holds by the principle of mathematical induction. □

Example 58. *For a nonzero real number x, if $x + \frac{1}{x} \in \mathbb{Z}$, then $x^m + \frac{1}{x^m} \in \mathbb{Z}$ for all natural numbers m.*

We will prove this result via induction, but notice that the base is automatically true as it is an assumption in the claim. Before we complete the proof, let us look at the following

$$\left(x^m + \frac{1}{x^m}\right)\left(x + \frac{1}{x}\right) = x^{m+1} + x^{m-1} + \frac{1}{x^{m-1}} + \frac{1}{x^{m+1}}$$

So from this we see

$$x^{m+1} + \frac{1}{x^{m+1}} = \left(x^m + \frac{1}{x^m}\right)\left(x + \frac{1}{x}\right) - \left(x^{m-1} + \frac{1}{x^{m-1}}\right)$$

and this tells us that this argument will be easiest by strong induction, as the result for $m+1$ will follow from the result at step m, step $m-1$, and step 1.

Proof. The result holds when $m = 1$ trivially. Using atrong induction we assume that $x^k + \frac{1}{x^k} \in \mathbb{Z}$ for all $1 \leq k \leq m$, but then

$$x^{m+1} + \frac{1}{x^{m+1}} = \left(x^m + \frac{1}{x^m}\right)\left(x + \frac{1}{x}\right) - \left(x^{m-1} + \frac{1}{x^{m-1}}\right)$$

shows why $x^{m+1} + \frac{1}{x^{m+1}} \in \mathbb{Z}$ by our induction assumption, thus the inductive step holds, and by strong induction we can conclude that $x^n + \frac{1}{x^n} \in \mathbb{Z}$ for all $n \in \mathbb{N}$ if $x + \frac{1}{x} \in \mathbb{N}$. □

Theorem 135. (The Fundamental Theorem of Arithmetic) - *Every natural number (except $n = 1$) has a unique prime factorization.*

Proof. (Existence): Show that every $n \in \mathbb{N}$, $n \neq 1$, has a prime factorization.

We proceed by way of contradiction. Assume that not every $n \in \mathbb{N}$, $n \neq 1$, has a prime factorization. Define the following set

$$S = \{n \in \mathbb{N} \mid n \text{ has no prime factorization}\}.$$

By our assumption above, $S \neq \varnothing$. Thus, by the WOP, S has a least element l. Well, l is either prime or l is composite.

- If l was prime, then l has a prime factorization, namely $l = l \cdot 1$. Thus, we can conclude that l is not prime.
- As l is composite, there exist $p, q \in \mathbb{N}$, with $p, q \neq 1$ and $pq = l$. Clearly, $p < l$ and $q < l$, as $p, q \neq 1$ and $p, q \in \mathbb{N}$.

As l is the least element of S, we have that $p, q \in S^c$. Thus, p, q both have prime factorizations.

$$p = n_1^{e_1} n_2^{e_2} \cdots n_s^{e_s}$$
$$q = m_1^{f_1} m_2^{f_2} \cdots m_r^{f_r}$$

for $n_1, \ldots, n_s, m_1, \ldots m_r$ prime numbers and $e_1, \ldots, e_s, f_1, \ldots, f_r \in \mathbb{N}$. As $l = pq$, we have that

$$l = n_1^{e_1} n_2^{e_2} \cdots n_s^{e_s} m_1^{f_1} m_2^{f_2} \cdots m_r^{f_r}.$$

Thus, l has a prime factorization, so $l \in S^c$. Then $l \in S \cap S^c = \varnothing$, which is a contradiction.

Now that we know that every natural number n has a prime factorization, let us see why it is unique.

(Uniqueness): We will prove this portion by strong induction. The first natural number that is not 1 is 2, and 2 has a unique prime factorization as it is prime, so this proves the base case.

Now we will assume the inductive hypothesis, which by strong induction is that claim that for all $1 \leq k \leq n$, that k has a unique prime factorization. We will try to deduce that $n + 1$ has a unique prime factorization.

So let us assume that $n + 1$ has two prime factorizations, i.e. there is are primes $q_1, q_2, \ldots q_u$ and $p_1, p_2, \ldots, p_m$ and natural numbers $r_1, r_2, \ldots, r_u$ and $s_1, s_2, \ldots, s_m$ with

$$p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m} = n + 1 = q_1^{s_1} q_2^{s_2} \cdots q_u^{s_u}$$

Now as p_1 divides the left hand side, it must divide the right hand side. This means that $p_1 = q_k$ for some k in $1 \leq k \leq n$ as these are prime numbers. Without loss of generality, assume that $p_1 = q_1$. By dividing both sides by p_1 we have

$$p_1^{r_1 - 1} p_2^{r_2} \cdots p_m^{r_m} = p_1^{s_1 - 1} q_2^{s_2} \cdots q_u^{s_u}$$

In other words if we define $l = p_1^{r_1 - 1} p_2^{r_2} \cdots p_m^{r_m} = p_1^{s_1 - 1} q_2^{s_2} \cdots q_u^{s_u}$ then we have

$$n + 1 = p_1 l$$

thus $l < n + 1$. So by our induction hypothesis, l has a unique prime factorization, so it must be the case that $m = u$, that $p_1, p_2, \ldots, p_m$ and $q_1, q_2, \ldots, q_u$ are the same list of primes, so up to relabeling we have $p_k = q_k$ for $1 \leq k \leq m$, and that $r_k = s_k$ for $1 \leq k \leq m$. So, using this unique prime factorization of l we have

$$n + 1 = p_1 l = p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m}$$

is the unique factorization of $n + 1$. Thus the induction step holds, so by the principle of strong induction, every natural number has a unique prime factorization. $\square$

WOP $\iff$ PMI $\iff$ PSI

We will prove this in the following way, we will prove WOP $\iff$ PMI, and WOP $\iff$ PSI.

First Part: WOP $\iff$ PMI.

Proof. $\Rightarrow$ We first prove that WOP $\Rightarrow$ PMI. The proof given will be by contradiction. Thus we assume WOP and $\neg$PMI and deduce a contradiction. Let us first find what $\neg$ PMI is. Using the statement formulation of PMI.

$$\begin{gathered}\neg[(P(1) \wedge (P(n) \implies P(n+1))) \implies (P(m)\ \forall m \in \mathbb{N}).] \iff \\ \neg[\neg(P(1) \wedge (P(n) \implies P(n+1))) \vee (P(m)\ \forall m \in \mathbb{N})] \iff \\ P(1) \wedge (P(n) \implies P(n+1)) \wedge \neg(P(m)\ \forall m \in \mathbb{N})\end{gathered}$$

Thus when we assume WOP and $\neg$ PMI we have

- The well–ordering principle on $\mathbb{N}$
- $P(1)$ is true.
- $P(n)$ true implies that $P(n+1)$ is true.
- There exists some $m \in \mathbb{N}$ such that $P(m)$ is false.

On with the proof. Let us start by defining the following set,

$$S = \{n \mid P(n) \text{ is true}\}.$$

Now, by assumption, $S \neq \mathbb{N}$. Thus $S^c \neq \varnothing$. So, by the well–ordering principle, as $S^c \neq \varnothing$, we have that S^c has a least element in the naturals, call it l. By assumption, as $P(1)$ is true, we have that $l \neq 1$. Now, since $l \in \mathbb{N}$ and $l \neq 1$, we have that $l - 1 \in \mathbb{N}$.

Now, remember, l is the least element of S^c. Thus, $l - 1 \notin S^c$, so $l - 1 \in S$. Thus $P(l-1)$ is true. Then, by assumption, as $P(l-1)$ is true, $P(l-1+1) = P(l)$ is true. Thus, $l \in S$. So, $l \in S$ and $l \in S^c$. This is a contradiction.

$\Leftarrow$ Now show PMI $\Rightarrow$ WOP. We will proceed with a direct proof. But, first a definition.

Define the statements, $P(n)$, for every $n \in \mathbb{N}$,

$$P(n) : \text{A nonempty subset } S \subseteq \mathbb{N} \text{ of size } |S| = n \text{ has a least element.}$$

The base case $n = 1$. A set S with size $|S| = 1$ has only one element. Thus, S has the form $S = \{m\}$ for some $m \in \mathbb{N}$. Clearly, S has a least element, namely m itself as there are no other elements in S to compare m too.

Now, assume the induction hypothesis, $P(n)$, i.e. we will assume any subset $T \subseteq \mathbb{N}$ of size n, $|T| = n$, has a least element. And let us show why a set with $n+1$ elements has a least element. Thus, let S be an arbitrary subset of $\mathbb{N}$ with $|S| = n + 1$. Thus, we can list the elements of S,

$$S = \{r_1, r_2, \ldots r_{n+1}\}.$$

What we are going to do is pick an element from S. Without loss of generality, say we pick r_1. Now we look at $S\backslash\{r_1\}$. As we have taken one element out of S, this set has size n, i.e. $|S\backslash\{r_1\}| = n$. Thus, by the induction hypothesis, as $S\backslash\{r_1\}$ has size n, we have that $S\backslash\{r_1\}$ has a least element, call it l. Now, one of two things happens.

1). $r_1 \leq l$, if this is the case, then as l is the least element of $S\backslash\{r_1\}$, then for any element $z \in S\backslash\{r_1\}$

$$r_1 \leq l \leq z,$$

and thus r_1 is the least element of S.

2). $l \leq r_1$, if this is the case, then as l is the least element of $S\backslash\{r_1\}$ and $l \leq r_1$, we have that l is the least element of S.

Thus, in either case, S has a least element. And as S was an arbitrary subset of $\mathbb{N}$ of size $n+1$, we have that $P(n+1)$ is true. Thus, by PMI we have that $P(m)$ is true for all $m \in \mathbb{N}$, in other words any set S of size $|S| = m$ for m any natural number, we have that S has a least element.

So, we're done right? **Well, no**. What about the even numbers?

$$E = \{2, 4, 6, \dots\}.$$

The size of the even numbers is clearly not finite, i.e. the size of the even numbers is not equal to m for any $m \in \mathbb{N}$, so what we proved above does not apply to the set of the even numbers.

What we proved above is not the well–ordering principle of $\mathbb{N}$, but a weaker version of it. We showed that any set with a finite number of elements has a least element. So, what happened? Did Induction fail? Is the PMI not strong enough to prove WOP?

The answer: Induction did not fail, and PMI is strong enough to prove WOP, but I gave the proof above to make an example. Induction is a powerful tool, but it can be limited due to the choice of statements $P(n)$ it is being applied to. The reason the proof above only gave a weak version of WOP was because the statements $P(n)$ were too weak to encapsulate information about infinite subsets of $\mathbb{N}$. So, to summarize, *induction can be limited by the choice of statements it is applied to.*

So, we try again, we will prove PMI $\implies$ WOP by contradiction. So, we assume the principle of mathematical induction and the negation of the well ordering principle. Thus we will assume the existence of a set S, that is a non-empty subset of the naturals and that S has no least element.

We then define the following collection of statements dependent on the natural numbers,

$$P(n) : \{1, 2, \dots, n\} \cap S = \varnothing$$

Let us check $P(1)$. It must be the case that $1 \notin S$ otherwise 1 would be the least element of S as 1 is the least element of $\mathbb{N}$. So it must be the case that

$$\{1\} \cap S = \varnothing$$

So, $P(1)$ is true.

Now, let us assume that $P(n)$ is true, i.e.

$$\{1, 2, \dots, n\} \cap S = \varnothing$$

From this we must have that $n+1 \notin S$, otherwise $n+1$ would be the least element of S since S contains none of the elements $1, 2, \dots, n$, so

$$\{1, 2, \dots, n+1\} \cap S = \varnothing$$

and we see that $P(n+1)$ is true. Thus by the principle of mathematical induction, we have $P(m)$ is true for all $m \in \mathbb{N}$, thus we have

$$\{1, 2, \ldots, m\} \cap S = \varnothing, \qquad \forall m \in \mathbb{N}$$

Thus we have

$$\varnothing \neq S = \mathbb{N} \cap S = \left[\bigcup_{n=1}^{\infty} \{1, 2, \ldots, n\}\right] \cap S = \bigcup_{n=1}^{\infty} [\{1, 2, \ldots, n\} \cap S] = \varnothing$$

and this is a contradiction. □

Let us now see why WOP $\implies$ PSI. We will again prove this by contradiction, thus we will assume for the statements P and the set S defined by

$$S = \{n \mid P(n) \text{ is true}\}$$

- The well ordering principle for $\mathbb{N}$.
- That $P(1)$ is true.
- That $\bigwedge_{k=1}^{n} P(k)$ true implies that $P(n+1)$ is true.
- That $P(m)$ is not true for all $m \in \mathbb{N}$.

Proof. By our last statement $S \neq \mathbb{N}$ and thus $\mathbb{N} \backslash S \neq \varnothing$. Thus by the well ordering principle $\mathbb{N} \backslash S$ has a least element l. By the second bullet point we have that $l \neq 1$, so $l > 1$. This means that $l - 1 \in \mathbb{N}$ and by definition of l being the least element of $\mathbb{N} \backslash S$, we have

$$\{1, 2, \ldots, l-1\} \cap S = \{1, 2, \ldots l - 1\}$$

Thus $P(k)$ is true for all $1 \leq k \leq l-1$. So by the third bullet point, we have that $P(l)$ is true, thus $l \in S$, but then

$$l \in S \cap [\mathbb{N} \backslash S] = \varnothing$$

and this is a contradiction. Thus we have WOP $\implies$ PSI. □

We lastly need to prove that PSI $\implies$ WOP. This effectively the same proof as PMI $\implies$ WOP except we define the collection of statements dependent on the natural numbers by,

$$P(n) : \{n\} \cap S = \varnothing$$

Thus we see the connection between our previous proof and the assumption in strong induction by

$$\left[\bigwedge_{k=1}^{n} P(k)\right] \iff \{1, 2, \ldots, n\} \cap S = \varnothing.$$

So the 'strong' part of the strong induction just helps us recover the previous statements we had in the PMI $\implies$ WOP proof.

Remark 12. *In fact this is a general trick, given statements dependent on the natural numbers, i.e. $P(k)$, if we define a second collection of statements dependent on the natural numbers by*

$$Q(n) = \bigwedge_{k=1}^{n} P(k)$$

then strong induction with $P(k)$ is clearly equivalent to induction with $Q(k)$.

Appendix: Equivalence Relations

Introduction & Examples

There are times in which equality in the sense we have known our whole life is far too restrictive. In particular, we would like to collect objects and consider them equal if they share certain properties even if the two objects considered are not the exact same object. You have likely done this your whole life without giving a name to this process. Think for example of the word 'dogs'. When used, it describes all animals that we would call a dog even though no two dogs are exactly the same physical being. And I can refer to the collection of all dogs by just saying the word 'dogs' without any loss in communication. In this section we will formalize this notion.

Definition 67. *An* equivalence relation *on a set S is a relation on S that is reflexive, symmetric, and transitive.*

- **reflexive** - *A relation $\sim$ is reflexive on S if $x \sim x$ for all $x \in S$.*
- **symmetric** - *A relation $\sim$ is symmetric on S if $x \sim y$ implies $y \sim x$ for all $x, y \in S$.*
- **transitive** - *A relation $\sim$ is transitive on S if $x \sim y$ and $y \sim z$ implies that $x \sim z$ for all $x, y, z \in S$.*

Remark 13. *As mentioned above, an equivalence relation will generalize the notion of equality, or two elements 'being equal', and any good notion of equality should precisely be reflexive, symmetric, and transitive.*

Example 59. *Let $S = \{$people living in the United States$\}$. We define the relation $\sim$ on S by saying that for $x, y \in S$, $x \sim y$ if x lives in the same state as y. We claim that $\sim$ is an equivalence relation on S. Thus we check*

i). Take $x \in S$, then as x clearly lives in the same state as itself, $x \sim x$. Thus $\sim$ is reflexive.

ii). Take $x, y \in S$ and assume that $x \sim y$. Thus, x lives in the same state as y. So, clearly, y live in the same state as x. So $y \sim x$, and thus $\sim$ is symmetric.

iii). Take $x, y, z \in S$ and assume that $x \sim y$ and $y \sim z$. So, x lives in the same state as y and y lives in the same state as z. Thus x lives in the same state as z, so $x \sim z$. Thus, $\sim$ is transitive.

So, $\sim$ is an equivalence relation on S.

Example 60. Modular classes of integers (mod p) *The set we will work with in this example is the integers, $\mathbb{Z}$. Let $p \in \mathbb{Z}$, we define $\sim_p$ on $\mathbb{Z}$ by the following: we say $x \sim_p y$ if $p \mid x - y$, i.e. if there exists $m \in \mathbb{Z}$ such that $x - y = mp$. We claim that $\sim_p$ is an equivalence relation.*

i). Given $x \in \mathbb{Z}$, we have that $x - x = 0$, i.e. $x - x = 0p$. Thus $p \mid x - x$, and so $x \sim_p x$. Thus $\sim_p$ is reflexive.

ii). Given $x, y \in \mathbb{Z}$ and assume that $x \sim_p y$. Thus $p \mid x - y$, and so there exists $m \in \mathbb{Z}$ such that $x - y = mp$. Well, then $y - x = (-m)p$ and $-m \in \mathbb{Z}$. So, $p \mid y - x$ and thus $y \sim_p x$. So $\sim_p$ is symmetric.

iii). Given $x, y, z \in \mathbb{Z}$ and assume that $x \sim_p y$ and $y \sim_p z$. Thus, $p \mid x - y$ and $p \mid y - z$. In other words, there exists $m, n \in \mathbb{Z}$ such that $x - y = mp$ and $y - z = np$. And so,

$$x - z = (x - y) + (y - z) = mp + np = (m + n)p$$

And as $m + n \in \mathbb{Z}$ we have that $p \mid x - z$, and so $x \sim_p z$. Thus, $\sim_p$ is transitive.

Remark 14. *Generally if a set S has an equivalence relation $\sim$, then we will write $(S, \sim)$ to say that we are looking at S with the structure of the equivalence relation $\sim$ on it. Do be careful though, as people often surpress the equivalence relation (and other structures) and just write S and expect you to be aware of the equivalence relation by surrounding context.*

Well, we have this new notion of equality coming from this new equivalence relation structure, so what do we do now? Given a set S and a an equivalence relation $\sim$ on S, we can classify or seperate the elements of S into subsets of elements that considered equivalent under $\sim$.

Definition 68. *Given a set S and an equivalence relation $\sim$ on S we call the* equivalence class *of an element $x \in S$ the set of all $y \in S$ with $x \sim y$.*

$$[x] = \{y \in S \mid x \sim y\}.$$

We call x the representative *of the class $[x]$.*

Example 61. *Let us look at the real numbers $\mathbb{R}$ with the equivalence relation of $=$ on it.*

In this case, for any $x \in \mathbb{R}$, we have

$$[x] = \{x\}$$

i.e. the equivalence class of x just contains x itself. This is what we mean when we say $=$ is the most rigid or strongest of equivalence relations. Equals, $=$ is restrictive, an object will only be seen as equal to object if they are identical, i.e. the equivalence class is not different from the object itself really. Under this paradigm we can not group multiple objects by a more general property, we must see all distinct objects as distinct.

Example 62. *Once again, define $S = \{$people living in the United States$\}$ and the relation $\sim$ on S by saying $x \sim y$ if x lives in the same state as y. We saw earlier that $\sim$ is an equivalence relation. Now, for example, I live in California (currently). Let us find the equivalence class of Bob.*

$$[Bob] = \{y \in S \mid y \sim Bob\}.$$

Thus, the set of all people y such that $y \sim Bob$, i.e. the set of all people y that live in the same state as Bob. Well, as Bob lives in California, so

$$[Bob] = \{y \in S \mid y \text{ lives in California}\}.$$

And thus $[Bob]$ is just the set of all people living in California, and Bob is the representative of this set.

Remark 15. *i). For example, let's say there is another person living in California named Kurt. Well, similar to the above $[Kurt]$ is just the set of all people living in California with the representative Kurt. Thus, we see*

$$[Bob] = [Kurt].$$

The important thing to note is here is that the equivalence classes $[Bob]$, $[Kurt]$ *are equal even though Bob and Kurt are not the same person. (Bob* $\neq$ *Kurt) Note however that Bob* $\sim$ *Kurt. So unlike our prior example with* $=$, a more general equivalence relation let's us see objects that share a certain property as identical even though the objects themselves are not identical.

ii). In the example above, Bob was the representative of $[Bob]$. *In many cases, it is simpler to refer to* $[Bob]$ *by the property all elements in this set share instead of by the name of a representative. Thus, we will describe* $[Bob]$ *not by the name of some element* $x \in S$, *but by the property of all* $x \in [Bob]$ *share. Namely we will call*

$$[Bob] = [California],$$

as all people in the equivalence class of Bob have the property of living in California. The important thing to note here is that California $\notin S$, *and therefore the object California does not fit into the definition of a representative, but we will use it regardless as it is a better descriptor of all elements in the equivalence class of Bob.* We often use the common property all objects in an equivalence class share to describe the class, even if the property does not make sense as an element in the original set.

Example 63. *Once again, let* $S = \{$*people living in the United States*$\}$ *and define the relation* $\sim$ *on* S *by* $x \sim y$ *if* x *and* y *have the same age. It can be shown that* $\sim$ *is an equivalence relation. Similar to the previous example, it is easier to describe equivalence classes in this examples by a different object than an explicit representative.*

To be more clear, if Steve is 19 years old, as the equivalence class containing Steve is precisely the set of all 19 year olds, it makes more sense to describe the equivalence class of Steve by 19 than by the representative Steve. Once again, 19 is not technically a representative as $19 \notin S$, *but it is a better descriptor of all elements of* $[Steve]$ *than Steve.*

Example 64. *Let* $S = \mathbb{Z}$, *and fix* $p \in \mathbb{Z}$. *Recall from the previous example that we showed* $\sim_p$ *is an equivalence relation on* $\mathbb{Z}$. *Let us find the equivalence class of the number 1.*

$$\begin{aligned}[1] &= \{y \in \mathbb{Z} \mid y \sim_p 1\}. \\ &= \{y \in \mathbb{Z} \mid p \mid y - 1\} \\ &= \{y \in \mathbb{Z} \mid y - 1 = mp \text{ for some } m \in \mathbb{Z}\} \\ &= \{y \in \mathbb{Z} \mid y = mp + 1 \text{ for some } m \in \mathbb{Z}\}\end{aligned}$$

Thus,

$$[1] = \{y \in \mathbb{Z} \mid y = mp + 1 \text{ for some } m \in \mathbb{Z}\}.$$

In general, we have

$$[n] = \{y \in \mathbb{Z} \mid y = mp + n \text{ for some } m \in \mathbb{Z}\}.$$

It is important to notice that the equivalence class of 1 (or n*) is precisely the collection of all integers that are 1 more (*n *more) than a multiple of* p.

To be more clear, in the case that $p = 5$, *we have*

$$\begin{aligned}[1] &= \{y \in \mathbb{Z} \mid y = 5m + 1 \text{ for some } m \in \mathbb{Z}\} \\ &= \{\dots, -9, -4, 1, 6, 11, \dots\}\end{aligned}$$

So, $[1]$ *is just the collection of all integers that have a remainder of* 1 *when divided by 5. In fact, all of the equivalence classes mod 5 can be thought of this way. There are only* 5 *equivalence classes (as there are only 5 possible remainders)*

$$[0], [1], [2], [3], [4],$$

While it is true that $[3] = [8]$ *as* $3 \sim_5 8$*, we prefer using the representative 3 when describing this class as 3 is the remainder when dividing by 5, not 8. So, in general, when dealing with the classes of mod p, for some* $p \in \mathbb{Z}$*, the classes should be thought of as collections of integers that have the same remainder when divided by p, and the representative of each class will just be the remainder.*

We have seen one of the uses of equivalence classes. The integers $\mathbb{Z}$ *is an infinite collection of numbers, but when thought of under the notion of equality brought about by* $\sim_p$*, we see that* $\mathbb{Z}$ *breaks into p pieces,* $[0], [1], \ldots, [p-1]$*. So, equivalence classes can be used to significantly reduce the size of infinite sets. (In other words, from the view of division, there are really only p numbers.)*

Example 65. *Let* $I([a,b])$ *denote the colllection of integrable functions over the interval* $[a,b]$*, we define* $\sim$ *on this set as* $f \sim g$ *if and only if*

$$\int_a^b [f(x) - g(x)]dx = 0$$

then one can check that $\sim$ *is an equivalence relation on* $I([a,b])$*.*

In particular, if $f(x) = 0$ *is the zero function on* $[0,1]$*, then inside of* $I([0,1])$ *we have that*

$$[f] = \left\{ g(x) \in I([0,1]) \mid \int_0^1 g(x)dx = 0 \right\}$$

i.e. the equivalence class of $f(x)$ *is the collection of all functions* $g(x)$ *on* $[0,1]$ *whose integral is 0 over* $[0,1]$*. Note that this contains a lot of nonzero functions:*

- *For example the function*
$$g(x) = \begin{cases} 1 & x = \frac{1}{2} \\ 0 & x \neq \frac{1}{2} \end{cases}$$
and generally any function with only a finite number of point discontinuities away from the 0 function.

- *For example, a variant of the Dirichlet function*
$$h(x) = \begin{cases} \frac{1}{b} & x \in \mathbb{Q} \cap [0,1],\ x = \frac{a}{b},\ gcd(a,b) = 1 \\ 0 & x \in [\mathbb{R} \backslash \mathbb{Q}] \cap [0,1] \end{cases}$$
which is discontinuous everywhere on $[0,1]$*.*

- *The function* $\sin(2\pi x)$*, and generally any function* $h(x)$ *in which* $h(x) = F'(x)$ *for some function* $F(x)$ *with* $F(0) = F(1)$*.* [80]

So, we see that the equivalence class of f contains many functions, including functions with discontinuities and functions that are as smooth or a differentiable as we please, but from the perspective of integrating from 0 to 1, they are all identical, all equal to 0.

[80] This is an early version of canceling/ignoring divergences that comes from h having a potential function with particular boundary properties

The following result collects some properties of equivalence classes. In particular it says explicitly what we have already remarked before, which is two objects can have an identical equivalence class even if the objects are not identical, but it also tells us that unrelated or 'unequal' elements under the equivalence relation have disjoint classes. This is similar to the notion of partitioning a set and we will see this connection later

Theorem 136. *Let S be a set, and let $\sim$ be an equivalence relation on S. Then,*

a). $[x] = [y] \iff x \sim y$.

b). $[x] \cap [y] = \varnothing \iff x \nsim y$.

Proof. Proof of a). $\Rightarrow$ Assume that $[x] = [y]$. Then $x \in [x] = [y]$, and thus $x \sim y$.

$\Leftarrow$ Assume that $x \sim y$. Take $z \in [x]$. Then $z \sim x$ and $x \sim y$, hence by transitivity of the relation $\sim$, we have that $z \sim y$. Thus, $z \in [y]$. This shows that $[x] \subseteq [y]$. Taking $z \in [y]$ and following the same proof shows that $[y] \subseteq [x]$. Thus, $[x] = [y]$.

Proof of b). $\Leftarrow$. We proceed by proof by contrapositive. Thus assume that $[x] \cap [y] \neq \varnothing$. Thus, there exists $z \in [x] \cap [y]$. So, $z \sim x$ and $z \sim y$. By symmetry of $\sim$ we have that $x \sim z$, and by transitivity $x \sim z$ and $z \sim y$ imply that $x \sim y$. As $[x] \cap [y] \neq \varnothing$ implies $x \sim y$, the contrapositive states that: If $x \nsim y$, then $[x] \cap [y] = \varnothing$.

$\Rightarrow$ We prove this result by contradiction. Assume that $[x] \cap [y] = \varnothing$ and $x \sim y$. As $x \sim y$, part a). shows that $[x] = [y]$. But then

$$[x] = [x] \cap [y] = \varnothing.$$

This is a contradiction as $x \in [x] = \varnothing$. Thus, it must be that $[x] \cap [y] = \varnothing$ implies that $x \nsim y$. $\square$

Remark 16. *What this result says is that equivalence classes have no overlap. Either two classes are equal because their representative are equivalent or they share no element in common and their representatives are not equivalent. We saw this in our examples above.*

- *When $S = \{$people living in the United States$\}$ and the equivalence relation $\sim$ was two people living in the same state, we saw that the classes of $\sim$ were precisely the fifty states of the united states and that no two states share an element.*
- *When $S = \mathbb{Z}$ and the equivalence relation was $\sim_p$, we saw that the equivalence classes were $[0], [1], [2], \ldots, [p-1]$. It is clear that no two distinct classes share an element as any integer only has one remainder when divided by p.*

Equivalence classes have one remaining property that is very nice. I'll first explain this in the abstract and then give an explicit example. Let S be a set and $\sim$ an equivalence relation on S. We have seen earlier that any equivalence class $[x]$ can have multiple representatives defining the same class, but let us suppose there is a natural choice of representative of each class. Call $R = \{x \in S \mid x \text{ is a representative of } [x]\}$ where we have picked one representative for each equivalence class. Thus, R is the set of representatives. [81] Then S can be represented in the following manner

$$S = \bigcup_{x \in R} [x].$$

[81] The existence of this set came from the Axiom of Choice.

and because of the previous theorem $[x]$ and $[y]$ are disjoint if $x \neq y$ (remember as $x, y \in R$, $x \neq y$ is enough to justify $x \nsim y$ as x and y are the sole representatives chosen for the classes $[x]$ and $[y]$ respectively.)

Thus, an equivalence relation $\sim$ on a set S gives us a way to break S into pieces (the classes under $\sim$) that do not overlap or share elements. And, all of S is broken into pieces, any element $z \in S$ lies in some equivalence class (namely $[z]$ itself, but z may not be the chosen representative of $[z]$.)

Example 66. *Let us return to the modular arithmetic example on $\mathbb{Z}$. Thus, fix $p \in \mathbb{Z}$ and define $\sim_p$ as $x \sim_p y$ iff $p|(x-y)$. The classes of $\sim_p$ are $[0], [1], \ldots, [p-1]$ and each class has a natural representative. The representative of $[i]$ for $0 \leq i \leq p-1$ is i itself as i is the remainder when any element $x \in [i]$ is divided by p. Thus, the set of representatives is $R = \{0, 1, \ldots, p-1\}$. And so,*

$$\mathbb{Z} = \bigcup_{k \in \{0,1,\ldots,p-1\}} [k].$$

In particular, when $p = 5$, the natural representative of $[18]$ is 3. The set of representatives is $R = \{0, 1, 2, 3, 4\}$, and

$$\mathbb{Z} = \bigcup_{k \in \{0,1,2,3,4\}} [k] = [0] \cup [1] \cup [2] \cup [3] \cup [4].$$

It turns out that this concept is used often in mathematics. It is very useful to be able to break a set (especially if it is infinite) into pieces that do not overlap in a manner in which each piece is described by some property and every element of the set falls into one and only one piece. Of course, this being an important property and all, it has a name

Definition 69. *Given a set S, a* **partition** *of S is a collection of sets $\{A_i\}_{i \in \Lambda}$ with the properties*

i). $\bigcup_{i \in \Lambda} A_i = S$.

ii). $A_i \cap A_j = \varnothing$ if $i \neq j$.

Theorem 137. *Given a set S, any equivalence relation $\sim$ on S induces a partition of S.*

Proof. We pretty much proved this result in our remark and theorem from earlier, but I will restate the important parts here. For $\sim$, we let x_i be the chosen representative of the class $[x_i]$ for $i \in \Lambda$ where Λ is some index set. Then we have

a). $\bigcup_{i \in \Lambda} [x_i] = S$.

b). $[x_i] \cap [x_j] = \varnothing$ if $i \neq j$

Thus we see that $\{[x_i]\}_{i \in \Lambda}$ is a partition of S. □

These concepts, equivalence relation and partition, feel very similar, and in fact the result above states that having one gives the other. So, of course the natrual question is does this connection go the other way, i.e. does a partition on a set S induce an equivalence relation on S. The answer is in the affirmative as we will see below.

Theorem 138. *For a set S and a partition $P = \{A_i\}_{i \in \Lambda}$ on S, define a relation $\sim_P$ on S by the following. For $x, y \in S$, call $x \sim_P y$ if x and y lie in the same partition element, i.e. $x, y \in A_j$ for some $j \in \Lambda$. If $\sim_P$ is defined in this manner, then $\sim_P$ is an equivalence relation. Also, for $x \in S$, if $x \in A_i$, then $[x] = A_i$ with respect to the equivalence relation $\sim_P$.*

Proof. To prove our result we must simply check that $\sim_P$ is reflexive, symmetric, and transitive.

i). Take $x \in S$. As P is a partition, $x \in A_k$ for some $k \in \Lambda$. As x is clearly in the same partition element of itself, we have that $x \sim_P x$. Thus, $\sim_P$ is reflexive.

ii). Take $x, y \in S$ and assume $x \sim_P y$. Thus, x and y lie in the same partition element. Clearly, y is then in the same partition element as x, so $y \sim_P x$. Hence $\sim_P$ is symmetric.

iii). Take $x, y, z \in S$ and assume $x \sim_P y$ and $y \sim_P z$. Thus, x and y are in the same partition element $x, y, \in A_k$ for some $k \in \Lambda$. Similarly, y and z are in the same partition element $y, z \in A_l$. As $y \in A_k$ and $y \in A_l$, we have that $A_k \cap A_l \neq \varnothing$. Thus, by the contrapositive, it must be the case that $k = l$. Thus, $x, z \in A_k$. So, $x \sim_P z$, and thus $\sim_P$ is transitive.

Given $x \in S$, as $\{A_i\}_{i \in \Lambda}$ is a partition of S, it is clear that $x \in A_j$ for some j. By the definition of $\sim_P$, if $y \sim_P x$, then $y \in A_j$. Thus, $[x] \subseteq A_j$. Similarly, by definition of $\sim_P$, and $y \in A_j$ has $y \sim_P x$, thus $A_j \subseteq [x]$. □

Well-Definedness

Suppose we had a function mapping a set X into a set Y, i.e $f : X \to Y$. Further suppose that X had an equivalence relation $\sim$ defined on it. For shorthand call $[X]$ the set of equivalence classes of X under $\sim$. At this point, there are two natural questions that could be asked.

1). Does f extend to a function from the set of equivalence classes on X to Y i.e. does the mapping $f : [X] \to Y$ make sense as a function.

2). Does f require any extra properties for it to carry the equivalence relation structure from X to Y. In other words, X has an equivalence relation, and f maps X to Y, does f map the equivalence relation on X over to Y somehow.

Let us look at question 1. The answer has to do with how we define a function on a set of equivalence classes. For $[z] \in [X]$, the natural way to define $f([z])$ is given as follows.

$$f([z]) = f(z).$$

Well, in the equivalence relation $\sim$ on X, say $w \in X$ and suppose $w \sim z$and $w \neq z$. Thus, w is another representative of $[z]$ that does not equal z. So, we could define $f([z])$ as

$$f([z]) = f(w).$$

Thus, we reach our issue with question 1. What if $f(w) \neq f(z)$? Remember the defining property of a function is that one input maps to one output. As both w and z can claim to be the representative of $[z]$ (as can any $y \in X$ with $y \sim z$.), if $f(z) \neq f(w)$, then $f([z])$ has two distinct outputs associated to the input $[z]$. In this case f can not be extended as a function on $[X]$.

This case describes what we would require for f to extend to a function on $[X]$. We require that $f(x) = f(y)$ for any two elements $x, y \in X$ with $x \sim y$, i.e. elements in the same equivalence class must map to same output for the extension of f to be defined. This is precisely the notion of **well–definedness**.

Definition 70. *Given a set X with an equivalence relation $\sim$ and a function $f : X \to Y$, we say that f is* **well-defined** *on $[X]$ if the output of $f([x])$ is independent of the choice of representative of $[x]$. This is equivalent to the notion that f can be extended as a map, $f : [X] \to Y$ and the extension is also a function.*

Remark 17. *Often times the well definedness of f on $[X]$ is very much dependent on Y even though Y does not show up in the definition. This will come up momentarily.*

Example 67. *Let S once again be the set of people living in the United States, and let $\sim$ be the equivalence relation on S given by $x \sim y$ if x, y live in the same state. Let Y be the set of colors (visible spectrum). We can define a function $f : S \to Y$ by mapping each person to their eye color (well, technically we should define f to be the average RGB color value of the left and right eye as some people have heterochromia) Anyways, it is clear that f is a function. Now, we ask does f extend to a map from $[S]$ to Y. In other words* is f well–defined on $[S]$?

Well, clearly not. I'm sure at least two people in this class have different eye colors, and yet each one of us could act as a representative for the equivalence class of California. Thus f is not independent of the choice of representative from our class. So f is not well–defined on $[S]$.

Example 68. *Another example by which we need the notion of well–definedness is something you have also known your whole life: Addition! The symbol $+$ is really a function. As a function $+$ takes in two integers and returns one integer. More formally,*

$$+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}, \qquad +(m, n) = m + n,$$

i.e. $+$ is a function from the cartesian product of $\mathbb{Z}$ with itself to $\mathbb{Z}$.

Well, what if $\mathbb{Z}$ has the equivalence relation $\sim_p$ put on it? Does addition $+$ extend? Is $+$ well–defined on the set of equivalence classes of $\mathbb{Z}$ under $\sim_p$.

When we are often dealing with the set $\mathbb{Z}$ under the equivalence relation $\sim_p$, we use the shorthand $\mathbb{Z}_p$ to mean the set of equivalence classes of $\mathbb{Z}$ under $\sim_p$.

To be less formal for a moment, we know we can add whole numbers and get a whole number back. Right now we are just asking if there is a sensible way to reliably add equivalence classes in $\mathbb{Z}_p$.

So, as a map is $+ : \mathbb{Z}_p \times \mathbb{Z}_p \to \mathbb{Z}$ well defined? In other words, is $+$ independent of the representatives chosen for two equivalence classes. In short: NO.

Recall for $[m], [n] \in \mathbb{Z}_p$ then the extension of $+$ is defined to be

$$+([m], [n]) = +(m, n) = m + n.$$

And for example, let's let $p = 5$ and $m = 3$, $n = 2$. Then $8 \in [3]$ is another representative of $[3]$, and $12 \in [2]$ is another representative of $[2]$, and we can clearly see that $+$ is not independent of the choice of representative,

$$+([2], [3]) = +(2, 3) = 2 + 3 = 5$$
$$+([2], [3]) = +([12], [8]) = +(12, 8) = 12 + 8 = 20.$$

Thus $+$ is NOT well defined as a map from $\mathbb{Z}_p \times \mathbb{Z}_p$ to $\mathbb{Z}$.

This is what I meant in the note above, the image space Y, in our case, $Y = \mathbb{Z}$, greatly effects whether or not functions are well-defined. So, we ask a similar question, but change where the map, $+$, goes to.

As a map, is $+ : \mathbb{Z}_p \times \mathbb{Z}_p \to \mathbb{Z}_p$ well defined? YES.

Take $[m], [n] \in \mathbb{Z}_p$. Let $r, s \in \mathbb{Z}$ be different representatives for the classes $[m]$ and $[n]$ respectively. Thus, $r \in [m]$ and $s \in [n]$. This implies the existence of $k, l \in \mathbb{Z}$ such that

$$r - m = kp$$
$$s - n = lp$$

Thus, we see that r, s have the form $r = kp + m$ and $s = lp + n$. We now ask if $+$ is independent of the representatives of $[m]$ and $[n]$ chosen. Well,

$$+([m], [n]) = +(m, n) = m + n$$
$$+([r], [s]) = +(r, s) = r + s = kp + m + lp + n = m + n + (k + l)p$$

Thus, by looking at the difference,

$$+([r], [s]) - +([m], [n]) = (k + l)p.$$

We see that $+([m], [n])$ and $+([r], [s])$ differ by a multiple of p. Thus in the image space $Y = \mathbb{Z}_p$, both $+([m], [n])$ and $+([r], [s])$ belong to the same class, i.e.

$$[+([m], [n])] = [+([r], [s])].$$

Thus, $+ : \mathbb{Z}_p \times \mathbb{Z}_p \to \mathbb{Z}_p$ is well defined. In particular the above showed that as $+$ is independent of the representatives chosen we have a simpler formula for modular addition,

$$[m] + [n] = [m + n].$$

Do note that the other way to view this well defined addition map is

$$+([a], [b]) = [+(a, b)]$$

i.e. there is some commutativity between the notion of class and the function $+$ itself, i.e. the sum of the classes is the class of the sum. In fact, the answer to question 2 tells us we really should not expect any other answer.

Turning to question 2 now. Suppose we have a function $f : X \to Y$ and X has the equivalence relation $\sim$ on it. We can define the relation $\sim_f$ on Y by the following: For $c, d \in Y$, we say $c \sim_f d$ if and only if:

- Either both c and d are not in the range of f
- Or $\exists a, b \in \text{Dom}(f)$ with $f(a) = c$ and $f(b) = d$ and $a \sim b$.

We can see that $\sim_f$ is an equivalence relation on Y by

1. *Reflexive* - For $a \in Y$, if a is not in the range of f then $a \sim_f a$. Otherwise $a = f(x)$ for some x in the domain of f and $x \sim x$ as $\sim$ is an eq. rel, so $a \sim_f a$.

2. *Symmetric* - For $a, b \in Y$ if we assume that $a \sim_f b$, then either a, b are both not in the range of f and so $b \sim_f a$ automatically. Or, there exists x, y in the domain of f with $f(x) = a$, $f(y) = b$ and $x \sim y$. However as $\sim$ is an equivalence relation on X, it is symmetric, so $y \sim x$ and thus $b \sim_f a$, so $\sim_f$ is symmetric.

3. *Transitive* - For $a, b, c \in Y$ and we assume $a \sim_f b$ and $b \sim_f c$. If a is not in the range of f then b can not be in the range either by our assumption, and this means c can not be in the range of f by the second assumption, so in this case $a \sim_f c$. Otherwise, there exist $x, y, z \in X$ with $f(x) = a$, $f(y) = b$ and $f(z) = c$. By our assumption $x \sim y$ and $y \sim z$, so by the transitivity of $\sim$ we have $x \sim z$ which implies that $a \sim_f c$, so $\sim_f$ is transitive.

And so we see that $\sim_f$ is an equivalence relation on Y. If we call the collection of equivalence classes of $(X, \sim)$ by $[X]$ and the collection of equivalence classes of $(Y, \sim_f)$ by $[Y]_f$, then we see that $f : X \to Y$ can be extended to $f : [X] \to [Y]_f$. In fact, this is the opposite picture of question 1. In question 1 we were looking for requisite properties of a function f to make sure it would be well-defined on a particular set Y, but in this question what we find is that by using f to construct a particular equivalence relation on Y, we will always get a well defined map in the extension $f : [X] \to [Y]_f$.

In particular, under $\sim_f$ the equivalence classes of Y are $Y \backslash \text{Ran}(f)$ and for a y in the range of f, i.e. there exists x such that $y = f(x)$ we have

$$\begin{aligned}
[y]_f &= \{z \in Y | z \sim_f y\} \\
&= \{z \in Y | \exists d \in \text{Dom}(f)\ d \sim x\} \\
&= f([x])
\end{aligned}$$

i.e. the equivalence classes of $\sim_f$ are one class formed of elements not in the range of f, and then the images of the equivalence classes of X.

Interestingly enough, this process also works in reverse. Given $f : X \to Y$ and Y having an equivalence relation structure $\sim$ is enough to induce an equivalence relation $\sim_f$ on X. We define $\sim_f$ by saying for $x, y \in X$, $x \sim_f y$ if $f(x) \sim f(y)$. In other words, we consider points in the domain equivalent if the values they map to in the image are considered equivalent under the equivalence relation $\sim$ on Y. And note that this does allow for an extension $f : [X]_f \to [Y]$ in a well defined manner.

Appendix: Construction of $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$

Construction of $\mathbb{N}$

We will not go through the formal construction of the natural numbers here. We will cover the basic contruction of the naturals and leave the more important details to a later lecture (or a supplemental lecture). **fix this**

The natural numbers are described as very particular sets, for example,

$$0 = \varnothing, \quad 1 = \{\varnothing\} = \{0\}, \quad 2 = \{\varnothing, \{\varnothing\}\} = \{0, 1\}$$

Specifically, 0 is defined as the empty set and every other natural number is described as the set of all previous natural numbers. To make this more precise we have ...

Definition 71. *For a set x, we define the* successor *of x, often written $succ(x)$ or x^+ as*

$$x^+ = x \cup \{x\}$$

From this we can see that $1 = 0^+$ and $2 = 1^+ = 0^{++}$ so on and so on. This shows how we can build the natural numbers up from the empty set, however we will move on from here to other number systems. The remaining properties of the natural numbers: the inductive property, the algebraic structure of addition and left/right cancellation, and the linear ordering of $<$ on the naturals will be left for later

This later component may be very intense set theory wise, formally the natural numbers is the 'least' element of all inductive sets containing $\varnothing$, i.e. the naturals is typically taken as an abstract intersection of a large number of inductive sets.

Construction of $\mathbb{Z}$

We will define the integers using a very specific equivalence relation. In particular we will effectively define the operation of subtraction through addition. So we will begin with the set $\mathbb{N} \times \mathbb{N}$, i.e. the cartesian product of the natural numbers with itself, and we define the following relation $\sim$ on $\mathbb{N} \times \mathbb{N}$.

For two pairs $(a, b), (c, d) \in \mathbb{N} \times \mathbb{N}$ we will define $(a, b) \sim (c, d)$ if and only if $a + d = b + c$. We will check now that this is an equivalence relation on the set $\mathbb{N} \times \mathbb{N}$.

- *Reflexive* - For $(a, b) \in \mathbb{N} \times \mathbb{N}$ we have that $a + b = b + a$ as addition is commutative on $\mathbb{N}$ [82]. Thus $(a, b) \sim (a, b)$ and thus $\sim$ is reflexive.
- *Symmetric* - For $(a, b), (c, d) \in \mathbb{N} \times \mathbb{N}$ assume that $(a, b) \sim (c, d)$, so $a + d = b + c$. By commutativity we have that $a + d = d + a$ and $b + c = c + b$, thus $c + b = d + a$ which shows that $(c, d) \sim (a, b)$. Thus $\sim$ is symmetric.
- *Transitive* - For $(a, b), (c, d), (e, f) \in \mathbb{N} \times \mathbb{N}$, assume that $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. Thus $a + d = b + c$ and $c + f = d + e$. By adding these together we have

$$a + d + c + f = b + c + d + e$$

[82] this will be shown in the section on the construction of $\mathbb{N}$

As $d+c=c+d$ and by the commutativity of addition we have

$$a+f+c+d=b+e+c+d$$

And using right cancellation of addition [83] for $c+d$ we have $a+f=b+e$, thus $(a,b)\sim(e,f)$, so $\sim$ is transitive.

Now that we have that $\sim$ is an equivalence relation on $\mathbb{N}\times\mathbb{N}$, we can look at what the equivalence classes are for a general element. So, let us look at the equivalence class of (x,y).

$$\begin{aligned}[(x,y)] &= \{(c,d)\in\mathbb{N}^2 \mid (x,y)\sim(c,d)\}\\ &= \{(c,d)\in\mathbb{N}^2 \mid x+d=y+c\}\\ &= \{(c,d)\in\mathbb{N}^2 \mid y=x+(d-c)\}\end{aligned}$$

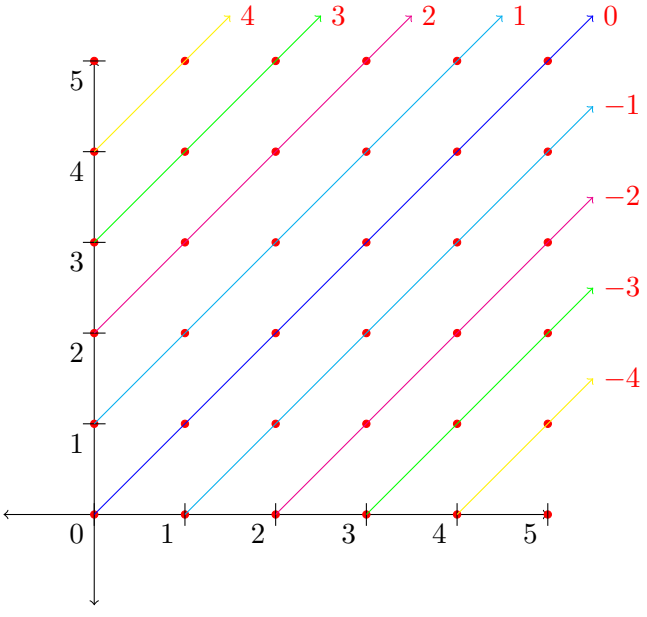
and so we can see that the equivalence class of (x,y) is precisely the line of slope 1 with y-intercept $y-x$ intersected with $\mathbb{N}^2$. In the following figure, we can see $\mathbb{N}^2$ drawn with the equivalence classes drawn as well. In this case, we identify each equivalence class with the y-intercept of the line, i.e.

$$[(m,n)]=n-m$$

In this sense we have defined the integers $\mathbb{Z}$ as $(\mathbb{N}^2,\sim)$. From this we can see

- The positive whole numbers along the y-axis, i.e. $[(0,n)]=n$.
- Zero at the origin $[(0,0)]=0$
- The negative whole numbers along the x-axis, i.e. $[(m,0)]=-m$.

Figure 1: Equivalence relation on $\mathbb{N}^2$



[83] Another property to be proved in the section on the construction of $\mathbb{N}$

What is left is to check the additive, multiplicative, and order structures on $\mathbb{Z}$. So let us check addition first.

Addition

Addition is already well defined on $\mathbb{N}^2$ in the same way we define vector addition, i.e.

$$(a,b)+(c,d)=(a+c,b+d)$$

i.e. componentwise addition. We want to show that this extends to a well defined operation on $\mathbb{Z}$, i.e.

$$[(a,b)]+[(c,d)]=[(a+c,b+d)]$$

in other words we have to make sure addition of the classes $[(a,b)]$ and $[(c,d)]$ is independent of the representatives taken from each class. So, let $(a,b),(A,B)\in[(a,b)]$ and $(c,d),(C,D)\in[(c,d)]$, and we have

$$(a,b)+(c,d)=(a+c,b+d),\qquad (A,B)+(C,D)=(A+C,B+D)$$

As $(a,b)\sim(A,B)$ we have $a+B=A+b$ and similarly we have $c+D=C+d$, by adding these we have

$$\begin{aligned}a+B+c+D&=A+b+C+d\\ a+c+B+D&=A+C+b+d\end{aligned}$$

and this shows that $(a+c,b+d)\sim(A+C,B+D)$. And this shows that the sum of $[(a,b)]$ and $[(c,d)]$ is independent of the choice of representative. Thus, addition on $\mathbb{Z}$ is well-defined by $[(a,b)]+[(c,d)]=[(a+c,b+d)]$.

The associativity and commutativity of addition follows from the associativity and commutativity of addition on $\mathbb{N}$, we also see that $[(0,0)]$ is the zero element as

$$[(a,b)]+[(0,0)]=[(a,b)],\qquad [(0,0)]+[(a,b)]=[(a,b)]$$

for any $[(a,b)]\in\mathbb{Z}$. We also have negatives or additive inverses for elements: for an integer $[(a,b)]$ its additive inverse is given by $[(b,a)]$ as

$$\begin{aligned}[(a,b)]+[(b,a)]&=[(a+b,a+b)]=[(0,0)]\\ [(b,a)]+[(a,b)]&=[(b+a,b+a)]=[(0,0)]\end{aligned}$$

In this sense subtraction is defined by

$$[(a,b)]-[(m,n)]=[(a,b)]+[(n,m)]$$

using $-[(m,n)]=[(n,m)]$.

Multiplication

Unfortunately, defining mulltiplication on $\mathbb{Z}$ is slightly more difficult. We can not simply multiply component wise to define multiplication instead we will define multiplication as

$$[(a,b)]\cdot[(m,n)]=[(bm+an,bn+am)]$$

and see that the multiplicative identity will be given by $[(0,1)]$. Multiplication on the integers will inherit associativity and commutativity from the natural numbers. Checking the well-definedness of this operation will be left to the reader.

Order Structure

We can extend the strict ordering $<$ on $\mathbb{N}$ to the integers by defining

$$[(m,n)] < [(p,q)] \iff m+q < p+n$$

Checking the transitivity of this relation is very similar to checking the transitivity of the original relation $\sim$ above. By nature of this definition, we have that $\mathbb{Z}$ will have the trichotomy law as $\mathbb{N}$ does and thus $<$ is a linear ordering on the integers.

This ends our discussion on the formal construction of $\mathbb{Z}$, moving forward we typically do away with the notation of $[(m,n)]$ and simply replace it with $n-m$.

Remark 18. *This conversation is not 'total'. The integers will also inherit left/right cancellation for multiplication of non-zero elements as well, but this is not proven here.*

Construction of $\mathbb{Q}$

Following how we constructed the integers, we will seek an equivalence relation on $\mathbb{Z}\times\mathbb{Z}$ to effectively define on division through multiplication. However, because division by 0 is not allowed, we will need to make one small alteration before we start, so we define

$$Q = \{(a,b) \in \mathbb{Z}^2 \mid b \neq 0\}$$

We now define the relation $\sim$ on Q by $(a,b) \sim (c,d)$ if and only if $ad = bc$. Let us check that is an equivalence relation.

- *Reflexive* - For $(a,b) \in Q$ we have $ab = ba$ from the commutativity of multiplication on $\mathbb{Z}$, thus $\sim$ is reflexive.
- *Symmetric* - For $(a,b),(c,d) \in Q$ and assuming $(a,b) \sim (c,d)$ we have $ad = bc$. By the commutativity of mulltiplication in $\mathbb{Z}$ we have $cb = da$ which shows $(c,d) \sim (a,b)$. Thus $\sim$ is symmetric.
- *Transitive* - For $(a,b),(c,d),(e,f) \in Q$ and assume that $(a,b) \sim (c,d)$ and $(c,d) \sim (e,f)$. Thus $ad = bc$ and $cf = de$. By multiplying both and making use of commutativity we have

  $$afcd = becd$$

  By definition of $(c,d) \in Q$ we have that $d \neq 0$, thus by right cancellation of multiplication in $\mathbb{Z}$ we have

  $$afc = bec$$

  Unfortunately, we can not directly make use of right cancellation in this case as there is no condition on c other than being an integer. What we can see is the following

  - *Case 1* - If $c \neq 0$, then we can use right cancellation and we have $af = be$ and thus $(a,b) \sim (e,f)$.
  - *Case 2* - If $c = 0$, then we have $ad = 0$ and $de = 0$, and as $d \neq 0$ this implies that $a = e = 0$, and in this case we trivially have $(a,b) \sim (e,f)$ (really $(0,b) \sim (0,f)$).

  In either case we see $(a,b) \sim (e,f)$ and thus $\sim$ is transitive.

We will define the rationals $\mathbb{Q}$ to be $(Q, \sim)$. Specifically each rational number will be an equivalence class of elements from Q. Informally we will think of

$$[(a,b)] = \frac{a}{b}$$

For example when we look at the equivalence class of $(1,2)$ we have

$$\begin{aligned}[(1,2)] &= \{(a,b) \in Q \mid (1,2) \sim (a,b)\} \\ &= \{(a,b) \in Q \mid 2a = b\}\end{aligned}$$

i.e. this equivalence class consists of all pairs from Q in which coordinate 1 is twice that of coordinate 2, i.e. all fractions who have a factor a in common and can be reduced to $\frac{1}{2}$. This is also why earlier, when we made use of rational numbers we included the condition $\gcd(a,b) = 1$ as this gave a manner of picking out a specific representative from each equivalence class.

Let us now check the addition, multiplication, and order structures on $\mathbb{Q}$.

Addition

Addition of two equivalence classes will be given by

$$[(a,b)] + [(c,d)] = [(ad + bc, bd)]$$

and this may appear strange at first until you remember that we require a common denominator for addition of fractions. What is left is to check that definition of addition is well-defined, i.e. independent of choice of representative.

So, let $(a,b), (A,B) \in [(a,b)]$ and $(c,d), (C,D) \in [(c,d)]$, the question is if

$$(ad + bc, bd) \sim (AD + BC, BD)$$

by our assumption we have $aB = bA$ and $cD = Cd$ so

$$\begin{aligned}(ad + bc)BD &= adBD + bcBD = aBdD + cDbB \\ &= bAdD + CdbB = ADbd + BCbd = (AD + BC)bd\end{aligned}$$

and this shows that $(ad + bc, bd) \sim (AD + BC, BD)$ thus addition is well defined in this manner. Similar to our discusiion about $\mathbb{Z}$, we will see that addition on $\mathbb{Q}$ is associative and commutative as addition on $\mathbb{Z}$ is.

The additive identity of 0 is given by $[(0,1)]$ as

$$\begin{aligned}[(a,b)] + [(0,1)] &= [(a1 + b0, b1)] = [(a,b)] \\ [(0,1)] + [(a,b)] &= [(1a + 0b, 1b)] = [(a,b)]\end{aligned}$$

And every rational number $[(a,b)]$ has an additive inverse given by $-[(a,b)] = [(-a,b)]$ as

$$\begin{aligned}[(a,b)] - [(a,b)] = [(a,b)] + [(-a,b)] &= [(ab - ba, b^2)] = [(0,1)] \\ -[(a,b)] + [(a,b)] = [(-a,b)] + [(a,b)] &= [(-ab + ba, b^2)] = [(0,1)]\end{aligned}$$

Multiplication

Multiplication of two equivalence classes will be given by

$$[(a,c)] \cdot [(b,d)] = [(ac, bd)]$$

and we need to check that this definition is well defined. Thus we take $(a,b),(A,B) \in [(a,b)]$ and $(c,d),(C,D) \in [(c,d)]$ and we need to check that

$$(ac, bd) \sim (AC, BD)$$

By assumption we have $aB = bA$ and $cD = dC$ and thus

$$acBD = aBcD = bAdC = bdAC$$

and this shows that $(ac, bd) \sim (AC, BD)$. Associativity and commutativity of $\mathbb{Q}$ will follow from the associativity and commutativity of multiplication on $\mathbb{Z}$.

The multiplicative identity of 1 is given by $[(1,1)]$ as

$$[(a,b)] \cdot [(1,1)] = [(a,b)]$$
$$[(1,1)] \cdot [(a,b)] = [(a,b)]$$

And for any $a \neq 0$, i.e. any nonzero rational, the multiplicative inverse of $[(a,b)]$ is given by $[(b,a)]$, i.e. $[(a,b)]^{-1} = [(b,a)]$ as

$$[(a,b)] \cdot [(b,a)] = [(ab,ba)] = [(1,1)]$$
$$[(b,a)] \cdot [(a,b)] = [(ba,ab)] = [(1,1)]$$

Order Structure

We can extend the strict ordering $<$ on $\mathbb{Z}$ to the rationals by defining

$$[(m,n)] < [(p,q)] \iff mq < np$$

Checking the transitivity of this relation is very similar to checking the transitivity of the original relation $\sim$ above. By nature of this definition, we have that $\mathbb{Q}$ will have the trichotomy law as $\mathbb{Z}$ does and thus $<$ is a linear ordering on the rationals.

Typically you will call the rationals with the structures we have found above a *totally ordered field*, $(\mathbb{Q}, +, \cdot, <)$.

Appendix: Dedekind Cut Construction of $\mathbb{R}$

Under construction

multiple proofs of AP in r (do in dedekind cut section),

Maybe make some mention of first order versus second order structures, interchanging of quantifiers.

Appendix: Surreal?

Under construction

mention the bit about how hypperreals let you 'be' at a limit. Multiple versions of completeness

Add in links to Korner and Propp paper on reverse analysis. Maybe an appendix on versions of completeness and equivalence proofs (don't go too crazy)

Teach limits and other things via hyperreals at some point?

ordinals, infinitesimals, surreals

References

[A] Tom M. Apostol, *Calculus, Vol. I*

[C] John B. Conway, *A first course in analysis.*

[R] Walter Rudin, *Principles of Mathematical Analysis.*

[S] Robert Strichartz, *The Way of Analysis.*

[T] Terence Tao, *Analysis I.*

[E] Herbert B. Enderton, *Elements of Set Theory*

[W] Stephen Willard, *General Topology*

[JB] John C. Baez, *The Octonions*

[LW] Peter A. Loeb & Manfred Wolff, *Nonstandard Analysis for the Working Mathematician*

[FM] François Monard, *105A Lecture notes, Winter 22*

[JP] James Propp, *Real Analysis in Reverse*